Data patterns: an exploration of big data and (mis)information collection

Project Proposal

by Sarah Hontoy-Major Presented to prof. Sabine Rosenberg CART 451 AA Oct. 5th, 2022

Data shape our lives; whether or not we know it, it is a fact of modern human life. From the moment one enters the digital realm, one gives in a considerable portion of personal information for entities to use as they mostly see fit. As Mimi Onuoha (2016) explains in her essay 'the point of collection', data implies a relationship from object of collection to collector, from users accessing free or cheap content and products to the companies providing them with it while profiting from collecting their personal data. Many people are somewhat aware that free or cheap digital services, like social media or online music platforms, use and sell data to better their services and have a profitable business model. However, an hypothesis is that users and buyers are much less aware that retailers use similar technologies to track online behaviors and even predict trends (Shein 2022). Nonetheless, very few people truly realize the extent of the data these entities are fighting over so dearly. In this context, the purpose of this project is to get a comprehensive understanding of the data collected from online users through technologies and policies already put in place by current retailers, as well as creating an adequate vessel of communication to transpose the information in a more transparent and accessible fashion.

As a fully vertically integrated company, the online fashion retailer Shein makes an interesting case study for this project and consequently inspires its audience. By their allusive privacy policy, numerous tracking technologies, marketing algorithm and ERP system, the fast-fashion giant is able to power a relentless positive feedback loop of rapidly moving and short lived wearable products (Shein 2022, Norna 2020). The ever growing business model is only able to sustain itself by amassing equally growing datasets from its customers' online behavior and personal data to predict trends and configurate their production line in real time (Norna 2020). It is therefore young paying customers, mostly women, interacting with social media and fashion content, buying fast-fashion not out of necessity, but from interest, that is the core audience of this project. An hypothesis is that they are much less likely to know the extent to which their data is being tracked by such retailers, and to neglect its effect by lack of accessibility to digestible information about it.

In fact, Hito Steyerl (2016) goes as far as arguing that "data "farming" and "harvesting," "mining" and "extraction" are embraced" to created products that are mere "hallucinations [from] automated apophenia". This means that not only are entities collecting, extracting, and using personal information to recognize patterns and forecast buying behaviors and trends, but customers would also be finding the process more convenient overall. One of the goal of this research project is therefore to get a better understanding of users' willingness to share data if they knew about it through a transparent transaction. If online retailers' customers knew to what extent companies they were buying cheap products from were "farming" their personal data, would it discourage them from buying from them? Would they disregard it for the convenience of it? These are questions this project aims to answer.

Through an online platform, the goal is to grab the attention of users and let them interact with clearer information about privacy, tracking and other such policies of online entities gathering their data and using them in various ways. However, the goal is not only informative, as the platform would conveniently be exploiting the same or similar technologies, though through much more transparent means. The users would also have option(s) as to whether or not to share more personal information in exchange with some kind of gift.

This method of data collection aims to get a more comprehensive understanding of two things; firstly, how willing users are to give their data in exchange for services and products in a transparent transactions, and secondly, what kind of data companies and online entities have access to from users. This phase of the project, and more specifically the data that is found using these technologies, will infer a dataset used to feed a model chosen according to the context and usability of said found data. The following steps can only be realistically planned and foreseen once data is collected.

Many artists and researchers have recently been interested in shedding light on different data collecting and other computer assisted processes, as well as how they impact the society we live in. Mimi Onuoha, previously mentioned, and her work with her Pathways project (2015) exemplify many methods of this current project, namely an accessible, readable and straightforward data visualization user interface which still builds a narrative that viewers can relate to. Indeed, by tracking 4 weeks' worth of mobile data from 4 different groups of people, Onuoha was able to recontextualize otherwise dry data into powerful and intimate testaments of people's lives, down to the birth of a family's first child and its consequential change of dynamics through the entire family tree. Pathways is able to not only create more transparency as to the data that is being tracked by our cellphones unknowingly, but it also does it through a friendly, captivating and welcoming environment so viewers are less likely to want to detach themselves from what they are learning or browsing through.

Aaron Koblin (2006) also touched upon similar topics when creating his 'The Sheep Market' experiment. By paying more than 10,000 people 0.02$ USD to draw a "sheep facing left" through Amazon's Mechanical Turk solution, his goal was to show viewers what 'you get what you pay for' really comes down to in our current crowd-sourced society. Rather than an explicit vessel for his opinions, his project simplifies complex steps required to understand the issue at hand and lets viewers make up their mind as to what that means for them. What is interesting here is how Koblin, just like Onuoha, creates an accesible output from complex technologies digestible by a wider range of people who otherwise do not access this information.

Paolo Cirio (2006) has done a similar undertaking with his Obscurity project in a more mischievous way. By 'stealing' millions of mugshots, blurring them, shuffling the attached data between pictures, and posting the results on his well-indexed website, he was able to create a lot of noise or *trash data*. By using unconventional affordances of data, he was able to blur mugshot websites from which he had 'stolen' the data from. Such websites have extensive databases of biometric data of people charged and convicted with crime which they advertise and profit from through 'removal fees' (Cirio 2006). In this instance, Cirio is able, through humorous *trolling* methods, to shed light on problematic use of datasets while maintaining a straightforward, clear and concise access to information for people who otherwise would not access these complex notions.

All these artists have been able to transpose in much more transparent systems the notions of data collection they wanted to shed light on, however none have come close to examining how these systems of collection impact the objects we consume, buy, and wear on a daily basis. By translating these themes into an accessible platform for online retailers' customers to find digestible and accessible information about their own collected online data, the current project also aims to understand what exactly is being collected from these users and speculate on ways it could, or could not, be used by the entities having access to the same information.

References

Cirio, Paolo. *Obscurity.* 2016. https://paolocirio.net/work/obscurity/

Norna. *How Shein revolutionised retail for the 21st century*. 2020. https://norna.ai/news/?p=378

Onuoha, Mimi. *Pathways*. 2015. https://mimionuoha.com/pathways

Onuoha, Mimi. *The Point of Collection*. Data & Society: Points, February 10, 2016. https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa#.y0xtfxi2p

Shein. *Shein Privacy Police.* Shein.com. June 1st, 2022. https://ca.shein.com/Privacy-Security-Policy-a-282.html

Steyerl, Hito. *A Sea of Data: Apophenia and Pattern (Mis-)Recognition*. e-flux journal, April 2016. https://www.e-flux.com/journal/72/60480/a-sea-of-data-apophenia-and-pattern-mis-recognition/

Reading list and other resources

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press, 2020.

Han, Byung-Chul, and Daniel Steuer. *Non-Things: Upheaval in the Life-world*. English edition. Cambridge, UK: Polity, 2022.

On Paul Slovik's *psychic numbing* phenomenon and how it operates through screen-based activities;