Data-Driven

Agitated nerves. Gnawing unease. Overwhelming vulnerability. Feelings that bear down on me with distressing force as I delve into my research to write this paper. Scathing and too many in numbers to even consider counting are the articles and posts outlining the amass of information that I have surrendered about myself by simply participating in the technological advances of our world. Following a few hours of falling down the search engine rabbit hole, I delete every accessible trace of myself from the digital world, exit my Internet browser, call mts to cancel my payment for their surveillance services, drown my laptop in the bathtub, and make myself an incredibly practical tin foil hat(imagine something resemble that of a shiny origami swan)... or at least it would have calmed my anxiety significantly if I had. -I more likely than not, logged onto my Instagram account to watch videos of cats.

The news is littered with articles outlining the breaches of confidentiality, trust, and ethical obligation by companies participating in our shiny, new, data-driven utopia. Facebook, Google, Apple, and other prominent technology companies are featured weekly for a variety of alleged unethical behaviors and/or breaches of their hoards of highly personal data. A quick Internet search offers fairly comprehensive articles from many news outlines in the vein of "I Downloaded the Information That Facebook Has on Me. Yikes," [1] and "How to download a copy of everything Facebook knows about you" [2], which most unsurprisingly express their findings in an uneasy tone. New York Times writer, Brian Chen, was in particular surprised to find that Facebook retained all 764 names and phone numbers on his iPhone's address book; an option that was recommended when downloading Facebook's Messenger app assist to 'connect you and your contacts'. In response to this revelation, Brian writes that he had hoped Messenger would "hold on to the relevant contact information only for people who were on Messenger. Yet Facebook kept the entire list, including the phone numbers for my car mechanic, my apartment door buzzer and a pizzeria" [1].

The methods in which this mass of data is collected include, but are not limited to, our willful disclosure of personal information for social media sites and credit card numbers for online purchases, the logging of current locations and device specifications when accessing companies services, and storing cookies -essentially identifying tags- on Internet browsers. Aside from the data we have consciously divulged to our Internet overlords, such as our names, birthdays, and favorite Harry Potter Movie characters, many websites -social media sites in particular- keep record of the date, time, and location of every log-in, every interactive behavior- such as watching a video,

commenting or liking a post, every photo uploaded- interactions with ads and their conversion to successful purchases, and the list goes on.

Online articles divulging the stalker-like behaviors of our most prominent technology giants such as Facebook, Google, and Apple are in depth, and plentiful. Seemingly appropriate for the astounding level of interaction they receive worldwide. There is however, one type entity in this goliath of observational data that does not appear to be under current public scrutiny; our Internet Service Providers, or ISPs, are essentially the gatekeepers to Internet access and the steep majority of these companies track every single click you make. Not only does your ISP have a comprehensive history of your online activity and corresponding location information expanding back to the day you began using their services, [3] but because of the explicit contract you have made for their services, they also know exactly who you are and where you live. Many ISP companies are even known for behaviors such as looking at the kind of traffic is coming from your machine(aka, web-browsing, video chatting, online gaming) and choking your Internet speeds if they see any BitTorrent traffic [4](whether the content being downloaded is an illegal copy of The Spice Girls new album, or a legal copy of "The Mueller Report: The Musical").

Scroll down on the main page of The Rubicon Project website and you will see a proudly displayed animation of a counter rapidly increasing from 0 to 1 000 000 000.... this number referring to the over 1 billion people around the world that they currently track. The Rubicon Project is an advertising company which creates a comprehensive profile of you by using cookies to track your activities around the web. This information is then used to process bids on ad-space on websites and apps that you view in order to provide advertising that is personalized to you; referred to as Interest Based Advertising, or IBA [5].

You are automatically opted in to IBA. It is easy to opt-out, it is a simple check box and 1 button, no guilt trips or hour of digging on their website to uncover the function, yet the way this works is by literally giving you another cookie. A cookie that tells the company not to process the data from the other cookies they have stored on your devices. How perfectly, ironic. The way in which The Rubicon Project functions is not a rarity in its field. Countless companies just like this, each hoarding unconceptualizable amounts of data, are now the backbone of our advertising industry. (And yes, if you're wondering, I did opt-out.)

The staggering majority of our personal data is collected and used as fodder by, or for sale to organizations involved with IBA. The widespread adoption of this method of advertising has infested our entire online world resulting in a significant increase in the amount of money companies shell out for online advertising and the IBA industry growth appears to be increasing dramatically, with a 21.7% between the years of 2016 to

2017 [6], yet, new research suggests that publishers make just 4% more using IBA vs the use of non-targeted ads [7].

I dug to the bottom of my backpack looking for the black felt-tipped marker I wished to use while drafting this section of my writing. I personally find something as simple as the width of the lines on the page, or the feeling of a particular ink on a particular style of paper can provide tactile feedback significant enough to influence my train of thought. I like to believe that I am a firmly rooted individual, and fairly resistant to influence, yet, the principles of behavioral psychology and the apparent influence of my term paper grade by a felt pen I did not find, seem to point otherwise.

Research published in 2016 in the Journal of Consumer Research found that not only do targeted ads encourage you buy a product they can also change the way you think about yourself. Across four studies, the researchers found that "behaviorally targeted ads lead consumers to make adjustments to their self-perceptions to match the implied label; these self-perceptions then impact behavior including purchase intentions for the advertised product and other behavior related to the implied label" [8].

It is commonly known that modern advertising practices are based on the principles of behavioral psychology and yet few of the practices required in the ethical collection, storage, analysis, and application of psychological data [9] appear to be upheld in this massive behavioral experiment currently being performed on our lives; as if they were not first and foremost a delicate analytical system based on scientific research and the application of said research.

In March of 2018 it became known to the public that Cambridge Analytica (CA), a political consulting firm had fraudulently harvested personal data from over 87 million Facebook users [10]. In 2014, a Facebook quiz claiming to be a research app used by psychologists titled "This Is Your Digital Life," was uploaded to Facebook under an academic licence. This app designed data scientist and psychologist, Dr. Aleksandr Kogan, had been commissioned by CA in response to their recent contract with the 2016 Trump Campaign for President of The United States [11].

If data scientist and whistleblower, Chris Wylie, had not come out publicly in 2018, divulging the goals and methods Cambridge Analytica, the company he worked for at the time of the data scandal, it is unlikely that this event would have ever came to public light. Facebook made no attempts to notify users of the breach of their privacy, yet they report that they had become aware of the fraud in 2015, removing Dr Kogan's app and demanding that all the harvested data be deleted [12]. Wylie says that he was only formally asked to delete the data in 2016, a year after Facebook found out about their misuse of the data. Facebook requiring no other confirmation of compliance than Wylie's signature on a form to satisfy that the data had been deleted [13].

In interview Wylie explains how CA's methods could be described as the application of psychographic techniques on large scale data to develop "cultural weapons," and that they had applied "some of the same techniques that the militaries use on ISIS, on the American Electorate" [13]. Cambridge Analytica claims that none of the illegally harvested data was used in the services they provided to the 2016 Trump Campaign, and yet their psychographically developed campaign still appears to have had the ability to apply targeted ads in a weaponized fashion with the aim of influencing the results of the 2016 United States Election.

Chris Wylie admits he was instrumental and at the heart of developing CA's systems; and even though the deployment of behavior altering techniques are not illegal he accepts his guilt in their development. "I was naive," "I made a big mistake," Wylie admits; and it feels somehow, as if he has been taken advantage of as well [14]. Computer scientists have no formal authoritative organization or leading ethical code, and globally there exists very little government legislation defining what practices are considered acceptable, and not. That is to say, yes, Wylie likely had the freedom to refuse the work (costing him his job), but I am positive the CA would have found a replacement happy to help them develop techniques for their less-than-ethical, albeit, perfectly legal, psychological weapons.

Previously, the website of CA's parent company, SCL, had featured a page listing the countries in which their services had played a role in influencing, Kenya, India, Italy, and Ukraine among them; their self-proclaimed track record was "more than 100 election campaigns in over 30 countries spanning 5 continents." Both Cambridge Analytica and their parent company, the behavioral research and strategic communications lab, Strategic Communication Laboratories(SCL), have shut down operations since the scandal broke surrounding their interference in the 2016 Presidential Race.

In Canada, I consider us lucky to live in a social and political climate that I would consider to be both fairly stable and tolerant. Canadian's privacy concerns do appear to be growing but increased regulation of the tech industry does not appear to be a serious government priority, at least at the moment. The EU currently seems to be progressing as the global leader in the regulation of data protection and privacy, enacting the General Data Protection Regulation(GDPR) in May 2018; with the aim of giving the control of collected personal data to all individual citizens. This regulation makes enforceable by law, among many other requirements, the use of appropriate technical and organisational measures to protect collected personal data, the default use of the highest possible privacy settings, and the clear disclosure by any processors of personal data as to the basis and purpose for the processing [15]. Unlike Canada's current privacy

policies, overseen by the Privacy Commissioner of Canada(which holds no legally binding authority), violators of the GDPR liable for fines topping out at 20 million pounds or up to 4% of their annual worldwide turnover of the preceding financial year, whichever is larger [16]. As of May 17, 2019, a total of 91 fines had been imposed due to the GDPR, totaling roughly $62, 815, 221 [17].

This may be an important step taken in the right direction for our global privacy yet the divide between our policy makers and our technological developers seems wider than ever. In a hearing to discuss Google's alleged "massive abuse of intimate personal data" I watched Texas Senator, Ted Poe, hold up his iPhone and ask the CEO of Google, Eric Sunder, "I have an iPhone. If I go over there does Google know if I have moved from here to over there? Yes, or no?" Sunder attempted to respond with the appropriate technical answer that he cannot say for sure as it is possible, depending on the device's current applications and settings. This resulted in Poe cutting the Google CEO off, accusing him of complicating what should be a simple yes or no answer, apparently completely unaware that his iPhone is in fact not a Google product.

Throughout my online journey, the number of articles I came across suggesting positive results from our massive data hoarding problem were considerably sparse. The large amount of sensationalization and oversimplification that the news industry is known for seems to have accomplished what it does best, hooked me in.

There now exist companies, such as, AncestryDNA and GEDmatch, who offer their services analyzing and storing what many would consider to be our most intimate data, our DNA. By the time it occurred to me to initiate a search focusing on possible DNA data hoards I was slightly less than enthusiastic, yet, the immediate results delivered surprising news. In April 2018, the suspected Golden State Killer was caught after 30 years at large due to a distant relative who uploaded DNA on GEDmatch [18].

California investigators were finally able arrest 72-year-old former cop, Joseph DeAngelo, guilty of committing 12 murders, more than 45 rapes, and upwards of 120 burglaries between 1974 and 1986. An article by Business Insider describes how investigators had used data, in compliance with the policies of GEDmatch, that the users themselves make public to connect with relatives in order to break the case [18]. This was a ground-breaking case; the first time this method had been used on a public DNA database to crack a case, and as of April 2019 has assisted in identification of suspects in over 59 cold cases of rape and murder [19]. The cracked cases include that of John Russell Whitt who has been arrested and charged with the 1989 murder of his 10-year-old son, Robert Adam Whitt, and his son's mother, Myoung Hwa Cho [20]; serial rapist, Darold Wayne Bowden, for the rape of 6 women between 2006 and 2008 [21]; and John D. Miller, for the 1988 rape and murder of 8-year-old April Tinsley [22].

Data from GEDmatch and similar companies who analyze DNA has also become a resource for pharmaceutical industry's genetic drug research [23]. No matter your personal opinion on the pharmaceutical industry, it's easy to agree that our DNA data

hoarding tendencies appear to be developing into a valuable resource pool for accomplishing tangible positive results.

In April 2019 the Utah State Department was successful in their use of GEDmatch's data in order to identify a 17-year-old boy who broke in to a church and strangled a 71-year-old woman, subjecting GEDmatch to scathing criticism. At the time, GEDmatch's terms of services stated that police enforcement was able to use their public data in order to "solve violent crimes of murder and rape." In this case, the victim was lucky enough to neither be raped by her attacker nor die from her injuries, therefore, this case did not meet the requirements to use GEDmatch's data. Curtis Rogers, the founder of GEDmatch, was contacted by the State Department's and ultimately swayed by their plea. Considering the severity of violence involved, Rogers agreed to make an exception and assist the police with the case [24].

GEDMatch has come under fire for this decision due to criticism from a small but very influential group of individuals. The lovely human being Judy G. Russell, author of the blog 'The Legal Genealogist,' which had previously praised GEDmatch, bluntly stated in an article in response to the situation, withdrawing her recommendation for the company, that "GEDmatch can no longer be trusted." Russell explains that "GEDmatch has broken its own word, its own contract with its users, set out in its own terms of service." Inflecting essential words in her article with **bold text**, she attacks GEDmatch's decision, conceding that most users would likely have agreed with the judgment, "**But they weren't asked**", "somebody **else** made that decision **for** them." [25] Russell's points are valid as they relate to GEDmatch's undeniable responsibility to uphold their contracts with their clients, and her tone offers no sympathy, only condemnation, for their role in convicting another dangerous, violent criminal.

In response to the backlash that was furiously incited by a small group of individuals GEDmatch has, responsibly, changed their policies. Users of GEDmatch are now automatically made to be unsearchable by law enforcement and may choose to opt-in to law enforcement searches, a reasonable policy. This change has resulted in the law enforcement's searchable database of over previously more than 1.2 million data points now currently sitting at 20,000 opted-in users. This change has left less the database at less than 2% of its previous size, and effectively grounded this powerful investigative method for the foreseeable future.
GEDmatch has also expanded their terms of service to cover all FBI requirements for violent crimes. The current policies would have covered the Utah Police Department's controversial case.

I will admit I watch too many super hero movies. And as our technology advances at an alarming rate, plotlines where 'super-technologies' are employed on a massive scale by supervillains appear more realistic every day.

The fact that ‘Google knows that I stayed up till 3am watching fail compilation videos when I was supposed to be writing my term paper’ is unlikely to result in any real repercussions, and I do not believe the majority of our technological goliaths pose a direct threat to the public, however, I do believe we should fear their ignorance. The denial, or disregard for what is now possible with our hoards personal data; the refusal to face the uncomfortably, unexpectantly intimate reality of the digital age we now live in has left us desperately vulnerable.

Works Cited

[1] C. X. Brian, "The New York Times," 11 April 2018. [Online]. Available: https://www.nytimes.com/2018/04/11/technology/personaltech/i-downloaded-the-information-that-facebook-has-on-me-yikes.html. [Accessed 26 June 2019].

[2] T. Haselton, "CNBC," 23 March 2018. [Online]. Available: https://www.cnbc.com/2018/03/23/how-to-download-a-copy-of-facebook-data-about-you.html. [Accessed 26 June 2019].

[3] PrivacyPolicies.com, "PrivacyPolicies.com," 2019. [Online]. Available: https://www.privacypolicies.com/blog/isp-tracking-you/. [Accessed 24 June 2019].

[4] W. Gordon, "lifehacker," 21 12 2011. [Online]. Available: https://lifehacker.com/what-does-my-internet-provider-see-when-im-downloading-5870042. [Accessed 20 6 2019].

[5] The Rubicon Project, "Rubicon Project," 2019. [Online]. Available: https://rubiconproject.com/. [Accessed 22 June 2019].

[6] IAB, "IAB," May 2018. [Online]. Available: https://www.iab.com/wp-content/uploads/2018/05/IAB-2017-Full-Year-Internet-Advertising-Revenue-Report.REV_.pdf. [Accessed 27 June 2019].

[7] N. Lomas, "TechCrunch," 1 May 2019. [Online]. Available: https://techcrunch.com/2019/05/31/targeted-ads-offer-little-extra-value-for-online-publishers-study-suggests/. [Accessed 25 June 2019].

[8] C. A. Summers, R. W. Smith and R. W. Reczek, "An Audience of One: Behaviorally Targeted Ads as Implied Social Labels, Journal of Consumer Research," June 2016. [Online]. Available: https://doi.org/10.1093/jcr/ucw012. [Accessed 26 June 2019].

[9] M. Krause, D. Corts, S. Smith and D. Dolderman, An Intoduction to Psychological Science, Toronto, Ontario: Pearson Canada Inc., 2018.

[10] I. Thomson, "The Register," 18 March 2018. [Online]. Available: https://www.theregister.co.uk/2018/03/18/facebook_confirms_cambridge_analytica_stole_its_data_its_a_plot_claims_former_director/?page=2. [Accessed 24 June 2019].

[11] M. Rosenberg, "The New York Times," 22 April 2018. [Online]. Available: https://www.nytimes.com/2018/04/22/business/media/cambridge-analytica-aleksandr-kogan.html. [Accessed 24 June 2019].

[12] K. Granville, "The New York Times," 19 March 2018. [Online]. Available: https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html?module=inline. [Accessed 26 June 2019].

[13] C. Cadwalladr, "The Guardian," 18 March 2018. [Online]. Available: https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump. [Accessed 26 June 2019].

[14] Channel 4 News, "YouTube," 17 March 2018. [Online]. Available: https://www.youtube.com/watch?time_continue=241&v=zb6-xz-geH4. [Accessed 27 June 2019].

[15] Wikipedia, "Wikipedia," 28 June 2019. [Online]. Available: https://en.wikipedia.org/wiki/General_Data_Protection_Regulation. [Accessed 28 June 2019].

[16] Wikipedia, "Wikipedia," 25 June 2019. [Online]. Available: https://en.wikipedia.org/wiki/General_Data_Protection_Regulation. [Accessed 26 June 2019].

[17] S. Fogg, "Termly," 17 May 2019. [Online]. Available: https://termly.io/resources/articles/google-gdpr-fine/. [Accessed 27 June 2019].

[18] H. Brueck, "Business Insider," 27 April 2018. [Online]. Available: https://www.businessinsider.com/golden-state-killer-caught-because-relatives-dna-online-2018-4. [Accessed 22 June 2019].

[19] Wikipedia, "Wikipedia," 26 June 2019. [Online]. Available: https://en.wikipedia.org/wiki/List_of_suspected_perpetrators_of_crimes_identified_with_GEDmatch. [Accessed 27 June 2019].

[20] V. Bridges and T. Grubb, "The News & Observer," 13 May 2019. [Online]. Available: https://www.newsobserver.com/news/local/article230338529.html. [Accessed 26 June 2019].

[21] N. Rojas, "Newsweek," 23 September 2018. [Online]. Available: https://www.newsweek.com/ramsey-street-rapist-dna-evidence-genealogical-data-darold-wayne-bowden-north-1088667. [Accessed 26 June 2019].

[22] K. Swenson, "The Washington Post," 16 July 2018. [Online]. Available: https://www.washingtonpost.com/news/morning-mix/wp/2018/07/16/i-been-watching-you-a-child-killer-taunted-little-girls-with-terrifying-notes-police-say-after-30-years-dna-led-to-an-arrest/?utm_term=.9f24966eff5b. [Accessed 26 June 2019].

[23] N. Martin, "forbes," 4 Dec 2018. [Online]. Available: https://www.forbes.com/sites/nicolemartin1/2018/12/05/how-dna-companies-like-ancestry-and-23andme-are-using-your-genetic-data/#6a2ada926189. [Accessed 28 June 2019].

[24] E. Levenson, "CNN," 9 June 2019. [Online]. Available: https://www.cnn.com/2019/05/27/us/genetic-genealogy-gedmatch-privacy/index.html?no-st=1561864860. [Accessed 25 June 2019].

[25] J. G. Russell, "The Legal Genealogist," 15 May 2019. [Online]. Available: https://www.legalgenealogist.com/2019/05/15/withdrawing-a-recommendation/. [Accessed 26 June 2019].

[26] R. Brandom, "The Verge," 25 May 2018. [Online]. Available: https://www.theverge.com/2018/5/25/17393766/facebook-google-gdpr-lawsuit-max-schrems-europe. [Accessed 26 June 2019].

[27] D. V. Boom, "cnet," 14 February 2018. [Online]. Available: https://www.cnet.com/news/huawei-zte-fbi-chris-wray-nsa/. [Accessed 27 June 2019].