Big Data, Dopamine and Privacy by Design

Thomas W. Deutsch

Introduction: Privacy and Mom's Cupcakes

Human computation obviously requires the engagement of the people doing the computation, and people's willingness to participate has a lot to do with the value they derive from the engagement and the "cost" of participating. Most of us would associate the "cost" of participating being the time spent but there is a second, and often under appreciated expenditure, that of the person's privacy. How privacy is handled varies substantially across the spectrum of human computation, and the people doing the computation may not be in full control of the privacy decisions they are making. There are several reasons for this, but we are going to focus our attention on an underexplored considerations based on how we, as people, actually function and make decisions. To explore the dynamics here, like all good things in life, we are going to start by talking about cupcakes.

Imagine no one told you that Mom's homemade buttercream cupcakes were bad for you—you'd eat them to the point of exploding. OK, maybe that's just me but you get the idea. This is why we label our food's nutritional content since in theory an informed consumer is a healthy consumer, or at least one making good long-term choices. In practice we know that doesn't work so well—the obesity rates in the USA as confirmation of that. So why do people engage in the irrational behavior of eating both unhealthy food and unhealthy amounts of it? It turns out that we do that for the same reasons that many consumers struggle with the notions of privacy in an increasingly virtual world. That struggle has significant implications for the future of human computation and the problems that we are collectively trying to solve. To put a fine point on it—people's willingness to work on group problems and serve as part of a human sensor network will long-term depend on their ability to trust how their engagement and inputs to the project are handled.

T.W. Deutsch (✉)
IBM Information Management, San Jose, CA, USA
e-mail: tdeutsch@us.ibm.com

P. Michelucci (ed.), *Handbook of Human Computation*,
DOI 10.1007/978-1-4614-8806-4_69,

To help understand what is happening here and how it relates to big data, digital engagement and the need for privacy in human computation, let's turn back to Mom's buttercream cupcakes for a minute. Now, to be sure they taste good (especially the vanilla cake with strawberry frosting ones), but at some level we know we probably should not be eating too many of them, or eating them too frequently. So why do we do often over indulge in too large position sizes or too often make bad nutritional choices? Well in some cases we just can't help it because of how our brains work. As it turns out we are wired to be susceptible to responding to certain food types and components in a way that in an age of surplus becomes counterproductive.[1] Fat, sugar and salt all trigger physiological reactions quite separate from our purely rational experience of eating the cupcake. Simply stated, the same physiology and brain wiring that has allowed us to survive to this point is not especially well-equipped to handle our new circumstances of surplus.[2] It is starting to become evident that the dynamics that contribute to our poor food choices have parallels in our ability to self-regulate in our digital engagements.

Challenges in How We Make Decisions: Temporal Discounting and Neurobiology

To help explore these challenges, there are two important concepts that need to be introduced. The first is the idea of temporal discounting; the second is the idea that our neurobiology and neurochemistry are in play here without our being aware of it. Temporal discounting refers to our brains tendency to discount further away events from near term ones thereby making even smaller (if fleeting) rewards now appear more valuable than larger rewards in the future.[3] Here again the digital implications are harder to grasp as at some level nearly everyone understands that eating too much can cause unpleasant issues but our long term cost of trading off privacy are not as immediately apparent as, say, an upset stomach. As one of our Editors pointed out "even eating three cup-cakes has a near term impact on how you feel as well as poorly understood long-term implications. In the digital world, we almost never have immediate consequences for poor decisions—they are always long term "costs".

Strategies in the physical world for dealing with temporal discounting—such as walking around the neighborhood before going into McDonalds to give your brain time to better weigh the true "cost" of that chocolate milkshake you are craving—don't always work so well in the digital world. It can be a bit challenging to go walk around the neighborhood to 'cool off' before using a mapping service on a mobile device if you are turning to the mapping service since you don't know your way

[1] http://www.ncbi.nlm.nih.gov/pubmed/3135745

[2] http://www.ncbi.nlm.nih.gov/pubmed/3300488

[3] http://www.ncbi.nlm.nih.gov/pmc/articles/pmc1382186/

around the neighborhood. The immediacy and intimacy of our digital engagements, which clearly contributing to the usefulness and likability of the engagements, does pose challenges in self-regulation and decision making.

The Neurobiology and neurochemistry factors are play here are much more complex than the temporal discounting challenge outlined above. Neurobiology and neurochemistry deal with how the biology and chemistry of our brains impact our behaviors,[4, 5] To illustrate some of the considerations at play here, we are going to discuss our neurobiology. Before we go any further, I hasten to point out I am going to use some of that system functioning, such as the role of dopamine, as a stand in for a much more complex set of neurobiology and neurochemistry considerations. Mapping out all of that complexity, especially given that it is a fast-evolving area of scientific inquiry, is out of the scope of this article (as well as not being my field of expertise, so please take what follows with a grain of salt.).

As it turns out, how we make decisions is under much less of our conscious control than we realize. In some cases, we make decisions before we even become aware we are making a decision, or as Soon et al. summarizes "a network of high-level control areas can begin to shape an upcoming decision long before it enters awareness."[6] Our neurobiology works to shape decisions without our being consciously aware of it, and it does this very-very quickly. This happens through a complex set of interactions, but to single out as an example one component of this we are influenced by the amount of dopamine in our systems. Dopamine is an organic chemical that serves as a neurotransmitter that our bodies synthesis in response to simulation and it plays an important role in how we respond to situations. More specifically, "….midbrain dopamine systems are involved in processing reward information and learning approach behavior."[7]

Going back to the challenge of Mom's cupcakes can help illustrate some of the neurobiology at play here. Just the thought of eating the cupcakes can trigger neurochemical reactions that make us want to eat them that much more.[8] The actual act of eating one invokes neurobiological feedback loops that encourage us to eat yet more. Yet as challenging as the cupcakes are, in some ways the digital challenge is even harder to manage. First, unlike mom's cupcakes, there is no natural satiation mechanism whereby (after, say, five or six cupcakes, maybe less if you aren't me) you actually get full. Our saturation point of experience is far higher in digital engagements than cupcakes. Even more challenging is that Mom's cupcakes don't get more and more appealing as you eat more of them as they are a fixed experience. That is to say, the cupcakes don't change their behavior to be even more appealing and thus trigger another round of reinforcement. Highly intimate and

[4] http://www.merriam-webster.com/dictionary/neurobiology

[5] http://www.merriam-webster.com/dictionary/neurochemistry

[6] http://www.rifters.com/real/articles/NatureNeuroScience_Soon_et_al.pdf

[7] http://jn.physiology.org/content/80/1/1.abstract?ijkey=9149a8c097da470088e9a355b467daa4a58ebd5c&keytype2=tf_ipsecsha

[8] http://www.ncbi.nlm.nih.gov/pubmed/15987666

personalized applications, increasingly enabled by big data technologies, however, do exactly that. More engagement leads to even better ability to micro-segment your likes and preferences, and that will change your experience to become even more intimate and pleasing. Whereas Mom's cupcakes don't whisper in your ear that some chocolate milk would be especially tasty right now, the more engaging digital experiences can do the equivalent of that. More digital engagement leads to more insight about you, which in turn leads an even more engaging or socially invocative experience. The more engaging experience is, especially if it is social attachment in nature, the more to appears triggers neurochemical reinforcement[9] that, whatever privacy you have surrendered for that experience was worth it, potentially without the consumer every being aware they had made a decision to do just that before consciously doing it.

Social Interactions and Potential Impact on Privacy Choices

The idea of moderation is especially challenging in the digital realm. It is well understood that we are responsive to the dopamine reward we get from social interactions and it appears this is true when the social engagement happens to be a digital one rather than a physical one.[10] Digital interactions are even more intense when we process them as "intimate," since they are more engaging, and the more engaging the greater the involvement of dopamine.[11] So when we are faced with a choice of an impersonal or socially intimate experience, we choose an intimate one, at least partially since we get the reward of the chemical "hit". All of this may happen far more rapidly than we have time to understand the ramifications of our choices, since "dopamine concentrations are now known to fluctuate on a phasic timescale (subseconds to seconds)."[12] As noted above, the notion of choice and rational handling of the privacy issues comes into clear question when our neurochemistry is moving faster than consumers ability to make temporal tradeoffs (which, as noted above, is an iffy proposition anyway).

In the natural world that intimacy, however, can't scale beyond a relatively low number of connections at a given time due to time/space/personal network limitations. Technology can overcome those natural limitations so we're engaging at a volume and pace never possible before. This similar to overeating in a new age of surplus creates a volume of reinforcing neurobiological events we struggle to effectively manage. That, in turn, sets up a feedback loop wherein surrendering privacy increases the likelihood of a reward, which in turn rewards the surrender of privacy and so on. So while at some level, many people can intellectualize that surrendering

[9] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889690/

[10] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889690/

[11] http://www.dnalc.org/view/2385-The-Neurobiology-of-Love.html

[12] http://www.clinchem.org/content/49/10/1763.full

our privacy should require some contemplative thought on the balance of what we surrender versus what is gained and the long term implications of such a trade-off, it simply may be very difficult to overcome the brain chemistry that says, "this feels good, more please." Long term considerations of privacy may not stand much of a chance at that point compared to the neurobiology at work especially give the temporal discounting issues. Privacy is hard to value, and unlike food that comes with at least some basic nutritional information, there is no third-party reference point that people involved in human computing efforts can turn to for help in making a privacy related decision.

The Role of Commercial Models to Which We Have Become Accustomed

If that were not enough of a challenge, there is another one which has proven to be a quite well-established consumer behavior: customers like free stuff. Free email, mapping, social sites, free music, free hosting, free just about everything. And of course none of that is truly free. As was seminally expressed in the following blurb "If you are not paying for it, you're not the customer; you're the product being sold."[13] Many of the activities that trigger the neurobiology that proves challenging have the double "incentive" of being supported by monetizing the consumer as the product. Now of course there is nothing wrong with that, but when combined with the other triggers we've discussed makes a powerful experiences that, short of handing out free beer (or wine based on your personal preferences, which of course we'll likely know) is about perfectly designed to functionally disincentivize privacy considerations. It is worth noting here that there is a tendency to still frame the "do I or don't I surrender privacy" in purely rational terms when that may not be how we actually make the decision. Our emotional reactions to engagement are powerful, and often influence our perception that we are making purely rational choices when we are not.[14]

Just as the food industry has learned to develop foods engineered to take advantage of our neurobiology (think sweet and salty mix in ice cream so have many of the most utilized Internet sites. To be clear, they are responding to consumer preferences. Consumers want easier, consumers want more relevant, consumers want to be better entertained. That, of course, is the challenge. The benefits from using well designed (from an engagement point of view anyway) sites are immediate, the privacy trade-offs not immediately apparent and almost always involves stopping the behavior we enjoy to read Terms and Conditions of site usage that is not, shall we say, quite so engaging and thus not as rewarding. The neurobiology of this *is* different than in our physical lives. As a good friend of mine said "No one gets a dopamine hit from having the grocery store track their purchases through a loyalty card".

[13] http://www.metafilter.com/95152/Userdriven-discontent#3256046

[14] http://metablog.borntothink.com/wp-content/uploads/2011/07/1994-Damasio-Descartes-Error.pdf

While human computing is not confined to social sites and advertising funded sites/applications, one could argue that a majority of human computing comes as a byproduct of those activities. Problematically, it is those sites that often have the least transparent privacy policies and are designed to present an experience that invokes a neurobiological response that overrides a deliberate or methodical privacy-oriented decision making.

Examples such as http://www.patientslikeme.com/ where there is the potential for advancing understanding through shared information processing (in form of shared experiences) depend upon deeply personal information quite possible that is to be de- anonymized. The question of a person engaging in human computation can consider privacy and how their information will be shared when dealing in a social, experience related to their (or loved one's) health is debatable.

The Shortcomings of Anonymization

Anonymization has been presented as a way around this but as it as it turns out, anonymization is not very anonymizing in the age of big data. Anonymization—the basic tenant of decoupling the data from the common unique identifiers of phone number, user ID, email, or name doesn't hold up to a world enabled by big data technologies. Big data technologies offer both an expanded range of data gathering as well as increased processing power to dig into the data in more depth. One need not always dig that far however, as Ohm warned us about in 2009[15] and de Montjoye et al. recently reminded us of how "anonymized" data sets can still allow for very precise identifications of people.[16] This presents a substantial privacy challenge as our most common approach to building applications assumes that anonymity can be counted on to protect privacy, and many of the most commonly used application would simply seem to function properly if all the data that could be used for undoing anonymization were removed. It is also unclear if commercial entities could track down all the potentially de-anonymizing data in their anonymized data sets. As was recently observed:

> Removing forgotten information from all aggregated or derived forms may present a significant technical challenge. On the other hand, not removing such information from aggregated forms is risky, because it may be possible to infer the forgotten raw information by correlating different aggregated forms.[17]

It is also worthwhile to keep in mind here that the data being collected is often critical to the services being provided, as well as generating positive results from the human computing effort. As David Myers summarized in an especially well written observation: "Mobile operators' datasets help keep their networks running.

[15] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006

[16] http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html

[17] http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/the-right-to-be-forgotten

Location-based services don't work without location. We even hope big data capabilities will help us fight diseases and socio-economics problems. And, most importantly, despite the fact that most people in the U.S. and European Union insist they want better data privacy, we see time and again that this desire doesn't translate into action—people still give up their data without much consideration." Rock, meet hard place. Not surprisingly this debate has surfaced recently surrounding sugary drinks where our biological challenges to moderation and resulting personal and societal costs all require difficult trade offs.[18]

As with the great soda debate of 2013 in NYC[19], it seems pretty clear at this point there are not any easy answers here in the digital space. Users appear unlikely to spontaneously demand privacy baring some traumatic mass event, and the commercial models based on data collection have become firmly and widely embedded. To reference Myers again "we are not going to stop all this data collection, so we need to develop workable guidelines for protecting people."[20] It is unlikely that we are going to quickly evolve to a point where our neurobiology is not an issue to be considered in our online engagements, yet doing nothing does not appear to be an option. Voluntary solutions like Do Not Track,[21] which is both a technology and policy approach to giving users more control over their privacy, remain works in progress with uneven implementations.[22] Do Not Track has spawned related ideas on dealing with the issues outlined above, including the notion of Privacy By Design.

Privacy by Design Principals

Privacy By Design, an initiative by the Information and Privacy Commissioner of Ontario, Canada,[23] lays out seven key tenants that are designed to introduce some privacy protection by default. Without getting into the role of free will in all of this, Privacy By Design attempts to help protect us from ourselves by codifying an approach to the systems we interact with. The key tenants of Privacy By Design are[24];

1. ***Proactive*** not Reactive; ***Preventative*** not Remedial
   The *Privacy by Design* (PbD) approach is characterized by proactive rather than reactive measures. It anticipates and prevents privacy invasive events *before* they happen. PbD does not wait for privacy risks to materialize, nor does it offer remedies for resolving privacy infractions once they have occurred—it aims to *prevent* them from occurring. In short, *Privacy by Design* comes before-the-fact, not after.

[18] http://www.nytimes.com/2012/06/06/opinion/evolutions-sweet-tooth.html?_r=1&

[19] http://www.nytimes.com/2012/09/14/nyregion/health-board-approves-bloombergs-soda-ban.html?_r=0

[20] http://gigaom.com/2013/03/25/why-the-collision-of-big-data-and-privacy-will-require-a-new-realpolitik/

[21] https://www.eff.org/issues/do-not-track

[22] https://www.eff.org/issues/do-not-track

[23] http://www.privacybydesign.ca/index.php/about-pbd/

[24] http://privacybydesign.ca/about/principles

2. Privacy as the ***Default Setting***
   We can all be certain of one thing—the default rules! *Privacy by Design* seeks to deliver the maximum degree of privacy by ensuring that personal data are automatically protected in any given IT system or business practice. If an individual does nothing, their privacy still remains intact. No action is required on the part of the individual to protect their privacy—it is built into the system, *by default*.
3. Privacy ***Embedded*** into Design
   Privacy by Design is embedded into the design and architecture of IT systems and business practices. It is not bolted on as an add-on, after the fact. The result is that privacy becomes an essential component of the core functionality being delivered. Privacy is integral to the system, without diminishing functionality.
4. Full Functionality—***Positive-Sum***, not Zero-Sum
   Privacy by Design seeks to accommodate all legitimate interests and objectives in a positive-sum "win-win" manner, not through a dated, zero-sum approach, where unnecessary trade-offs are made. *Privacy by Design* avoids the pretense of false dichotomies, such as privacy vs. security, demonstrating that it is possible to have both.
5. **End-to-End Security—***Full Lifecycle Protection*
   Privacy by Design, having been embedded into the system prior to the first element of information being collected, extends securely throughout the entire lifecycle of the data involved—strong security measures are essential to privacy, from start to finish. This ensures that all data are securely retained, and then securely destroyed at the end of the process, in a timely fashion. Thus, *Privacy by Design* ensures cradle to grave, secure lifecycle management of information, end-to-end.
6. ***Visibility*** and ***Transparency***—Keep it ***Open***
   Privacy by Design seeks to assure all stakeholders that whatever the business practice or technology involved, it is in fact, operating according to the stated promises and objectives, subject to independent verification. Its component parts and operations remain visible and transparent, to users and providers alike. Remember, trust but verify.
7. ***Respect*** for User Privacy—Keep it ***User-Centric***
   Above all, *Privacy by Design* requires architects and operators to keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and empowering user-friendly options. Keep it user-centric.

None of this is easy or problem free to implement. As noted earlier many existing applications pre-date these notions and may not be able to function if tracking/tracing data were removed. It is unclear that separating the notions of privacy and security is commercially practical given how many applications have been designed. There is also the non-trivial issue of the potential need for a shift from the user's data being monetized to pay for the digital service if privacy is fully protected. The costs of not implementing, however, could be higher. If the potential of human

computing is blunted by concerns of privacy, who knows what we as a society we will forgo. It would seem that a reasonable next step is an honest conversation and full disclosure of how a human computing participant's information and activities will be utilized. In a free-market, people can vote with their time and there should be no shortage of human computing projects that both have worthy goals and manage to protect the participant's privacy. I hope this was a useful discussion, and I don't know about you but I'm craving a cupcake at this point.