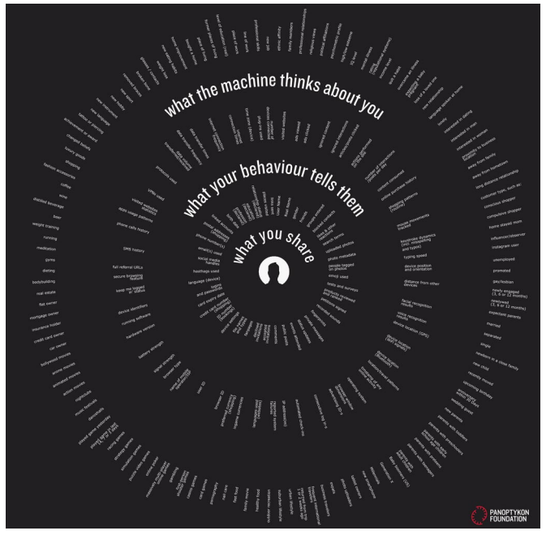
The near future of digital profiling

The goal of this essay is to explore three main issues from the users' side (lack of transparency in data collecting and processing, lack of control, and lack of feedback loop for confirmation bias) inside digital profiling, followed by my solution to these issues and the applications. My stance toward this topic is neither against the powerful internet platforms nor to introduce some technology that will block data collection—it is to imagine possible futures where digital profiling could become human-centered instead of advertising-driven.

ESSAY



15: Three layers of your digital profile
- panoptykon.org

We are currently facing potent social and economic shifts.[1] Covid-19 is not necessarily the protagonist of these shifts but rather a catalyst. It is accelerating the process of transforming offline life towards heightened cyber connection. The most significant change right now is working from home and the homebody economy [2] has suddenly become the near-real future. In addition, there are lots of other shifts happening online that average users [3] are unable to perceive and resonate with since they are a hidden, abstract, discrete, and changing phenomenon.[4]

At the front-end of the internet, cheaper, smaller, and faster chips [5] cybernate devices, bodies, and places to construct the 'Internet of Everything'.[6] Meanwhile, increasing features like cookies and the 'Like' button are designed to help track data.[7] At the back-end of the internet, digital profiling is intensively developed. Subjects related to digital profiling, such as data collection, data processing, algorithms, and applications, are well developed for activities such as advertising.

It's more important to understand the relationship between the individuals and the system and how to leverage the balance. Opting out of the system is one utopic solution,[8] although it's not practical. Unveiling the supply and demand of personal data:[9] More shared data piped into algorithms improves performance, better algorithm performance leads to higher advertising revenue and better-customized services, and higher revenue increases investment in algorithm development. Shoshana Zuboff used the word 'reciprocity' 16 times in her article "Big Other" to describe multiple levels of relationship between individuals and firms. The current situation is one of asymmetric reciprocity, as described in Varian's claim:[10] exchanging private information for new information and communication tools, which are essential requirements for social participation. From an individual's perspective, sharing data will cause issues such as data privacy, secondary usage, and targeted pricing. There are direct and indirect benefits and costs attached to sharing personal data.

Looking back in history, there is a spectrum between two extreme conditions: Intact privacy on the left, and no privacy on the right. The 'current' condition slowly moves from left to right. Its pace of seesawing [11] from one to the other is getting faster. (Does it also follow Moore's law? [12]) If full data privacy becomes pseudo-proposition in the future and blockchain-based digital identity like SoLid [13] and interoperability [14] will not be realized in the foreseeable future, then what should we expect for the near future? Research into the current tendencies from individual users and companies says that data collection is inexorable, but will become more permeative in the future.

Digital profiling

The digital profile is essentially the string of 1s and 0s that represent you in the systems. The current visual representation is seen in the ad setting page. Google, Facebook, Instagram have different user interfaces. However, they all point to the reality that our digital profile is an ad interest list based on our online trace and activities.

It would be naive, but understandable, to think that we have control over this profile in the first place. We have some control over what we agree to share, however, that is just the tip of the iceberg. It's not hard to imagine that by merely building connections between each piece of data, you could infer behavioral patterns. Even more so when algorithms take over this task. The map "Three layers of your digital profile" [15] done by

1: The last global crisis didn't change the world. But this one could.

3: 74% of Facebook users say they did not know that this list of their traits and interests existed
— Facebook Algorithms and Personal Data

5: Smaller faster cheaper over the future of computerchips

7: How am I being tracked? Behavioural Advertising 101

9: The intricate tale of demand and supply of personal data

11: In these extractive activities it follows the Street View model: incursions into legally and socially undefended territory until resistance is encountered.
— Big other: surveillance capitalism and the prospects of an information civilization

13: Interoperability is a characteristic of a product or system, whose interfaces are completely understood, to work with other products or systems, at present or in the future, in either implementation or access, without any restrictions
— Wikipedia
— When and how ict interoperability drives innovation

2: The rise of the 'homebody economy' means you can have all your food, alcohol, clothing and entertainment brought to your door.
— The 'homebody economy' gains steam in China amid Covid-19 pandemic

4: A Probe into the Symbol Form of Symbolized Design

6: Embracing the internet of everything to capture your share of $14.4 Trillion
— Cisco White Paper 2013

8: My experiment opting out of big data made me look like a criminal

10: Beyong Big data - Hal R. Varian P.5

12: Moore's law is the observation that the number of transistors in a dense integrated circuit doubles about every two years
— wikipedia

14: The ultimate goal of Solid is to allow users to have full control of their own data, including access control and storage location.
— Tim Berners-Lee

Panoptykon is an onion-like map which reveals the hidden layers behind the digital profiling. The first layer is the one you control or trigger. It includes your profile information, your posts, likes, search queries, and other types of personal interactions. In other words, it is your online trace. The second layer is one step further. It consists of your behavioral patterns like your typing speed, mouse movements, location pattern, voice recognition results, internet connection frequency, etc. These are not conscious choices you make, but rather the metadata which is embedded with the data you shared.[1] The third layer is composed of the interpretation of the first two layers. Your data will be preprocessed to remove noise and reduce complexity. After that, the data will be analyzed with multiple algorithms. There will be connections built between each piece of data to form specific patterns, and comparison with other users will be conducted to help evaluate the relevance and validity.

There are seven steps in the profiling process:[2] preliminary grounding, data collection, data preparation, data mining, interpretation, application, and institutional decision. The purpose of profiling is as follows:

> "it is not simply a matter of computerized pattern-recognition; it enables refined price-discrimination, targeted servicing, fraud detection, and extensive social sorting." [3]

From the perspective of tech companies, data collection helps them build profitable profiles that could help them link advertising to the targeted group of people. 'Internet' delivers people, "you are the end product." [4]

In the near future, I do not expect any fundamental change in digital profiling. Preliminary grounding identifies the goals of the analysis. It will remain profit-oriented. There are already changes happening outside the digital profiling occurring in law.[5] According to data protection regulation, GDPR requires companies to bring more transparency into tracking and profiling.[6] Thus, before blockchain-based digital identity and interoperability are realized, we could expect increasing transparency in digital profiling and a sense of control over personal data. Other than transparency and access to the digital profile, is it also plausible to expect feedback, like psychological mind-mapping, that could benefit users to some degree.

Confirmation bias

Confirmation bias is a state of intellectual isolation, people tend to resonate with information that helps confirm and enhance their beliefs or hypotheses. But such a phenomenon had been studied long before this term came out in history, which was long before the internet and algorithms were even invented.

"Filter bubble",[7] written by Eli Pariser, is one book that critiques algorithms that are enhancing confirmation bias. I agree with this opinion because the diminishing incentive of the subject is complemented by the algorithms. One used to search for information to confirm one's belief proactively. However now, one needs to receive the information recommended by algorithms reactively. The book "We are data" [8] brought up the idea of "data derivatives" [9] to critique algorithms that extrapolate the future based on present and the past. The book "Why We're Polarized," written by Ezra Klein, provides a political perspective.[10] Media, Congress, candidates, journalists, and voters will form a system in which there's a feedback loop that accelerates the process of adopting more polarized strategies. Political polarization is just one of the associated effects and outcomes of the confirmation bias.

1: Most digital cameras record the GPS coordinates of a photo you take in the EXIF metadata.
— Image EXIF data

2,3: Profiling (information science)
— Wikipedia

4:
Commercial television delivers 20 million people a minute.
In commercial broadcasting the viewer pays for the privilege of having himself sold.
It is the consumer who is consumed.
You are the product of t.v.
You are delivered to the advertiser who is the customer.
He consumes you.
The viewer is not responsible for programming—
You are the end product.
— Richard Serra "Television Delivers People" (1973)

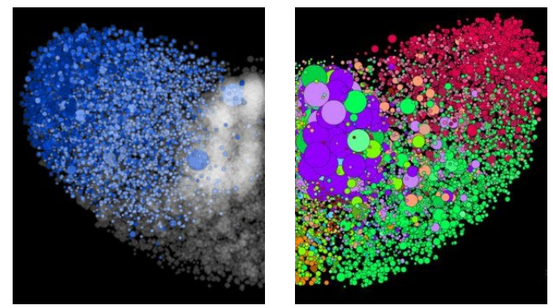
5: Transparent information, communication and modalities for the exercise of the rights of the data subject
— Art. 12 GDPR

6: EU Privacy Law Snares Its First Tech Giant: Google

8: We Are Data: Algorithms and the Making of Our Digital Selves
— John Cheney-Lippold

9: Data Derivatives: On the Emergence of a Security Risk Calculus for our Times
— Louise Amoore

10: Social media is one of those institutions, and in my view, is clearly a polarization accelerant. In the coming years it may prove a primary driver. But the bulk of the run-up in American party polarization predates social media, which means social media isn't core to the story.
— Interview with Ezra Klein



7: Left: Resist, anti-Trump
Right: The US political landscape
— This is what filter bubbles actually look like

Algorithms enhance the confirmation bias. It also causes bad user experiences. There are two different models being used amongst different platforms. One relies more on algorithm recommendation. The other one relies more on the user's following activities. Different platforms will adapt to or favor one of the models according to their strategy.[1] From the user's side, there have been complaints about content fatigue,[2] which is caused by an algorithm constantly pushing similar content. When Twitter applied the 'while you were away' algorithm, it pushed the 'best tweets' to the top of feeds, which caused complaints about restructuring timelines.[3] There is a button to opt-out from algorithms to have a regular timeline.[4] There is an option to avoid content fatigue, which is tapping the post/feed and selecting 'not interested'.[5] What is getting peculiar here is that if you want to prevent content fatigue, you have to give a click on 'not interested' to something overwhelming your feeds, which is your interest.

There is very little feedback conveyed from the platform to users, needless to say, feedback for confirmation bias. The feedback discussed here contains two layers: behavioral feedback and data analysis feedback. Eli Praiser mentioned in his Ted Talk [6] that structured physical spaces give people social feedback. But in the online environment, other than plain text with emoticons, there is a lack of body language and expression generated by other users, and it wasn't until recently that a dislike button was added on Facebook.[7] Shoshana Zuboff mentioned in his article "Big Other" [8] that tech companies

> "eliminate the need for, or possibility of, feedback loops between the firm and its populations."

The reason is to separate subjective meaning (revenue) from objective result (profiling). This is an asymmetric reciprocal relationship between users and firms which is reflected

> "in the fact that typical user has little or no knowledge of Google's business operations, the full range of personal data that they contribute to Google's servers, the retention of those data, or how those data are instrumentalized and monetized." [9]

Furthermore, concerns around using algorithm modifying behavior emerged after Facebook's emotional contagion experiment [10] was revealed to the public.

This essay is not trying to introduce a method to get rid of confirmation bias (which is incapable by ourselves [11]) but rather to propose the question of what we can do to nudge the situation. A good feedback loop might help change your behavior. Does the internet user need some feedback about their confirmation bias from the platform? How can information be designed to offer feedback for users which could intrigue them to conduct proactive searching? Are some of Spotify's algorithms[12] excellent examples of paradigm shifts since it allows users to steer slightly out of their comfort zone and expand their music listening?

1: What are the differences between instagram, facebook and twitter's algorithms?

2: Why YouTubers are feeling the burn

3: Put Down the Pitchfork - A Twitter Algorithm Won't Ruin Anything

4: About your Twitter timeline

5: What can I do if I see a post I don't like in Instagram Search & Explore?

6: What obligation do social media platforms have to the greater good?

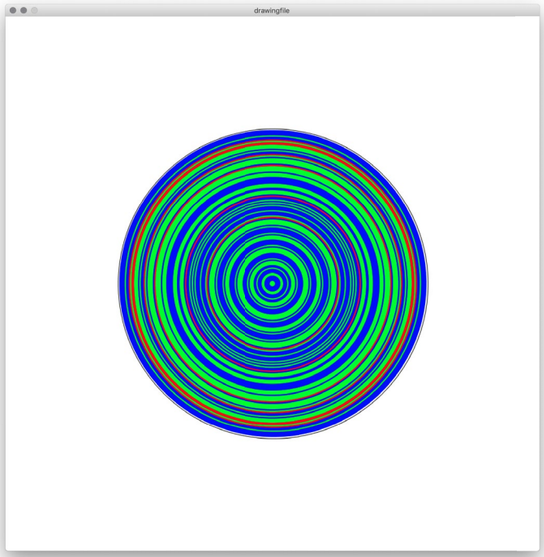
7: Facebook is finally rolling out a dislike button - sort of

8,9: Big other: surveillance capitalism and the prospects of an information civilization — Shoshana Zuboff

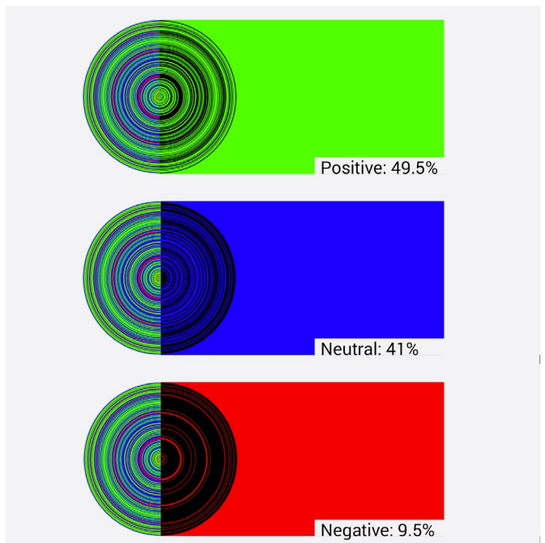
10: Experimental evidence of massive-scale emotional contagion through social networks

11: Myside bias in thinking about abortion

12: Tips for Getting the Most Out of Spotify



5: Platform: Twitter
Account: GigiHadid
Count: 116
Time: 2018 summer
- Mimesis 1.0



10: Data Distribution
- Mimesis 1.0

What I built?

In a nutshell, lack of control from the user's side, lack of transparency in data collecting and processing, and lack of feedback loop for confirmation bias are the three main issues in digital profiling. As mentioned before: the idea of stopping data collection and changing the goal of digital profiling may not be easy to realize because digital profiling serves as monetizing traffic based on algorithms. The degree of transparency needs more laws to level up. Although there are options for people who realize their confirmation bias to opt-out, the political correctness of the existence of those options is more important than their practicality.

Mimesis [1] starts by modifying the preliminary grounding, which is the first step of digital profiling, into constructing a feedback loop for users. The fundamental logic is to reuse existing data to conduct digital profiling for the second time. It falls in line with the concept brought up in SoLid: "reusing existing data."[2] Since some platforms offer public and friendly API, [3] it is possible to reproduce others' 'digital profiling' as a third party.

In this project, Twitter is the central platform used. Twitter API has three aggregated streams of Tweets, which are home timeline, user timeline, and mention timeline.[4] The home timeline consists of retweets and the user's tweets, which represent the user's thoughts. Mimesis 1.0 [5] pulls tweets from the Gigi Hadid home timeline and then pipes tweets into a sentiment analysis model which will spit out 1,0,-1 as sentimental results meaning positive, neutral and negative. After that, it uses Processing to match these digits to green, red and blue and draws circles from inside to the outside. Similar to how a tree ring [6] stores data about climate and atmospheric conditions, Mimesis 1.0 stores personal sentiment data. It is a visual representation of the internet user. It is an avatar,[7] as well as a simulacrum.[8] Mimesis 1.0 meets the requirement of substituting the current visual of digital profiling. The result of sentiment analysis is a representation of how a users' emotions flow in their expression. But it has a relatively loose connection with confirmation bias.

Thomas Goetz, executive editor of WIRED magazine, mentioned that a feedback loop involves four distinct stages: [9] evidence, relevance, consequence, and operation. The data comes in first: A behavior has to be calculated, recorded, and processed. Second, the information must be conveyed to the user in a way that makes it emotionally resonant, not in the raw-data form it was collected in. But if we don't know what to make of it, even persuasive evidence is useless, so we need a third stage: consequence. The knowledge needs to illuminate one or more paths forward. Finally, a clear moment must come when the person can recalibrate a behavior, make a decision, and act. Then the action is evaluated, and the feedback loop can run again; each step triggers new habits that get us closer to our objectives.

All data collected from the users after being processed is conveyed back to the user. Mimesis 1.0 did manage to complete the first two stages: providing users with behavior evidence which they can resonate with. It did illuminate the past pattern and left the future paths, such as being more positive or negative to the users. However, it is unable to break the loop of confirmation bias because the message unveiled in the graph [10] is not related to diversity. Diversity of perspective matters in confirmation bias. If there is a visual representation of confirmation bias, it will be the "filter bubble." [11] The color inside the bubble has much fewer diversity. Thus, what Mimesis needs to illuminate is the information relevant to the diversity of perspectives.

1: Mimesis is a term used in literary criticism and philosophy that carries a wide range of meanings which include imitation, nonsensuous, similarity, receptivity, representation, mimicry, the act of expression, the act of resembling, and the presentation of the self.
— Wikipedia

2: Developers will be able to easily innovate by creating new apps or improving current apps, all while reusing existing data that was created by other apps
— SoLid

3: The rise of APIs

4: Get Tweet timelines

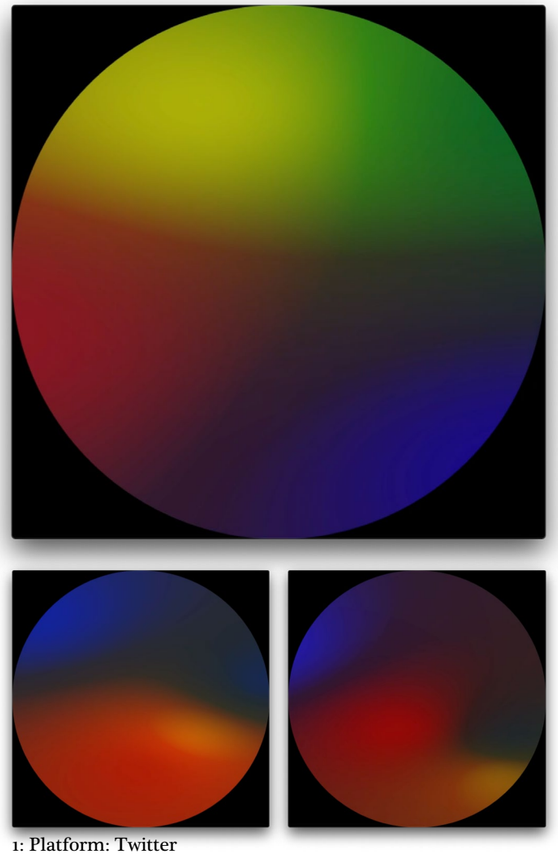
7: In computing, an avatar (also known as a userpic) is the graphical representation of the user or the user's alter ego or character. An icon or figure representing a particular person in a video game, Internet forum, etc.
— Wikipedia

8: 'It is the reflection of a basic reality. It masks and perverts a basic reality. It masks the absence of a basic reality. It bears no relation to any reality whatever: it is its own pure simulacrum.'
— Jean Baudrillard: Simulacra and Simulations

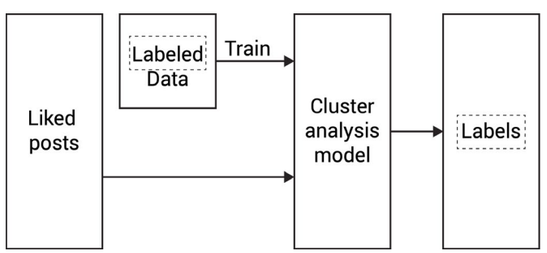
9: Harnessing the Power of Feedback Loops - Thomas Goetz

6: tree ring
— LintonArt Shop

11: Filter bubble
— Ted talk: Beware online 'filter bubbles'



1: Platform: Twitter
- Mimesis 2.0



7: Platform: Twitter
- Mimesis 2.0

Mimesis 2.0 [1] is the speculative result based on the same process as Mimesis 1.0, which is extracting data, conducting analysis, and drawing the result. The difference between 1.0 and 2.0 is not only replacing the sentiment model with the cluster analysis model [2] but also much higher complexity during the process. The MobileNet [3] model in RunwayML (a democratized machine learning tool for artists and designers) is the one that I have used and tested for 2.0. Since low fidelity models perform similar to the required cluster analysis model, which needs to be trained by enormous labeled datasets,[4] the list of which will refer to the popular synsets in imagenet (an image database organized according to the WordNet hierarchy). After data is piped into the model, it generates several clusters.[5] The names of these clusters are subsets of the labels in the data set, which will be mapped to colors.[6]

There are a lot of details in the process [7] unresolved in Mimesis 2.0. What is the definition of diversity? What is the benchmark of adequate diversities? For example, one user is mainly interested in animals, but the information about animals in his feed has depth and breadth. The other user has a wide range of interests, but the information about each topic lacks depth. Which scenario counts as having more diversities? At the same time, how does Mimesis 2.0 transform when the case changes from scenario 1 to scenario 2, and vise versa? Moreover, what is the process of color mapping? Is the result of mapping based on some degree of universal agreement? Can users decide how colors are mapped? Although there are two variables in color, saturation and variety, that could be mapped to depth and breadth, what is the criteria for mapping? Have all the questions above fallen in the realm of how to design a compatible label system for the stream of diversive information?

The ideal Mimesis should have a wide range of data input that could represent the user's epistemology and ideology. Input data should be analysed by a customized cluster analysis model which will be trained by a specifically designed training data set. In other words, this model should generate neither clusters that are too general, nor clusters that are too scattered. The system should find the balance between the amount of colors being used and an appropriate level of complexity which users can easily comprehend. The complexity of ideal Mimesis outweighs my imagination. But Mimesis 1.0 tests and proves the logic and process, and Mimesis 2.0 is a low-resolution speculation for the ultimate program. Thus, I believe it is manageable. Plus, tech companies already have 'Mimesis2.0' which is the ad interest list. It just needs some adjustment and one more step—drawing.

Goals for Mimesis
This is an information design project.
This is an avatar design project.
This is a UX design project.
This is a generative design project.
This is also a speculative design project.

There are multiple ways to define this project. Different definitions are pursuing different goals. I believe there are two major goals (based on Emotional Design [8]) in a design project: changing perception, and changing behavior.

Perception and behavior are tightly interconnected [9].Tim Brown from IDEO suggested designing simple digital tools to provide feedback as an additional tip to nudge people into new behaviors [10]. Thus, likely, the best way to change perception is to let users use the Mimesis rather than an informative one-time report. There are multiple ways to convert information design outcomes into tools

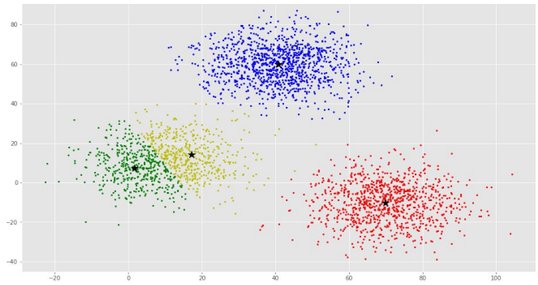
2: Cluster analysis is a class of techniques that are used to classify objects or cases into relative groups called clusters.
— Cluster analysis

3: Everything you need to know about MobileNetV3

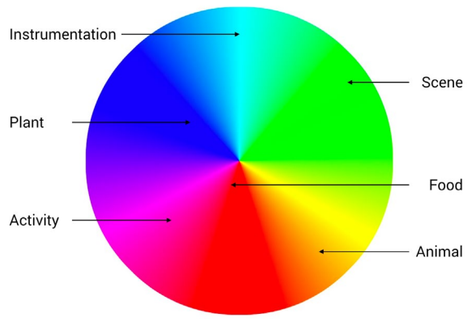
8: The three levels are visceral, behavioral, and reflective.
— Emotion Design - Don Norman

9: Beyond the Perception-Behavior Link: The Ubiquitous Utility and Motivational Moderators of Nonconscious Mimicry

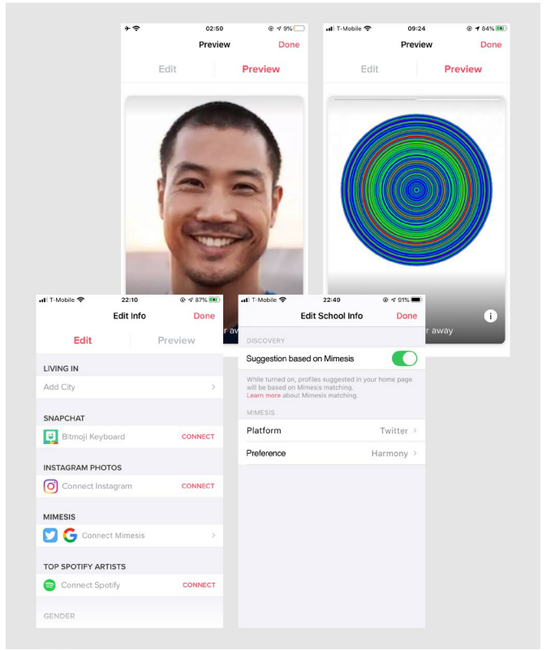
10: IDEO's Tim Brown on Using Design to Change Behavior



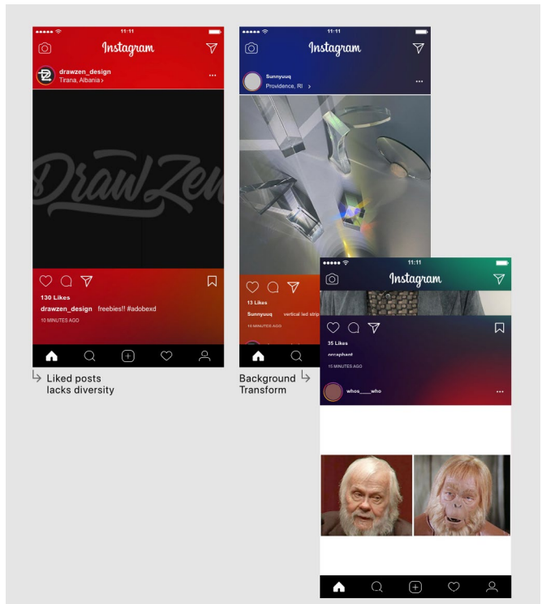
5: Example of clusters



6: Color mapping List resource-imagenet



3: Tinder
Mimesis 1.0



6: Instagram Home page
- Mimesis

for other activities. Pantone linked global environmental issues with the "Color of the Year 2019." [1] Designers, artists, and industries will keep reminding ordinary people of the issues when they keep applying the color in their work. Mimesis shares the same idea of generating customized results based on one's information with "Casa da Música logo generator." [2]

Mimesis 1.0, a customized graph, could be interpreted as an avatar. The applied platform is Tinder.[3] You can choose to connect Mimesis like Snapchat and Instagram photos. From a user's perspective, bridging Snapchat and Instagram with Tinder increases the plurality of information, though it might turn into a concern for some users. Mimesis shares the same idea offering Tinder users an option to show more about themselves. On top of that, Mimesis provides an alternative for matching algorithms.[4] In color practice, there are multiple choices to collocate different colors. What if matching algorithms could use Mimesis as a source to incorporate color combination theory [5] to expand users' options? Users' cards will be pushed to others based not only on age, distance, and gender preferences but also on their Mimesis matches.

For Mimesis 2.0 and the ideal Mimesis, the speculative application happens on platforms functioning as user's information resources. In this scenario, the demonstration [6] executes on Instagram's home page, which will be used as the background of the app. There are two different stages of the Mimesis. In the first stage, colors won't change but rather they'll slowly float. The second stage will be triggered by the posts that are liked by the user. The new liked post will be analyzed by the system, and the analysis outcome will decide to increase, decrease or change colors. Referring back to the four stages of the feedback loop: evidence, relevance, consequence, and operation, Mimesis embedded in the background functions as the relevant evidence of the user's liked posts. When users' likes are limited to a few categories of posts, the diversity of colors in Mimesis will diminish. This builds the consequence stage, which informs users of their 'bubble.' And finally, the 'bubble' nudges users to operate and completes the feedback loop.

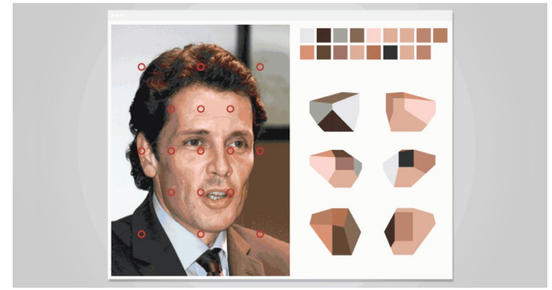
Finally, the speculative scenarios outlined here need to be a proof of concept that could generate revenues through further experimentation, research, and development of policy.

The idea put forward in this essay, that shifting the purpose of digital profiling as a third party, may draw worthwhile arguments about the practicality and policy issues. My point is to demonstrate my future visions or wishes for this technology and to propose a mutually beneficial strategy to tech companies.

4: Aside from your current location and gender, it's just your age, distance and gender preferences to start. Proximity is a key factor
- The method beind our matching

5: Complementary, harmony, calming, split, dynamic, vivid
-Color combinations

1: Living Coral
— Pantone color of the Year 2019



2: Casa da Música logo generator
— sagmeister