Interview: Massive Data Institute Postdoctoral Fellow Dr. Gaurav Sood on the Challenges and Opportunities of "Massive Data"

Jamie Obal and Austin Williams

From The New York Times and The Economist to Facebook, Twitter, and Reddit, the options for downloading and digesting news and information are endless in today's digital world. Technology is rapidly developing and being used in ways never previously imagined. Most significantly, technological progress has helped to address some of society's most pressing problems—finding new cures, building safer cities, and expanding economic opportunity. However, the pace at which technology, and the associated expansion of data collection, is moving has sparked a heated debate about how government should strike a balance between enforcing regulation, incentivizing competition, and protecting consumer privacy. For scholars dedicated to producing impactful policy research, the rise of big data marks a significant opportunity—and responsibility. Gaurav Sood joined the new Massive Data Institute (MDI) at the McCourt School of Public Policy in September 2014 as its first postdoctoral fellow. At the MDI, Sood plans to focus on estimating ideological positions of media sources by using a novel dataset of more than seven million news articles and television news transcripts. Sood is also exploring the effects of liberalizing regulations on broadband media. He recently spoke to the Georgetown Public Policy Review about his plans at the MDI, his perspective on the development of regulation within this field, his recently completed unpublished work on broadband Internet, and the growing influence of news media on public opinion.

As the first postdoctoral fellow of the Massive Data Institute (MDI) at the McCourt School of Public Policy, Dr. Gaurav Sood stands on the front lines of a significant change in academic research: the rise of big data. If the word "big" sounds vague, that is because it is. It is an imprecise term, just like "massive," which was adopted as a moniker for the Institute. Unfortunately, it is hard to find appropriate language to describe the large datasets that

social and political scientists are now using to deliver fresh understandings of society and human behavior. While the meaning of the term "big data" remains unclear, its benefits are clear. Louisville, Kentucky is combatting asthma with data by using GPS trackers in medical inhalers to see where residents experience the greatest difficulty breathing. The New York Police Department is deepening its understanding of where violent crime exists

Jamie Obal and Austin Williams interviewed Dr. Gaurav Sood on March 3, 2015. Jamie and Austin are Interview Editors at the Georgetown Public Policy Review and Master of Public Policy students at the McCourt School of Public Policy. Jamie, originally from Los Angeles, California, focuses on economic development and is interested in uplifting communities in urban and metropolitan areas. Austin, originally from McRae, Georgia, is pursuing an MBA at Georgetown as well to further his interest in public private partnerships on community development initiatives.

by crime-mapping aggregate criminal justice data. Amazon is boosting its profits by using customer data to recommend additional products that a customer "may also like." In a similar vein, the MDI is looking to leverage the power of massive data, with the goal of developing major advances in public policy by bringing together scientists like Sood, who can analyze these types of powerful knowledge resources, and policy practitioners from the government and nonprofit sectors, who can design and implement effective solutions.

The evolution from a brick and mortar world to a more digitized one is nowhere more apparent than at the home of the new MDI. "For the past two years, I've given up on reading conventional outlets. I don't spend too much time reading *The Washington Post*, or *The New York Times*. . . I find them to be quite dangerous actually. I am more liable to end up learning about how to spend 24 hours in Marrakech than the contents of Section 702 of FISA." Sood prefers browsing *The Economist* on his laptop instead. Technology has not just changed the way we consume news, but it has also inspired researchers, policymakers, and students to approach old problems with new innovative techniques and data-driven tools. Previously, terms like "scraping data," "cloud-computing," and "bootstrapping" were more likely to be found in a computer science class than a public policy one. However, with the big data revolution and technology blurring the lines between the physical and digital worlds, these technical terms are starting to be more commonplace at public policy schools. Classes such as Sood's "Introduction to Data Science" are ensuring that the next generation of policymakers is up to speed with the skills to collect, manage, and analyze large datasets.

REGULATION OF MASSIVE DATA

When Sood was pursuing an undergraduate degree in Computer Science at Rutgers University in the early 2000s, he rarely encountered studies with sample sizes of more than a few thousand. In the field of psychology, studies were typically published with sample sizes of less than 100. By the time he finished his PhD in communications at Stanford in 2011, however, data usage had exploded. Data has become massive. Indeed, Sood often works with information sets so large that they do not fit on a computer hard drive. Today, researchers often need to store information across a series of servers, a process he referred to as "data gymnastics."

With big data comes big responsibility. The rise of big data has made waves through the private sector where concerns over consumer privacy and exploitative business practices are raising eyebrows. Media providers ranging from Facebook to Direct TV have been scrutinized for leveraging private customer information for commercial gain. According to a 2014 Pew report analyzing public opinion on security and privacy, researchers found that 91 percent of the respondents on the same survey agreed that consumers have "lost control" over how personal information is accessed and utilized (Madden 2014). Yet 55 percent of the same survey participants agreed that they were willing to share personal information in exchange for free online services. Sood is a realist about these abuses. "It is the reality of a capitalist economy. We have ceded some freedom to businesses. Sometimes data analysis in the private sector enhances people's lives, and sometimes the impact is more negative. These are complex issues to debate." Bruce Schneier, a fellow at the Berkman Center for Internet

and Society at Harvard Law School, cautions of the dangers associated with this economic trade off. In a Ted Talk on security, Schneier—dubbed by *The Economist* as a "security guru"—calls security "a feeling and a reality" (Schneier 2010). He notes, "You can feel secure even if you're not, and be secure even if you don't feel it." Many Americans, according to Schneier, respond to the feeling of individual security, but fail to align their feelings with the reality of national security threats. Hackers exist all over the world, sometimes motivated by things other than profit and threatening the physical safety of Americans and the US government. Regardless of the complexities, Sood thinks governments will respond with more regulation around these issues in the future.

In the post-Snowden era that left Americans debating how much privacy they were willing to relinquish in exchange for national security and technological progress, the Obama Administration has strengthened its efforts to create a regulatory framework that protects consumers' privacy. As a follow up to the 2012 Consumer Data Privacy in a Networked World report, the Administration released a discussion draft of the Consumer Privacy Bill of Rights Act of 2015 (White House 2012; White House 2015). The proposed language aims to complement existing regulations by setting guidelines for how companies can collect and use personal data. However, privacy advocates argue that the draft legislation falls short. In a letter addressed to President Obama, a coalition of 14 consumer privacy groups, including the DC-based non-profit Center for Democracy and Technology, criticized the draft legislation "gives companies broad leeway" and should "afford stronger regulatory and enforcement authority to the Federal Trade Commission"

(Center for Data and Technology 2015). Congress is also weighing in on the privacy debate. In a Senate Commerce, Science, and Transportation Committee hearing on how the internet has heightened connectivity, Chairman John Thune advised policymakers to "resist the urge to jump head first into regulating this dynamic marketplace," and emphasized Congress' role to "ensure that any government efforts to protect consumers are tailored for actual problems and avoid regulatory overreach" (2015). Ranking Member Bill Nelson characterized the idea of "overregulating" as a "red herring," and urged for "conversation and cooperation between the FTC and the industry" in order to address concerns of consumer privacy and network security.

The Federal Trade Commission (FTC), the US federal agency tasked with protecting consumers and promoting competition, recognizes the privacy challenges involved with harnessing the power of big data. The FTC recently created the Office of Technology Research and Investigation, which will conduct investigative research on emerging technology issues including privacy, data security, connected cars equipped with internet access, smart homes, algorithmic transparency, emerging payment methods, big data, and the Internet of Things—the vast physical network of technology that enables devices to be connected to the internet. With technology moving at lightning speed and plugging in to almost every aspect of our daily lives, legal institutions like the FTC are encouraged to collaborate across sectors. In a keynote address at a Georgetown University forum entitled "Privacy Principles in the Era of Massive Data," Federal Trade Commissioner Maureen Ohlhausen called for "a coalition of academics, regulators, businesses, and con-

sumers" to tackle privacy concerns surrounding big data. Among the growing concerns for regulators is how big data can become a tool of exclusion. For example, algorithms can use an individual's neighborhood to generate different discounts for the same product and eligibility scores for housing or employment. Low income and underserved communities are particularly susceptible to this "digital redlining," and the 2014 White House Big Data Report warns that the "increasing use of algorithms to make eligibility decisions must be carefully monitored for potential discriminatory outcomes for disadvantaged groups, even absent discriminatory intent." Just as redlining in the financial lending markets was outlawed through anti-discrimination laws in the Community Reinvestment Act of 1977, we may expect – and hope – to see legislative action to ensure the same protections in the online marketplace.

JOURNEY TO GEORGETOWN

Sood's path into academia was not inspired by a passion for research or a desire to teach, but rather a thirst for learning. "It is a mystery," he says, describing his decision to pursue a PhD Academia in India—he emigrated from India more than a decade ago—which, like in many developing countries, is an unorthodox career choice. "Since no one among my friends or family had a PhD, it was something that I didn't understand completely. I sort of went into it blindly." And yet, blindly or otherwise, his curiosity has led him all the way to being a founding member of the MDI at Georgetown University.

Sood will use his time under the fellowship to pursue a range of research interests. "I broadly see myself as a social scientist. I don't feel

constrained in terms of what kind of things I should want to study. There are lots of topics that I'm curious about, and data can shed light on a variety of questions." Some of Sood's previous work has focused on political partisanship. While exploring correlations between how people feel about the Democratic and Republican parties and their policy positions, he had one of the "eureka moments" he hoped a life devoted to learning might bring: to his surprise, his analysis showed that these correlations tend to be relatively trivial. In other words, party choice in America often has little to do with our actual beliefs. In reality, the overlap between Democrats and Republicans on policy positions are actually quite extensive, but average Americans usually do not see the similarities. To quote Sood, "They think the other party lives on Pluto, which is not even a planet anymore, right?" The negative associations, even feelings of hate, for the opposing party are often not founded on deep moral or ideological differences. "We know that hatred and love between people is sometimes determined by really trivial things. For instance, color of skin, why small—many a times substantively immaterial—differences divide people deeply is one of the oldest questions in social science. Politics is just another example of that."

POLITICAL POLARIZATION, THE MEDIA, AND NET NEUTRALITY

The rise of partisan cable news has exacerbated the divide between Democrats and Republicans. In a 2013 study analyzing the effect of access to ideologically distinctive news sources, Georgetown researchers Daniel Hopkins and Jonathan Ladd concluded "citizens often respond to political messages from candidates and news outlets differently depending on

their partisan predisposition." Partisan cable news such as the liberal MSNBC or conservative Fox News reinforce the voting tendencies of voters who already share the network's ideological worldview. This phenomenon not only heightens the partisan news media, but it also leaves a more polarized electorate. McCourt School Professor Micah Jensen, whose research focuses on identity politics and political behavior, suggests that the "force that mobilizes members of groups to political action can also increase discriminatory attitudes and behaviors between groups; a process which may help to explain our increasingly polarized politics."

In some of his recent research, Sood expands his work on political polarization to examine how access to broadband Internet affects political attitudes. A seemingly trivial but relevant point is that people with access to broadband internet consume more media than those without it. Americans spend much of their time consuming media. A 2013 survey of American time use by the Bureau of Labor Statistics found that Americans spent more than half of their leisure time—approximately 2.8 hours per day—watching TV, trumping time spent socializing with friends or attending social events. Nielsen, a global marketing research firm, released a US Digital Consumer Report (2014) revealing that the average American spends approximately 60 hours per week consuming news media across an average of four digital devices. Media consumption has "become a full time job," Sood reflects. The Telecommunications Act of 1996 gave local governments jurisdiction to regulate and price the rights of way—the physical space, wires, conduits, poles, and corridors passing through public land that enable broadband providers to build an internet infrastructure in a given neighborhood. Exploiting differences in rights of way and broadband availability, Sood discovered that "going from no internet to dial-up to broadband has a sizeable effect on media consumption." His research suggests that regulations that lower the cost of broadband internet polarize rank-and-file partisans, likely by increasing their exposure to partisan news media (2015). "Broadband access causes people to consume a lot more media, much of it non-political. But the little additional political media they consume polarizes them."

From a supply perspective, greater broadband access increases competition. This increased competition "depletes the quality of news outlets." With news providers looking to stand out from the competition, viewers should expect to find a greater menu of entertainment and sports news—not better political coverage. Americans anxious to satisfy their appetite for streaming high-definition videos on YouTube and Netflix, or joining an online gaming session, may rejoice over the Federal Communications Commission (FCC) net neutrality ruling in 2015 to regulate high-speed internet. Classifying broadband internet as a public utility, this ruling prevents providers from charging higher fees for faster Internet speeds. The ruling has left regulators at opposite sides of the table. Dwelling on his experience as an entrepreneur, FCC Chairman Tom Wheeler defended the ruling, citing that a fast, fair, and open Internet is imperative to innovation and human expression. On the other hand, FTC Commissioner Joshua D. Wright testified that the net neutrality ruling "does not make sense from an economic perspective" and leaves consumers "worse off" (2015). Wright criticizes the FCC for threatening "to strip the FTC of

its jurisdiction to regulate broadband providers of its consumer protection mission" by classifying Internet as a public utility. Wright dismissed Wheeler's "gatekeeper" justification, challenging that "no broadband provider can be viewed as a gatekeeper to anything when there is viable competition from other broadband providers." When weighing in on the net neutrality debate, Sood believes business concerns are the primary motivator. Broadband service providers are eager to please their digital customers, but for a price. "All this comes down to is the ability to charge for how much people consume," Sood remarked. With top cable trade groups representing service giants such as AT&T and Verizon expected to sue the FCC over net neutrality, aspiring YouTube stars, Netflix binge watchers, and avid gamers may need to accept that their virtual victory could be short-lived.

Our media-dependent culture is contributing to our inability to relate to those with different views. We choose our media sources and curate our online networks in ways that limit our exposure to opposing perspectives. And unfortunately, the information we are hearing often misrepresents the underlying data, to the frustration of Sood.

"Academia allows you to pursue perfection. That pursuit defines us and broadens both the people who produce it and the people who consume it. I want to produce papers that achieve the highest standards of academic research." These ideals stand in contrast with the reality of most news media. Sood believes that the most important message that Americans can draw from the rise of massive data is to pay attention to facts and statistics. Ultimately, "It is very easy to be vague and misrepresent things very convincingly. Pay attention to phrasing, and decide based on probabilities rather than possibilities. Anything is possible, but it is what is probable that really matters." With the MDI, hopefully we can get closer to achieving this standard.

REFERENCES

American Time Use Survey – 2013 Results. 18 June 2014. Bureau of Labor Statistics. http://www.bls.gov/news.release/atus.nr0.htm

Center for Data and Technology. 2015. *Coalition Letter to President Regarding Consumer Privacy Bill of Rights.* https://cdt.org/insight/coalition-letter-to-president-regarding-consumer-privacy-bill-of-rights/

The Connected World: Examining the Internet of Things: Hearing Before the Committee on Commerce, Science, and Transportation, United States Senate, 114th Congress (2015). http://www.commerce.senate.gov/public/index.cfm?p=Hearings&ContentRecord_id=d3e33bde-30fd-4899-b30d-906b47e-117ca&ContentType_id=14f995b9-dfa5-407a-9d35-56cc7152a7ed&Group_id=b06c39af-e033-4cba-9221-de668ca1978a

Hopkins, Daniel J. and Jonathan M. Ladd. "The Consequences of Broader Media Choice: Evidence from the Expansion of Fox News." *Quarterly Journal of Political Science* 9 (2014): 115-135.

Madden, Mary. 2014. "Public Perceptions of Privacy and Security in the Post-Snowden Era." http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/

Schneier, Bruce. 2010. "The Security Mirage." TED. http://www.ted.com/talks/bruce_schneier?awesm=on.ted.com_Schneier&utm_content=awesm-bookmarklet&utm_medium=on.ted.com-static&utm_source=direct-on.ted.com

Sood, Gaurav. 2015. *The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect.* http://www.gsood.com/research/papers/BroadbandPolarization.pdf

The US Digital Consumer Report. 10 February 2014. Nielson. http://www.nielsen.com/us/en/insights/reports/2014/the-us-digital-consumer-report.html

White House. 2012. "Consumer Data Privacy in a Networked World: a Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy." https://www.whitehouse.gov/sites/default/files/privacy-final.pdf

White House. 2014. "Big Data: Seizing Opportunities, Preserving Values." https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

White House. 2015. "Administration Discussion Draft: Consumer Privacy Bill of Rights Act of 2015." https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf

Wright, Joshua D. 2015. Prepared Statement of Commissioner Joshua D. Wright Federal Trade Commission on Wrecking the Internet to Save It? The FCC's Net Neutrality Rule Before the US capture as "the subversion of regulatory agencies by the firms they regulate" (2014 p. 49). And Carpenter and Moss define it thus: when "regulation, in law or application, is consistently or repeatedly directed away from the public interest and toward the interests of the regulated industry, by the intent and action of the industry itself" (2014 p. 13).

GPPR

Georgetown Public Policy Review

www.gppreview.com
info@gppreview.com
@GPPolicyReview

17 89

McCourt School of Public Policy
37th & O Streets NW
Old North, Suite 100
Washington, DC 20057