Main Content

Background

v.0.0.1

1.1 Intro

We’re going to be talking about topics that cut across a wide range of disciplines. That means we need to have some familiarity with a bunch of different concepts. Here’s a very rough and very quick overview of some of them, just so we can all be on the same page.

1.2 Data

Basically everything we're going to discuss revolves around the use of data. So, what's data?

Very roughly, a piece of data (a datum) is a piece of information or a (purported) fact. More carefully, we'd probably want to say that data is information in the context of some use.[1] The temperature of the room you are in now is not data. It would be if we were adjusting the HVAC or running an experiment about how temperature affects reading ability.

1.2.1 Facts

Data represents facts. What are facts?

Facts represent a state-of-affairs which obtain. A state-of-affairs is the way part of the world is at a time. Take the sentence 'At 6 PM on 10 May 2019, Adam Swenson weighs 200 lbs'. This sentence is true because it says the world is a certain way and the world is actually that way. In other words, the sentence describes a possible state-of-affairs. Since that state-of-affairs obtains (it actually happens), the sentence describes a fact.

Consider three sentences:

> There is a taco on Mars.
>
> There is a taco in Adam's hand
>
> The dog needs attention

All three are states of affairs. The first two are not facts. There are no tacos on Mars; I am (sadly) presently taco-less. The third is a fact. He keeps nudging me, bringing over toys, and otherwise acting cute in wanton disregard of my need to write this. Be right back.

[1] Some writers invert this formula and define information as data plus use; we needn't wade into this fight here.

It is important not to confuse the fact with what actually makes it true. The thing that makes the fact that I weigh 200lbs true is that if you add up all the masses of all the atoms comprising my body at the specified time within the Earth's gravitational field, you will get 200 lbs.

1.2.2 Representations of facts

Can you own a fact? What could that even mean? Facts are abstract things. The fact that I weigh 200lbs can be expressed with different sentences. For example, 'if you make a big pile of all the things which weigh 200lbs and look through it, you will find Adam' (hopefully near the top). Thus no one can own a fact. Even if you owned me, you would not own the fact about my weight.

What you can own is a representation of a fact. Lets call that a datum. A datum can come in different forms (sentences, representations in a database). A datum (the singular form of 'data' which no one uses) is a (purported) fact.

If my weight is recorded in Hoolie's database and someone hacks into the database and erases the record, presumably Hoolie can sue her for destroying its property. I probably cannot. In some cases, the conditions under which the company acquires data gives the user rights to the data. The user agreement might say, for example, that if I cancel my account, Hoolie will delete all records related to me. In other cases, I may have no such right to the data Hoolie possesses.

1.3 Personal data

What's personal data? Data of a personal nature. All done. Moving on…

Okay, just kidding (although not by much). Here's a summary of what we are and aren't interested in from the Stanford Encyclopedia of Philosophy

Personal information or data is information or data that is linked or can be linked to individual persons. Examples include date of birth, sexual preference, whereabouts, religion, but also the IP address of your computer or metadata pertaining to these kinds of information. Personal data can be contrasted with data that is considered sensitive, valuable or important for other reasons, such as secret recipes, financial data, or military intelligence. Data that is used to secure other information, such as passwords, are not considered here. Although such security measures may contribute to privacy, their protection is only instrumental to the protection of other information, and the quality of such security measures is therefore out of the scope of our considerations here. [2]

Thus personal data is a proper subset[3] of data. Namely, it is data about a natural person.More importantly for what we'll be talking about, an often-used definition is the one found in the European Union's Data Protection Directive, namely

"Any information relating to an identified or identifiable natural person"[4]

Notice that this hinges on whether a piece of information or data can be explicitly related back to a person. [5]

1.3.1 Natural persons

Natural persons are contrasted with unnatural persons, for example, corporations; also, non-persons like rocks.

There may be borderline cases. If a robot turns out to be sufficiently like a human being that similar moral considerations should be extended to them, then perhaps the robot could become a natural person. For now, natural persons are limited to human beings.

1.3.2 Linkability

What it is to be 'about' a natural person? Is it enough that it represents a fact where a person is the subject. Or does the person have to be identifiable?

[2] https://plato.stanford.edu/entries/it-privacy/

[3] A fancy way of saying all personal data is data, but not all data is personal data

[4] EU Data Protection Directive (95/46/EC) Article 2(a) [ToDo: check formatting]

[5] See Sax 30

The ability to link a piece of data to a natural person is the decisive feature. There are 2 ways of making a link. Writers distinguish between referential and attributive uses.

Referential uses are made on basis of (possible) acquaintance relationship of the speaker with the object of their knowledge. For example, if someone says “the murderer of Tupac must be insane” while pointing at him in a courtroom, they are referring directly to a particular person. This is usually the sort of connection between a person and a piece of data which the law is concerned with.

Attributive uses, say something about a person without implying that we know anything about who they are. If I say “The murder of Tupac must be insane, whoever he is”, there is no implication that I’m actually talking about someone I can pick out. If personal data is understood this way, most data will not be protected under current regulations.

1.4 Algorithms

If you ask a computer scientist or mathematician what an algorithm is, you’ll get something like

An ordered set of unambiguous steps that produces a result and terminates in a finite time.

That’s more formal than we need. We’ll just say that an algorithm is a stepwise computational procedure for doing something.

Outside of technical contexts, indeed, often in the articles we’ll read, some writers use the term in narrower or more loaded ways. For example, some talk about algorithms as computational processes used to make decisions. Decision-making is a subset of the things we might do with algorithms.[6]

[6] For a discussion of this see Ref: *ALGORITHMIC HARMS BEYOND FACEBOOK AND GOOGLE: EMERGENT CHALLENGES OF COMPUTATIONAL AGENCY*. (2015). *ALGORITHMIC HARMS BEYOND FACEBOOK AND*

1.4.1 Adam’s attendance algorithm

For a simple example, here’s the attendance algorithm I follow at the beginning of every class:

- For each student, call out name. If answer, mark present

- When done, ask if anyone was missed

- If anyone answered ‘yes’, mark each answering student as present

If we were doing this with a python function, it might look something like (lines starting with ‘#’ or enclosed between triple quotes “””this is a comment””” are comments for humans to understand what’s going on and not part of the program):

```
def take_attendance(list_of_students):
“””Use at beginning of class to record which students are present”””

    for student in list_of_students:
        # Do the following for each student in the list
        answer = call_student_name(student)

        if answer:
            mark_present(student)

        # Do nothing if no answer

    # Now we're done calling the initial list
    missed_students = ask_if_anyone_was_missed()

    if len(missed_students) > 0:
        for student in missed_students:
            mark_present(student)
```

GOOGLE: EMERGENT CHALLENGES OF COMPUTATIONAL AGENCY (pp. 1–16).

Note that `mark_present` and `call_student_name` are two other functions which do what they're named.

1.5 Databases

Data of the sort which concerns us is stored in databases. There are many different formats, from single spreadsheets to complex relational databases with hundreds of tables.

In a relational database, the information will be broken up across a series of tables with another set of tables connecting them together.

[ToDo: Flesh this out or find video/explanation to suggest]

This is important because abstracting out the parts of the data (we separate the thing, the property, etc) allows us to connect things in novel ways. For example, we can query the database to learn new stuff. This enables us to more easily do data-mining.

1.6 Data mining

[ToDo: brief overviews of how some data mining techniques work]

1.7 Main Questions

Companies have collected data on their customers forever. Many of the hot data-mining algorithms have roots in statistical techniques that have been around awhile. Why are we so worried about this stuff now?

There are probably a lot of factors. One major set of changes involves the drastically declining costs of storage and computation. It used to be that if you wanted to keep data on something, the benefit of that data needed to outweigh the costs. Now, the marginal cost of storing data and processing it is basically trivial. In many cases, there is very little financial reason not to capture all the data you can and store it for an unlimited amount of time. [7]

Another factor is the availability of extremely individualized data. The internet enabled this to some degree. But the rise of social media and the ubiquity of smart phones makes it possible to gather a detailed profile of every potential customer.

Indeed, the wealth of very granular data has given rise to companies who specialize in what used to be called Knowledge Discovery in Databases (KDD). These companies aim at

> discovering non-trivial new insights in existing datasets, insights that cannot simply be observed in datasets or follow automatically from datasets, but insights that have to be extracted or generated since they do not 'lie at the surface' [Sax 27]

Nowadays, we call this big data.

To paraphrase an expert on the B.I.G., big data, big problems.

1.7.1 (Q1) When may an ethical company profit from the use of personal data?

As Sax notes

> big data's entrepreneurial potential resides in the fact that advanced mining techniques can extract/generate unanticipated, non-trivial, new, and (commercially) interesting insights. [Sax 27]

If that's true, then as he writes

[7] Check out how cheap Amazon Web Services is: https://aws.amazon.com/pricing/

> Big data's entrepreneurial potential is equally dependent on the legitimacy of the appropriation of these newly extracted/generated insights by commercial parties [Sax 27]

This is roughly our first main question:

(Q1) When may an ethical company profit from the use of personal data?

This extends to insights derived through machine learning and other analytical techniques done on users data.

Q1 Answer 1: When they own it

Q1 Answer 2: When we let them

1.7.2 (Q2) Harms of informational privacy violation

(Q2) When is a misuse of personal data significant enough to warrant moral condemnation, regulation, criminalization, or other forms of coercion to prevent?

Q2 Harms of informational privacy violation

1.7.3 (Q3) Responsibility

(Q3) How should we assign responsibility when people are harmed by algorithmic uses of personal data?

Q3: Responsibility

Q1 Answer 1: When they own it

v.0.0.2

2.1 Cases

2.1.1 Golden camper

Golden camper: Indigo has an uncanny sensitivity to features of terrain which are strongly correlated with gold being present beneath the surface. She doesn't know this. But something about the nearby rocks, plant life, and other features always make her feel comfortable and happier. Thus when she's out backpacking, she frequently camps in these places. Goldmember Inc has been flying drones over the area and notices her camping in spots they've already identified as gold deposits. At first they freak out, thinking that she's beat them to the gold. But then they realize that she's not digging anything up. Eventually someone suggests that they try digging for gold in other places they've seen her camp. They find gold there too. They thus repurpose their surveying drone to discretely follow her on her trip. They carefully map where she camps, where she takes rest breaks, et cetera.

2.1.2 Wheelbarrow

Wheelbarrow: We land on an uncharted, uninhabited island. Nothing on the island belongs to anyone. Brown cuts down a tree and builds a wheelbarrow out of the wood. Brown sells the wheelbarrow to Green.

2.1.3 Hoolie

Hoolie: Hoolie is a big-data company. They amass data from its users and other sources. They do some very sophisticated math to find previously detected correlations in consumer behavior. From these correlations, they create detailed consumer profiles which it can sell to marketers/advertisers. One product, Taco-targeting aids owners of taco stands in finding people most likely to be influenced by ads and coupons at particular times. For example, customers who like red vines, whiskey, and own cats are much more likely to purchase tacos on

Thursdays. If you own a taco stand, these are the customers in your area that you want to be sure you reach on Thursdays.

2.2 Intro

When you go to the grocery store, the store tracks what you buy. The degree to which you are individually identifiable depends on several factors, including whether you paid by credit card and whether you signed up for their loyalty program. But what's important for now is that they collect the data about you at the point-of-sale through their normal business operations. At that point, they own the data.

If you own something, you have a right to use it in certain ways. If some of those ways are profitable, it seems that you have a right to keep the profits. Thus our first potential answer to Q1 is:

(Q1A1) A company may profit from personal data which they own.

To understand whether this is a good answer, we need to start by thinking about the concept of ownership. That is, we need to think about property.

2.3 Ownership and rights to profit

If you own a car, you may sell it and keep the profits. If you own an apartment and rent it out, you get to keep the profits because. How does ownership fit with the right to profit? Let's start with some fairly obvious observations.

2.3.1 Necessary condition

One possibility is that ownership is a necessary condition.

(O-N) S may profit from the sale or use of x only if S owns x

This is false. Ownership is not a necessary condition of legitimate profiting. Suppose you rent an apartment and your lease allows you to sublet it. If you rent out your apartment while you are on vacation and make more money than your rent, ceteris paribus, you have a right to keep the profit.

That said, the 'ceteris paribus' (all things being equal) clause is doing a lot of work in this example. If your lease prohibits subleasing or it is illegal to do so, your legal right to keep the profit is undermined. Still, since you would have the right to the profit in some cases, (O-N) cannot be true.

2.3.2 Sufficient condition

Perhaps ownership is instead a sufficient condition

(O-S) If S owns x, S may profit from the sale or use of x

That's getting better. Though, clearly, this will need to be supplemented with legal and moral qualifiers on the kinds of use. Gun owners have no right to profit by killing people. That said, it does give us basically what we need.

2.4 Property / ownership

But what exactly do we mean by 'ownership'? The answer cannot be separated from philosophical justifications for property.

2.4.1 Property as a bundle of rights

Bundle of rights

Whenever we talk about property, we're talking about a bundle of rights. If you own a car, you get to determine who touches it, who uses it and how. You can destroy it. You can sell it.

Obviously, these rights are not absolute. They exist only within a larger system of rights and obligations. The right to exclusive use of your car doesn't mean that you can use it in any way you please. You must drive it on the correct side of the road. You cannot destroy it with explosives in the middle of a parking lot. You can't sell it to a child.

Importantly for what's to come, property rights are transferable. You can sell, donate, or gift your car to someone else. Such property transfers are, generally speaking, complete. Once the car you sold me is mine, I now have all the rights you did —exclusive use, et cetera. You no longer have any rights to it.[8]

That brings us to the question of how you get property. There are 2 possibilities: you were either the first owner or not.

2.4.2 Transfers

If you weren't the first owner, you get the property through a legitimate transfer. Someone sells or gives it to you. As long as they got it by a legitimate transfer and so on back to the original owner, it's yours.[9] We can summarize:

(T) If S1 acquires x from S2 through a legitimate transfer and S2 either created x or acquired it through a legitimate transfer, then S1 owns x

[8] This is, of course, a broad generalization. It is possible to put riders and other provisions into a sale contract —you could retain a right to drive it once a month. This might be something to keep in mind for later

[9] Note that if at some point the property was stolen, none of the subsequent owners legitimately own it. It doesn't matter if the theft occurred yesterday or generations ago. Recognizing this opens the door to one line of argument for reparations to African Americans for slavery.

But what about the first owner? For that, we'll turn to two flavors of a Lockean account. You can either make it out of something or make it out of nothing.

2.4.3 Locke

On Locke's account, broadly speaking, you can create property by taking an unowned resource and mixing your labor with it. That somehow makes it yours and gives you the exclusive right to control it.

Thus suppose we land on an uninhabited island which no one owns. I walk over to an orange tree, reach up and grab and orange. Because I've mixed my labor with it, it's now mine. In Wheelbarrow, Brown owns the wheelbarrow because she mixed her labor with the wood. Thus she may sell it to Green.

2.4.3.1 Account of property

We can summarize Locke's account of creating property rights:

(L) If r is an unowned resource and x is the result of S mixing her labor with r, then S owns x [10]

Assuming that we have some sort of system of market exchange we also accept something like

(L2) If A owns x, ceteris paribus, A has a right to profit from the sale or use of x.

[10] For simplicity, this leaves out the famous proviso: that this is true as long as S does not exhaust the supply of x.

Again, we'll assume that prohibitions on harm to others, immoral uses, and other restrictions are built in to the 'ceteris paribus'.

2.4.3.2 Right of use

Why does it matter that the resource labored upon is 'unowned'? Suppose I lend you my wheelbarrow to use. After you are done using it, you clean it up and paint some cool flames on it before returning it to me. You have mixed your labor with the wheelbarrow. But since you don't own it, you do not thereby acquire any property right to it. You cannot demand payment for the artwork. Indeed, if I don't like the flames, I can demand you restore it to it's previous un-enflamed state.

At the same time, if I have granted you the right to use my wheelbarrow in your between-bar-transportation business, I do not automatically acquire a right to your profits. That is, it is not true that:

X (L3) If A owns x, A has a right to the profits from any use of x which A permits

Obviously, that could be part of the rental agreement. But the right to profit does not follow automatically from ownership. If it was automatic, the tool and equipment rental business would be awesome. You would have a right to the profits your customers make from whatever they use your equipment to build.

2.4.3.3 Application to Q1

There's a clear difference between Wheelbarrow and Hoolie. The wheelbarrow is built out of a resource that no-one owned. That seems to raise a crucial question: Who owns personal data about S from which valuable insight V is extracted?

On first glance, it depends. Suppose the supermarket doesn't keep any record which ties you to the purchase you made. They know some customer bought all that beer. But they have no way of knowing it was you. That data is unquestionably their property. It may seem that this is very different from inputting your weight into a fitness tracking app. That data is always about you.

But I don't think any difference, if there is one, matters. Virtually any company whose lawyers have a pulse will have a privacy policy, terms and conditions, or other binding policy by which you license the company to do as they please with the information. Just like the person who borrows the wheelbarrow to run their bar transportation business, the company has the sole right to the profit.

2.4.4 Kirzner

Kirzner's account offers something different. Applying his view, the ownership of the data is (morally) irrelevant. The company owns the profits from the valuable insights because the insights are created out of nothing.

2.4.4.1 Kirzner's view

To see how Kirzner's view works, consider

Pizza arbitrage: Scarlet notices that Green is selling pizza for $1 on the north side of campus and Blue is selling pizza for $5 on the south side of campus. Being a smart business student, Scarlet buys a bunch of pizza for $1 and sets up shop on the south side of campus, selling it for $4.

What gives Scarlet a right to the $3 profit?

Approaching this in the Lockean way, it's probably because Scarlet owns the money which she used to acquire the pizza. She bought the pizza which gaver her the right to sell it for whatever her customers were willing to pay. That's a bit uncomfortable for the Lockean. The main intuition behind Locke's view is that property is connected to labor and effort. The fact that you worked on something gives you the right to it.

But in Pizza arbitrage, Scarlet didn't do much. She walked from one end of campus to the other carrying stuff. Plenty of people do that for free. She labored at taking people's money and handing them slices. But can those movements justify ownership of the vast wealth she amasses?[11]

On Kirzner's approach, the labor is irrelevant. What Scarlet did was create value (viz., $3) by recognizing the market opportunity. So, just like the person on the island chopping down a tree to make a wheelbarrow, she created something and therefore has a right to profit from it. But unlike Locke who focuses on the labor part of this, Kirzner makes the property right depend on creation.

Kirzner calls his view 'Finders keepers'. I find this strange. Shouldn't it be 'Makers keepers'? Though since he is focused on entrepreneurs, it makes more sense in that the entrepreneur finds the value by finding the market opportunity.

''In order to introduce plausibility to the notion of finders–keepers, it appears necessary to adopt the view that, until a resource has been discovered, it has not, in the sense relevant to the rights of access and common use, existed at all'' (Kirzner 1978: 17).

Importantly, the person selling the pizza for $1 can't complain that she got ripped off. When she sold the pizza it was worth exactly $1. The arbitrageur created the extra $3 of value. As Sax says:

"the entrepreneur has created –ex nihilo– the new use for oranges and has therefore created the additional value of $3....the additional value...was not, in any relevant sense, present in the oranges before...intervention." [Sax 28]

and

"the discovery of a hitherto *unknown market use* for an already-owned resource or commodity constitutes the discovery of a hitherto *un-owned* element associated with that resource or commodity." [Sax, 28]

Thus Kirzner's basic claims are:

[11] Fine. Think about the real example of a stock trader.

(K1) If S discovers a novel marketable use u for R with value v, S creates v

and

(K2) If S creates x ex nihilo, S owns x

Which we can summarize:

(K) If S discovers a novel marketable use u for R with value v, S owns v

That doesn't seem quite right. Dreaming up the marketable use isn't enough. You have to actually exploit the niche. Suppose you and I have the same idea. I sit on my butt. You work your butt off bringing it to life. I have no right to your profits. That said, I don't want to spend more time fixing this picture up. We're going to blow it up shortly.

2.4.4.2 Application to Q1

Sax's strategy in the paper is to come up with the best case he can for big-data companies having a right to profit from the insights they generate and then turn around and attack that case.

He thus brings in Kirzner's account to explain the right to profit. From that perspective, the profitable insights about consumer behavior aren't sitting there in the dataset waiting to be discovered . They aren't like iron waiting in the ground to be dug up and sold. They are created. That's what gives the big data company the right to profit.

To make this concrete, in the Hoolie case, the marketable insight —the taco-Thursday proneness of cat-owning whiskey-drinking red vine lovers—doesn't exist until Hoolie runs its analytics on the dataset. That process finds something valuable in the dataset. Since Hoolie did the finding, they own the valuable insight. Therefore they may profit from it.

2.5 Problems for Q1 A1

Now that we have a couple of answers to Q1 in terms of ownership on the table, let's turn to some problems. Spoiler alert: I wouldn't get too attached to these answers.

2.5.1 Divisibility Problem(s)

Notice that we haven't said much about personal data in particular. Everything we've said would apply a company which uses analytics on weather data to make highly targeted predictions for farmers, airlines, event-planners, et cetera.

The finders-keepers picture turns on creation of value. There has always been liquid in oranges. The valuable beverage orange juice appeared when someone recognized a market opportunity for it. Similarly, according to Sax,

> As long as the big data entrepreneur gets a hold of the original (personal) data in a just way, the entrepreneur is free to apply entrepreneurial insights and appropriate the additional value that she creates. Indeed, justice even requires that the entrepreneur is the legitimate owner of these new insights that are extracted/generated from the original data by the entrepreneur. Just like the original holder of the oranges was never the owner of the property of the oranges that allowed the entrepreneur to make orange juice out of the oranges, so the data subjects, whose data are used, were never the owners of those valuable insights that lie hidden in the data and that the big data entrepreneurs manage to extract. The data subjects providing the data cannot, in providing the data, be explicitly aware of the specific valuable insights that are hidden in their data. To see why, remember that these insights are in fact new non-trivial data, created out of the original data. The very nature of big data analysis is such that the newly mined insights do not follow directly from the original data, meaning that the original data subjects cannot, by definition, be aware of what emergent data can be extracted/generated from their personal data prior to the actual extraction via data mining. Due to this lack of explicit knowledge of all the unpredictable new insights that can be extracted from their personal data, the original data

> subjects can, under the 'finders, keepers' ethic, not be seen as the legitimate owners of these newly mined insights. The big data companies are the finders-creators of these new insights and their appropriation of the fruits of these new insights is therefore legitimate when the 'finders, keepers' ethic is accepted. [Sax 29]

and

> As I have argued, the 'finders, keepers' ethic depends on the idea that within the same goods, some of the properties can be owned by the original holder, while other properties, namely those allowing for applications the original holder is not explicitly aware of, are unheld at the very same time and can thus, after discovery, be appropriated by the finder-creator. This introduces a certain kind of divisibility to goods which is necessary for finders-keepers to function adequately. {Sax:2016bq} 29

I confess I don't entirely understand what Sax means by 'a certain kind of divisibility'. But I think we can get roughly his worry going by noting that, since we are basing the right to profit in ownership, people better be separable from their data. The closer any theory claims to saying people can be owned, the more aggressively we should reject it.

2.5.1.1 Technical problems

The first set of concerns about divisibility involve the technical ability to anonymize data. This doesn't require grand metaphysical claims like the ones we'll get to in a minute. Just the idea that, if we could completely sever the connection between a person and her data (for some uses), there would be no special concerns about personal data.

Oftentimes, the company doesn't need to be able to identify the people the data represents. If I just want to sell advertisers on a way of identifying whom to advertise to, I just need to find correlations between traits of people. I don't care who those people are. If I can replace everyone's name with, say, a unique identifier and then throw away the names, it seems like there's no difference between the personal data and a bunch of weather data.

Unfortunately, it is very hard to anonymize datasets. This is an active research problem for computer scientists. It matters a lot for, say, medical researchers to have a bunch of publicly available patient data. But those patients better never, ever be identifiable.

The problem is that for a variety of theoretical and mathematical reasons, it doesn't take too many data points to identify an individual. As long as there are clever computer scientists around, it is really hard to completely anonymize a dataset.

2.5.1.2 Metaphysical problems

Of course, no one has proven that datasets cannot be irreversibly anonymized. This is in fact a major ongoing research project. If it turns out that someone invents a way that irreversibly anonymizes personal data, then the problems raised so far will no longer apply. Thus Sax draws on work by Floridi to give the problems of divisibility metaphysical teeth; computer science won't save you here.

Floridi's view is quite provocative. We can summarize it as claiming that you are literally your data.

> Looking at the nature of a person as being constituted by that person's information allows one to understand the right to informational privacy as a right to personal immunity from unknown, undesired or unintentional changes in one's own identity as an informational entity, either actively – collecting, storing, reproducing, manipulating etc. one's information amounts now to stages in cloning and breeding someone's personal identity – or passively – as breaching one's informational privacy may now consist in forcing someone to acquire unwanted data, thus altering her or his nature as an informational entity without consent.[12] [195]

[12] Floridi, L. (2005). The Ontological Interpretation of Informational Privacy. Ethics and Information Technology, 7(4), 185–200. http://doi.org/10.1007/s10676-006-0001-7

[ToDo: Do more to explain and make seem less crazy]

[ToDo: Add summary of the solution to personal identity problems he thinks justify this]

2.5.2 Conception of justice presupposed

Sax also claims that if the business models of big data companies are justified on Lockean or Kirznerean grounds, then they are vulnerable due to the nature of the conception of justice they presuppose.

Both accounts suggest that the only questions we can ask about the legitimacy of ownership are historical: Did you legitimately create the thing or obtain it through legitimate transfers? If the latter, were all the transfers back to the beginning legitimate? If the answer is 'yes', then there's nothing else to say. That's really significant. These accounts of justice rule out (or make very tricky) questions about the externalities of ownership (e.g., does your ownership negatively impact others) and distributional questions (e.g., how much wealth should any one person control).

One reason this may matter is that the privacy costs and benefits related to big data may need to be assessed cumulatively. It may be difficult to see the problems (and benefits) if we look only at individuals and the transfers between them. We need to look at what those mean overall. If one company has your information and uses it to target marketing to you, that may be relatively unproblematic. But if every area of your life is subject to different companies gathering information on you, that looks potentially more problematic. However, it seems, Sax claims, that these problems cannot be assessed on the Locke/Kirzner accounts of property.

The proper account of justice is a long-running debate in political philosophy. For our purposes, we'll just note that the account is built on foundations which are, at best, still under construction.

Q1A2: Consent to data use

v.0.0.2

3.1 Intro

Perhaps the problems raised for ownership of data so far can be easily sidestepped. Suppose you are a gardener. Your chainsaw breaks right when you start cleaning up a fallen tree for one of your customers. I lend you mine while yours is in the shop. You make money from using my chainsaw and return it to me without a share of the profits, but with a gift of beer out of gratitude. I let you use my property so you could make money. The fact that it is my chainsaw gives me no right to your revenues. (Obviously, that could've been part of the lending agreement, but it wasn't). So, perhaps the problems we've seen with data ownership can be avoided by referring to the fact that customers agree to allow companies to use their data.

3.2 Notice and Consent

The standard model of consent embodied in privacy policies, terms of service, user agreements, et cetera often gets called 'Notice and consent'. This just does what it says on the label: The company notifies you about what data they are going to collect and what they are going to do with it. You agree (or don't use the service).

More broadly, the notice and consent model can be seen as extending standard ideas from contract law to these sorts of agreements. Generally speaking, as long as someone knows what they are agreeing to when they sign a contract, it will be enforceable. More importantly, courts will generally presuppose that you know what you agreed to when you signed the contract.

3.2.1 Consent intro

Let's get started by talking about consent in general. Once we've got a grip on that, we can turn to our particular topic.

3.2.1.1 Political obligation and consent

To get the hang of consent in this context, let's start with something completely different —the question of when people have political obligations. That is, when can a state (morally) demand that someone pay taxes, serve in the military, or otherwise do stuff.

3.2.1.1.1 Early days

As western Europe moved from feudalism to something more like democracy, the question of how we can justify citizens' obligations to their rulers took on a new significance. Where in the past , this could be answered via the alleged relationship between gods, kings, and nobles —god says the king gets to demand anything he wants from anyone in his lands, the king says I get to demand whatever I want from you, peasant, so give me your crops. Oh. And I'm gonna need you to fight in my armies.

Or, probably more commonly, just a naked exercise of the force and violence feudal rule was based on: I want your crops and military service. Why? Because I have swords and loyal retainers to use them on you. Questions?

But as the idea that political legitimacy and authority came up from the governed (I.e., land-owning white guys) rather than down from the heavens or just at the point of a sword, philosophers like Locke tried to explain how it is that the governed have moral obligations to their governments.

3.2.1.1.2 Explicit consent

The easiest way to get political obligations is to have explicit consent. Thus the easy case is something like naturalized citizenship. You hold your hand up or put it on a book or over your heart or whatever and say that you agree to be bound by the laws and obligations of the land. But that's not most people. Most citizens never agreed to be, well, citizens. Nor does it make sense of the political obligations non-citizens may have to the state in which they reside.

3.2.1.1.3 Tacit consent

For Locke, citizens and residents did in fact consent. Sort of. They didn't say they did, but they acted like it. This is the idea of tacit consent. It has a bunch of different versions. But, basically, if you accepted the benefits of living under a government —drive on the roads, get educated, survive because no steppe nomads sweep in and kill everyone in your town— then you have tacitly consented to be obligated.

Very few writers these days believe this form of tacit consent genuinely creates obligations. Folks like Nozick and Simmons have pretty well ways of spelling out how tacit consent would work like the so-called principle of fair play (if you benefit from the cooperative activities of others, you have an obligation to contribute).

For our purposes, that's fine. We just needed the idea of tacit consent to in order to start making sense of how users might be consenting to data use. Let's turn back to the topic at hand.

3.2.1.1.4 Wild west internet

Collection of data on the early internet may have relied on something like the idea of tacit consent.

There weren't a lot of rules and disclosures. The attitude was basically: You come to our website. Our website is ours. We get to keep any data you produce while using it. If you don't want that, then go away. No carefully crafted terms and conditions. No privacy policies. Just: if you use our stuff, you agree. Even if you weren't exactly clear on that.

But then money started being made. Websites were being put up by proper companies; you know, with corporate structures and everything. Once that happened, the legal department got involved. All fun stopped. As with other businesses, the web needed to be well-defined agreements between the company and its clients.

In particular, customers needed to be actually consenting to the use of their data. To understand that, we need a more robust form of consent.

3.2.1.2 Informed consent

What we probably want is something more like the notion of informed consent that we use in fields like medical ethics.

As we'll talk about more below, fields such as medical ethics have been hammering out what informed consent looks like in practice for a few decades now. For now, let's just say that for there to be genuine consent to anything, the person needs to be at the least: informed, able to make decisions, and the choice must be voluntary.

3.2.1.2.1 Informed

Informed consent requires information. It's right there in the name. More specifically, it requires accurate and understandable information. Again, both parameters vary.

3.2.1.2.1.1 Accuracy

Obviously, the information must be true. But our understanding of accuracy here needs to go a bit beyond all the sentences being true.

For one, when the goal is allowing people to genuinely agree to something, in some cases, that may require telling them things which are strictly speaking false. Oftentimes the way people will understand specialized information is through analogies, metaphors, and the like. An immunology student will groan at all the disanalogies involved in the claim that 'White blood cells are like pac-men roaming your body eating germs, right now they are very weak and tired', but that may be accurate enough for the leukemia patient to use in reasoning about treatment options.

3.2.1.2.1.2 Understandable / level of detail

Similarly, the information must be understandable by the patient. This bar will be set in different places for different patients, depending in part on the background knowledge they bring to the exam room.

For example, through my research on pain I spend way more time around medical journals than your average patient. I've had to explain to a surgeon that if he wanted me to sign the consent form, he would have to explain the procedure and risks as if he was talking to a colleague about my case. The point is not that I'm more comfortable with medical jargon than others. It's that I know enough about (some!) areas of medicine that I need a higher level of technicality for me to be able to make an informed choice.

3.2.1.2.2 Capable of deciding

Hopefully you still vaguely remember enough about autonomy and Holley's paper from before the first exam that you will not be surprised to hear that it's not enough to have the information in front of you, multiple options, and no gun to your head. If you are a very small child, your choice will not be genuine consent. You need to be capable of processing the information and understanding how the potential risks and benefits fit with your preferences and desires.

Thus whether you have the sufficient capacity for consent will depend to some degree on the nature of the choice. If the stakes are low, the required capacities will often be less. Children can consent to agreements around cookies and chores. When they are high, a higher degree of psychological stability, reasoning skills, background knowledge, and self-reflection will be needed.

3.2.1.2.3 Voluntary

As we've discussed throughout the semester, autonomous choice requires liberty —the ability to wiggle or refrain from wiggling as one chooses.

With Holley, we saw this built into his conception of a voluntary exchange with the non-compulsion condition. One must have more than one option for the choice to be legitimate.

3.2.2 What must be disclosed

Turning (finally!) to the topic at hand, what information will a user need to know to make intelligent choices about whether to agree to a company's privacy policy?

Very broadly, you need to know anything which is materially relevant to the decision about whether the benefits of the service you'll receive outweigh the costs (including risks) associated with the information you are giving up.

In a sense, what you need to know is anything which bears on the question 'How much might surrendering control over this data hurt me?' Of course, this is unanswerable. Once your information is out there, it's not coming back. Who knows how clever smear artists might twist your college grocery store purchase history during your run for president.

Fortunately, the question isn't just what the consumer would need to know. It's what the company seeking her business must disclose to her. Presumably the grocery store has no more understanding of future political character assassination risks than you do. (Or do they….)

So we can at least limit the scope of our question to information which bears on risks which the company is or could be aware of. Barocas and Nissenbaum give a partial list for what the user of a newspaper publisher's website would need to know :

> (1) Which actors have access…; (2) What information they have access to…; (3) What they do or may do with this information; (4) Whether the information remains with the publisher or is directly or indirectly conveyed to third parties; and (5) What privacy policies apply to the publisher as compared to all the third parties, assuming these are even known to the users. These still constitute…only

a subset of what a user might need to know in order to be meaningfully informed {Barocas:2009ws}

Notice that this list may not translate smoothly to other industries and uses of personal data. The use of your purchasing history by a grocery store loyalty program may require more or less information along different axes.

3.2.3 Strengths of notice and consent model

The notice and consent model is appealing for several reasons.

3.2.3.1 Spells out all details

The notice and consent model makes everything completely clear by spelling out all the details of the agreement.

Okay, so 'makes things clear' is not actually right. A lot of times, the reason why the 'leagalese' of contracts and legal documents is so confusing and difficult to understand is actually due to the company's lawyers making things as clear as they possibly can. That's what lawyers drawing up contracts and agreements are doing: They are trying to think of every possibility, state precisely what it is, and state precisely who owes what. We'll come back to this later.

This benefits the service provider by shielding them from various forms of liability that come from incomplete disclosures and by shifting risk onto the user.

It benefits the user by allowing them to pick and choose which companies they deal with on the basis of their preferences with respect to personal data privacy.

3.2.3.2 Gives consumer basis for trust

The notice and consent approach benefits the consumer by providing some legal basis for consumer protection and trust. With any rule-bound bureaucracy and especially the law, if it says in clear language in black and white on a piece of paper that the company will not do x, you can have decent confidence that they will not do x. Obviously, there are exceptions. It assumes people in the company know their own policies or that they are confident enough that their lawyers can keep the costs of paying lawsuits over their doing x low. But, on balance, we shouldn't ignore the consumer benefit of having everything spelled out in excruciating detail.

3.2.3.3 Opportunity for market differentiation

Leaving the management of personal data to be determined by be what companies and their users agree makes it possible for different companies in the same space to differentiate themselves through their privacy policies. For example, Apple seems to be putting a lot of effort —both marketing and engineering— to protecting privacy in their mobile devices.

Of course, consumer privacy advocates have long known that as a general matter this is not on its own sufficient. Consumers both chastise companies for misusing their personal data and are uninterested in paying directly for the services being funded by this alleged 'misuse'.

3.2.3.4 Other strengths

[ToDo]

Other strengths of the notice and consent model include

Individual

- Control of information: Allows individuals to evaluate options deliberately and decide whether to give or withhold consent

- Respects individual choice

Company

- Allows for more collection of data and thus better trained models / better services

Economy level

- Efficient market: Allows market to function efficiently whereby individuals can decide when the price is right
- Efficient allocation of services

3.2.4 Problems with notice and consent

To explore the problems facing the notice and consent model, it will help to do a bit of rough taxonomy to keep the issues straight. Let's distinguish between problems with the notice part and problems with the consent part.

Obviously, this division is highly artificial. Since genuine consent requires being informed, many of the problems with the notice part undermine the consent part. I'm cutting it up this way just to help organize, so don't worry too much about the taxonomy, at least not until we're done worrying about the cases.

3.2.4.1 Consent part

Let's start with problems affecting the consent portion of 'notice and consent'.

3.2.4.1.1 Inescapable / no real choice

Anytime we're talking about people having adequate options / choices, we quickly run into difficulties judging whether a person really has alternatives. Take for example social media companies such as Instagram. Do you have a choice other than accepting their policies?

On the one hand, it's tempting to say that people don't need social media, thus they always have the option of not using the sites. Obviously, if we're taking 'need' here in the sense of immediate survival —I need water— then we don't need social media. Indeed, humanity was doing okay for the millennia up to the demise of Friendster and MySpace (no one needed those early social media platforms).

But on the other hand, we are social animals. If all your friends are using Snaptagram, using Snaptagram may be the only way of interacting with them. Loss of social integration can be a terrible loss. You can survive without food for a few months. Going without friends for that long may not kill you, but it can still be a signifiant hit to your well-being.

Indeed, one of the value propositions behind social media companies is network effects: Once your site's user base is large enough, it sucks in everyone who wants to interact with them. In other words, once such companies exist at scale, people need to be able to use them.

3.2.4.1.2 Hard to judge risks / consequences of agreeing

It is very very very hard to understand the risks different uses of your personal information may pose. This is in part because there are tons of uses of personal data which exist right now. But the uses which exist now are just the beginning. Once data escapes into the wild, it does not return. Every person will have to anticipate what kinds of problematic uses may get dreamt up in the future. That's incredibly hard.

But the task is actually much harder. There will not be just one list of concerns which we could all look at and make individual choices about where we draw the line. It also matters who you are. As is unfortunately too common with society's laws, being older, whiter, and male-r tends to insulate people from ill effects. The middle aged white guy gets a warning; the young black woman gets a speeding ticket. There is no reason to think that these patterns will not translate to the ill effects of various uses of personal data.

For example, there are companies which purport to use machine learning on job applicants social media presence to determine whether they are likely to be good employees. Guess whether these companies target low wage jobs or Fortune 500 companies for C-suite positions. Thus there is no one set of risks from the use of personal data.

In addition to there being tons of uses, it is difficult to know how to weight the risks. As we talked about at the beginning of the semester, it isn't enough to know how bad something might be. You also need to know how likely it is and weight accordingly. A lot of the relevant data you would need to assess risks of different uses by currently existing companies is proprietary or not well documented.

And just to emphasize this again, data outlasts the use. Maybe things folks are doing now are fine. But maybe down the road, data collected today will have completely unforeseen uses. I'm sure that companies five years from now will be substantially more enlightened, ethical, and scrupulous about their customers' privacy. Ok. That's a lie.

3.2.4.1.3 Difficulty of knowing competence of person agreeing

So far, I've mostly talked about the problems of the notice and consent model from the consumer side. But there are concerns from the company side too. Suppose you run a well-intentioned company and want to make sure your privacy policies are clear and that your users really will understand the trade-offs you are presenting to them. How do you know whether your users are actually competent to agree to use your service? You can't give a test. You have to aim for your average consumer. You might be able to estimate the level of education needed for the language. But you would also need to ensure that the users are properly weighing the risks. It's hard enough to know when, as individuals, we ourselves are taking risks seriously but not overblowing them. Imagine trying to ensure this for millions of users from myriad walks of life from around the globe.

3.2.4.2 Notice part

Let's turn to problems with the notice part of the notice and consent model.

3.2.4.2.1 TL;DR

Probably the biggest problem is that people do not and cannot be expected to read policies. They are long, legalistic, and difficult for the average user to understand. Indeed, even specialists have trouble deciphering the implications of some policies [ToDo: add ref]. We might call this the TL;DR problem, if we were stuck using internet slang from 10 years ago.

Now, it may be tempting to dismiss these concerns as the customer just being lazy or dumb. But even if they were written in the clearest, most accessible language, the length alone is serious burden and barrier.

To illustrate this, let's do some rough math. A meta-analysis of studies estimating reading speeds in English comes up with approximately 238 words per minute for non-fiction.[13]

I downloaded and very roughly estimated the word count of Ralphs supermarkets privacy policy on 31 October 2019. It was approximately 2204 words. At the 238 wpm, it would take 9.26 minutes. Doing the same with Google's privacy policy (not the terms of service) yielded about 7311 words. That's 30.72 minutes of reading. The LA Times privacy policy came in at 4343 words. That's 18.25 minutes of reading.

Thus with just three websites, and only considering their privacy policies, we're now up to about an hour out of your life. Hopefully it's becoming clear how big a task this is. Multiply the average word count in privacy policies with the number of companies that the average consumer will need to interact with; I can't imagine that the result would be anything like a manageable amount of time which each consumer would need to commit to reading privacy policies.

3.2.4.2.2 Policy revisions

Even once you've read the policy and agreed to it, you'r e not done. Companies may change their terms at will. Sure, they will usually notify you and summarize the changes. That lowers the burden of reassessing whether you still want to agree. But its nonetheless a burden. Multiply those changes across the number of user agreements we are bound by and we're talking about a not insignificant amount of time.

3.2.4.2.2.1 Lock-in

Indeed, as Hoofnagle points out, other costs of the reassessment may rise. Suppose you've spent several years using a company whose privacy policies are satisfactory to you, there may be a significant cost to switching to a new company when the company's policy changes. As Nissenbaum summarizes the concern:

> In July 2007, Susan Wojcicki, Google vice president of product management for advertising, suggested that OBA [Online Behavioral Advertising] was "not something that we have participated in, for a variety of reasons," and that Google wanted to "be very careful about what information would or would not be used"

[13] https://psyarxiv.com/xynwg/

> for the purposes of advertising And yet, as we know, Google is now potentially the most dominant player in the OBA field. This about-face should give us pause…[because] users who relied on Google's aversion to behavioral targeting from 2000-2006 may "have already used Google for years and may have some lock in from adopting the company's many services" [On notice]

3.2.4.2.3 3rd party use

Perhaps the biggest concern —this will be a big driver of the Transparency Paradox below—arises from companies sharing and selling data to other companies.

Sure, you might decide that you can trust the company you're giving consent to. But what about the company who buys your data from the company who bought your data from the company you agreed to allow to harvest your data? Who's keeping an eye on them?

Once there is a market for personal data, businesses in possession of your data may change their partnerships at will. Customers can't really assess constantly changing web of business partners. Even identifying which companies these are will be prohibitively difficult. Assessing all of their privacy policies just multiplies all the problems we've already seen.

The online advertising space pours rocket fuel on the pace of such changing relationships. As Barocas and Nissenbaum note, brokers such as ad exchanges, who provide a marketplace for algorithmic auctions of ad impressions, muddy the waters with respect to the privacy policies actually governing your data.

> Ad exchanges replace semi- stable contractual relationships concerning the sale of impressions or transmission of user data with fleeting relationships based on real-time auctions that may nonetheless result in the equally permanent transmissions of user data. {Barocas:2009ws}

3.2.4.2.4 Legalese

Few of us can understand the policies. Let's talk a bit about why that is.

Obviously, these are legal documents, written by lawyers. That poses challenges to the consumer actually being 'notified'. Notice that in most areas of law this doesn't matter. The US legal system, excluding criminal law (sort of), presumes that if you have a legal dispute which you care enough about not losing, you will hire a lawyer.

But these aren't legal disputes. These are agreements between a consumer and a company. It is fairly unlikely that even rich customers run user agreements for social media companies past their lawyers. And they have lawyers.

3.2.4.2.4.1 Interpretation against body of case law

Why would you need a lawyer? Couldn't any of us pick up a contract and read carefully? Well, yes and no. On the one hand, a lot of dense legalese is actually just enumerating all sorts of scenarios and explaining what will happen in them. But on the other, any contract will be interpreted against the laws and (often more importantly) the case law of its jurisdiction. This is why people go to law school. Every profession has certain shared understandings, short hands, and fixed reference points.

If you're not convinced, here's an example from medicine. If you've ever seen a paper prescription[14] you might have puzzled over where the dr is saying how often or even how you should take it. For example, 'BID' means twice a day. If you are an amateur and relying on hasty google searches, you may confuse *p.r.n.* (as needed) with *p.r.* (stick it up your butt —per the rectum)[15] . (I confess I have waited years to find an opportunity to make this joke)

3.2.4.3 Other problems

While we are jumping up and down on the notice and consent model, let's get on the table some non-notice and non-consent problems it faces.

3.2.4.3.1 Economic externalities

[14] And, seriously, you really shouldn't. Paper prescriptions are a huge source of medical error since the pharmacist has to interpret your doctor's handwriting.

[15] https://www.drugs.com/article/prescription-abbreviations.html

Pressure toward either monopolies or data market saturation.

[ToDo: Figure out what this was supposed to be. This is from student suggestions during a previous class; it must have made sense then….]

3.2.4.3.2 Affects the value we put on privacy

Some writers worry that if people get too used to giving up their privacy, they won't value it as much. That in turn makes it easier for companies and other entities to chip away at our privacy, which makes people value it even less. Rinse and repeat.

That said, it's likely true that there is a lower limit on how much privacy we'll be willing to give up. The early utilitarian reformer Bentham thought that prisoners would benefit from being watched 24/7.[16] That would help them build good character since they'd learn to act like they are always in public. Unfortunately, these ideas did get adopted by prison architects who built prisons so that the prisoners never could tell if they weren't being watched. I say 'unfortunately' because Bentham was very very wrong about what happens to people when they think they are being watched all the time. It is not good for their character; it is probably not good for their sanity.

3.2.4.3.3 Spider lawyers

A team of highly skilled lawyers churning out carefully vetted agreements is as close to a Star Trek deflector shield as we are likely to find in nature. If the starship Enterprise was a company, Captain Piccard commanding 'shields up' would be a call down to legal.

However, whenever you're developing a policy, you are always navigating between the need to keep it flexible enough to handle the random problem cases that you will never dream up while precise enough so that everyone can understand what it means. If a company's policies are complex and detailed, the chance that something unenforceable or, worse, illegal in a particular jurisdiction will sneak in (or fail to be removed when laws change) grows.

[16] https://en.wikipedia.org/wiki/Panopticon

There are [ref] groups who use web-spidering techniques (spiders because they 'crawl' the web —don't look at me, I didn't make this one up) to digest user agreements by the thousands. If they find something actionable, they sign up as a user, light an expensive cigar and then file a lawsuit. Done correctly, the lawsuit will cost just enough that the company would lose money having it thrown out in court. So, detailed user agreements can pose some litigation risk for the company.

At the broader economic level, if there's enough of this sort of litigation, it can create economic inefficiency. That's not to say that economic inefficiency is always a bad thing. But it is something which we want to be attentive to when we are thinking about how to approach these problems at a national level.

3.2.5 Attempts to fix notice and consent

It does seem that policymakers, some sectors of the public, and some data-industry actors are aware of the problems with the notice and consent model and interested in repairing it.

How exactly the repairs might work depend on what we think the problems are.

3.2.5.1 Standardizing and simplifying policies

If you think the problems arise from the policies being too confusing for the consumer to understand, the fix might be to find ways to make the policies more standardized and simpler. Indeed, finding more intuitive and clear ways of conveying the information might also help; this will probably require help from experts in visual communication and other aspects of consumer psychology.

The model here might be nutrition labels. Nutrition labels on food are imperfect. But they do a far better job conveying important information to the average consumer than a biochemist's report.

Similarly, in the wake of the Great Recession, legislation [ToDo: NAME? Was this in Dodd-Frank?] standardized a number of consumer financial documents. All credit card terms and conditions need to be formatted the same, use the same time-frames for interest rates, et cetera. Again, this is in the service of helping consumers understand what they are signing up for by standardizing and simplifying a complex financial arrangement.

3.2.5.2 Stavra example

Don't believe me that this is an approach folks like? Here's an email I received from the cycling app Stavra on 11/13/2019, note the first item:

> Hi adam,
>
> Privacy. It's something we all care about, but making the time to read the fine print can be tough. So as part of this Privacy Policy update, we've created a way to understand the most important details at a glance.
>
> **Meet our new Privacy Label.** Like the nutrition label you've come to trust, we've created a no-nonsense list of need-to-know privacy facts. Take a look.
>
> **We didn't sell your personal information before, and we don't sell it now.** We're excited to give athletes even more clarity about how information is shared and how you can control your privacy settings. Read about this and other updates in our latest Privacy Policy, effective December 11, 2019. By continuing to use Strava after this date, you agree to our updated Privacy Policy.
>
> **The details you need, all in one place.** Our Privacy Center has what you need to make informed choices about your data.
>
> **A new privacy law from California.** We're giving all Strava athletes, regardless of your location, the same tools and controls as Californians under the California Consumer Privacy Act. And for athletes living in the European Economic Area, while nothing's changed about your rights these disclosures are always helpful to review.
>
> We take your privacy seriously and are committed to making sure you can hit record, upload to Strava, give kudos and cheer on your friends with confidence that your personal information is safe and sound.
>
> Cheers,
>
> The Strava Team

3.2.5.3 Opt out vs opt in

Alternatively, if you focus on the 'take it or leave it' nature of the user agreements, you might try to find ways to allow users to 'opt out' of data collection without losing all access to the service.

This is attractive because it recognizes the reality of consumer decision making. The economist's idealized (and presumably ideally wealthy) consumer makes carefully considered purchasing decisions in the ideal competitive marketplace and thus opts out of any agreement that is not in her interest. None of us are the idealized consumer; at least not all the time.

3.2.6 Transparency paradox

Let's summarize the task that lies before anyone hoping to fix the notice and consent model.

For someone to evaluate the risks and benefits of sharing their data with a company, they will need to know at least:

1) What information will be collected

2) How long the information will be retained

3) How the information will be anonymized, if at all

4) How information will be processed and used (group level statistics; individual targeting; etc)

5) Which 3rd parties the data will be shared with, and what information will be shared

Each of these will likely be complex and require detailed disclosures. Obviously, the level of detail and precise information involved will vary depending on the nature of the service and the nature of the data collected. So the task may be more difficult in some areas than in others.

Still, even the easy cases will be a significant challenge since it's not enough to simply disclose the information. The information must be conveyed to the consumer in ways that are relevant and meaningful, while still being digestible in a short period of time.

Nissenbaum thinks that this will basically be impossible. She argues that the notice and consent model is fundamentally misguided. The project of attempting to simplify, standardize, or otherwise fix up the way information about personal data is disclosed is doomed. This is due to what she calls the *transparency paradox*. The paradox is roughly the tension between two claims:

1) If a policy gives the consumer adequate information, they will not read or understand it.

2) If a policy gives the consumer understandable information, it will not adequately inform them.

3.2.6.1 Paradoxes

To keep the philosophy majors from going nuts, let's note that, strictly speaking, the transparency paradox doesn't involve a paradox in the standard philosophical sense.

When philosophers talk about paradoxes, we usually mean situations where (very roughly) we are forced to choose between two equally plausible but incompatible claims. The most famous and short paradox is the liar paradox. Consider the following sentence:

(LS) This sentence is false.

Is LS true? Well, if it is true it is false. So, it must be false. But, remember, false is equivalent to not true and vice-versa. (c.f., You: Are those tacos in the bag? Me: They are not not tacos). Therefore, if LS is false, it is true.

I'll wait while you pick up the pieces of your blown mind. Go[17] ahead. I'll be right here…

[17] Mind not blown? Okay, how about the weaponized version, Godel incompleteness? Here's some videos
Computerphile: History of undecidability pts1 -3
https://www.youtube.com/watch?v=nsZsd5qtbo4
https://www.youtube.com/watch?v=lLWnd6-vSGo
https://www.youtube.com/watch?v=FK3kifY-geM
Another video from Numberphile:
https://www.youtube.com/watch?v=O4ndIDcDSGc
Obviously, while interesting and guaranteed to make you the life of any party, none of this is relevant to our class.

3.2.6.2 The transparency paradox

The transparency paradox comes about because, if the notice (privacy policy) details everything, the average consumer will not read or understand it. At the same time, if the notice is readable and understandable, it will omit crucial information.

However, we need to be careful about what's generating the problem. Otherwise, the transparency paradox might seem trivial.

After all, simplifying complex information virtually by definition leaves stuff out. That's obvious. Indeed, the fact that we leave stuff out when we summarize isn't usually a problem. Suppose your friend asks you what Game of Thrones is about and you answer "It's a fictional version of the English Wars of the Roses but with dragons and zombies." There are certainly better answers. But if your friend knows a bit of history or knows how to use Google, it's a reasonably informative answer. She could make a decision on whether to check it out.

But if the transparency paradox showed that all attempts at summary are doomed, you'd have to reply "It is impossible to tell you. Go buy the books, subscribe to HBO, and we'll talk in a couple of weeks." Alternatively, no introductory textbook would ever be usable.

Put another way, to understand the Transparency Paradox, we need to know why summarized privacy policies will never be good enough for consumers to make choices about using a service. This is because, Nissenbaum claims, attempts to simplify these policies will necessarily leave out essential information. The question is thus what guarantees that this will happen when we are talking about privacy policies.[18]

3.2.6.3 What guarantees the paradox

[18] You might detect echos here of Holley's distinction between an ideal exchange and an acceptable exchange —in the former, the buyer knows everything about the product. But in the latter case, the buyer is informed enough (etc) to have a decent shot at there being a mutually beneficial exchange.

Nissenbaum thinks that the nature of the data collected and the nature of the industries doing the collection will provide the guarantee that we run into the Transparency Paradox.[19]

Big data techniques which attempt to extract novel insights from data require huge datasets. (Hence the name, 'big data'). Generally speaking, the more data a company can snarfle up, the more useful it becomes. Thus, except for companies like google who keep all their data in-house and sell access to the products of it, in most cases you would need to know not just how the company you are immediately doing business with will use your data.

Indeed, you need to know several layers of business relationships down. These are complex webs of companies, which often change. Thus you would need to know how all of these companies are handling your data.

That's what drives the transparency paradox. The structure of the industry entails that you need to know the policies of multiple companies. That fact alone seems to doom the simplification project.

3.2.6.4 Objection: informed consent works in medicine

So far it looks like Nissenbaum has pointed to a genuine tension. All parties to this problem can agree that it is hard to resolve this tension. But she is claiming that it can't be fixed; that we should think that the notice and consent model is doomed. Let's think through how sure we are that the tension cannot be resolved.

When someone points out a problem and claims that it can't be fixed, it's usually a good idea to look for analogous situations and consider whether we have solved the problem there. If we can find others where we have fixed the problem, that will cast doubt on her claim that the notice and consent model is doomed.

[19] Discussed in Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Daedalus*, *140*(4), 32–48. http://doi.org/10.2307/23046912?refreqid=search-gateway:4971cc18293f6167ab2b2c12057373be

One place we might look is the use of informed consent in medicine. The relevant information for consenting to a medical procedure is often intensely complex and technical. It very often involves weighing probabilities and fitting those probabilities together with other decisions. Human beings are pretty bad at that. Thus it looks like the transparency paradox arises for informed consent in medicine too.

However, we know that informed consent in medicine generally works. It's not perfect. But in most cases, the patient probably is well-enough informed to decide to undertake a medical procedure. We have decent guidance for practitioners on how to inform patients; similarly, human subjects review boards are reasonably sophisticated.

So, if the transparency paradox can be overcome in medicine, it looks like we needn't be completely pessimistic about privacy.

3.2.6.4.1 Response: Nope not fixed in medicine

Nissenbaum's response to this objection is straightforward: No. We haven't solved the transparency paradox for informed consent in medicine. Instead, we have a system that relies on other factors to ensure patient protection. She writes

> "these protocols work not because they have found the right formulation of notice and consent but because they exist within a framework of supporting assurances....It is not the consent form itself that draws our signature and consigns us to the operating table, but rather our faith in the system. We trust the long years of study...that physicians undergo, the state and board certifications, peer oversight, professional codes, and above all, the system's interest (whatever the source) in our well-being" [36]

Medince didn't solve the transparency paradox. It evolved institutions which allow us to trust doctors.

3.2.6.4.2 Response: Unknowability

Indeed, Nissenbaum and Barocas offer what seems like an even stronger version of this claim, viz., that the things you would need to know are fundamentally unknowable:

Complexity constitutes a challenge, generally, for achieving meaningful notice, but OBA is unlike surgery in two significant ways. First, given adequate time, training, and education, a person confronted with a medical decision could, in principle, fully understand what they were consenting to. In the case of OBA, however, there is a degree to which the tracking, analysis, and use (current and future) of data is not only difficult to grasp, but unknowable. As we noted above in our description of the capture and processing of information, there is potentially an unending chain of actors who receive and may make use of behavioral and other data. New companies bloom, novel analytical tools emerge, business relationships begin and end. In the currently preferred model, when people consent to OBA—or fail to opt out—they literally cannot know what they are consenting to. {Barocas:2009ws}

3.2.6.5 Alternative approaches to saving notice and consent

Suppose Nissenbaum is right that the institutions and norms around medicine are what create the ability for patients to make informed choices, not just the way the info is presented. Perhaps there are other ways to try to fix the model which follow in medicine's footsteps.

3.2.6.5.1 Privacy consultants

One suggestion, raised independently by several brilliant business ethics students, is to try to fix notice and consent by importing something like the trusted intermediary which did the work in medicine.

For example, maybe we just bite the bullet on the transparency paradox —we let the legal department keep writing the notice forms— and create an industry of privacy consultants who have the expertise to interpret them and provide a trustworthy opinion to consumers based on the consumers privacy preferences.

If you could trust that your consultant is evaluating privacy policies on your behalf based on your preferences, we would work around the paradox. As with medicine, professional standards of practice, social expectations, and other reenforcing norms would be crucial to making this work.

3.2.6.5.1.1 Concerns about privacy consultants

Of course, even if this strategy would in fact preserve the notice and consent model, we would then have to take up whether it is desirable.

Certainly, there will be economic concerns. It adds another layer of professionals required for ordinary people to conduct their lives imposes costs.

There will be social concerns. If all consumers remain subject to such personal data policies but only some can afford a privacy consultant, the harms of personal data use may fall disproportionately on some groups in society.

There will also be political concerns. As is perfectly normal, a group of professionals will form an interest group. Presumably, such professionals will advocate for consumer interests. But that advocacy will be inevitably filtered through their professional interests. The union covering CSU professors is a strong advocate for CSU students in state politics. Of course, we believe that the best way to help students is to empower (and pay well!) full-time professors.

3.2.6.5.2 Ratings organizations

Alternatively, public interest groups could take up the work of parsing privacy policies and making recommendations to consumers. A quick Google search revealed one such organization called Terms of Service; Didn’t Read: https://tosdr.org/

3.3 Nissenbaum's approach

If notice and consent is beyond saving for personal data, what should we do instead? How should we make policies that protect people’s data online?

Instead of trying to extend ordinary contractual practices to online privacy, Nissenbaum wants us to stop treating issues of online privacy (and presumably privacy in other data-snarfling environments) as something special which requires a new approach. Instead, she wants us to recognize what tools we already possess and work from what we already know. We are to start by recognizing that

"Online activity is *deeply integrated* into social life in general and is *radically heterogeneous* in ways that reflect the heterogeneity of offline experience." [37]

The online realm isn't separate from regular life. The issues it poses aren't generally distinct from what we already deal with. Instead of looking for new tools and new theoretical justifications, we should use what we already have.

3.3.1 Contexts

Let's start with what counts as a context. As sociologists and theorists have long recognized, in daily life we pass through many different socially defined contexts. Think of the differences in how you act / how you are expected to act within your roles as a student, a roommate, an employee, a child in a family, a parent in a family, a patient in the doctor's office, a teammate, a friend, a lover. Each of those contexts has it's own rules. It may be appropriate to pat your lover or teammate on the butt in congratulations. Your doctor will react poorly to this.

3.3.2 Contextual integrity

Just like rules about what forms of intimate personal contact are appropriate, contexts have rules about information flows. Indeed, she claims that every social context, has certain norms and assumptions about data sharing and use.

The basic idea on Nissenbaum's view is that when we move to the online realm, we should preserve the norms which already govern information in similar contexts. The main task of informational privacy from her perspective is to try to determine how we can extend our norms, policies, and laws which govern information in non-online contexts, to newer arenas. She writes:

A central tenet of contextual integrity is that there are no arenas of life not governed by norms of information flow, no information or spheres of life for which "anything goes." Almost everything—things that we do, events that occur, transactions that take place—happens in a context not only of place but of politics, convention, and cultural expectation. These contexts can be as sweepingly defined as, say, spheres of life such as education, politics, and the marketplace or as finely drawn as the conventional routines of visiting the dentist, attending a family wedding, or interviewing for a job. For some purposes, broad sweeps are sufficient. As mentioned before, public and private define a dichotomy of spheres that have proven useful in legal and political inquiry. Robust intuitions about privacy norms, however, seem to be rooted in the details of rather more limited contexts, spheres, or stereotypic situations. {Nissenbaum:2004uu} p.119

These contexts follow the information. Even when it leaks into another context.

One point of contrast with other theoretical accounts of privacy rights is that personal information revealed in a particular context is always tagged with that context and never "up for grabs" as other accounts would have us believe of public information or information gathered in public places. A second point of contrast is that the scope of informational norms is always internal to a given context, and, in this sense, these norms are relative, or non-universal. [125]

3.3.3 Informational norms

Nissenbaum believes that

technologies, systems, and practices that disturb our sense of privacy are those that have resulted in inappropriate flows of personal information. Inappropriate information flows are those that violate context specific informational norms (from hereon, ''informational norms''), a subclass of general norms governing respective social contexts. {Nissenbaum:2015ki} p.839

She argues that there are 2 broad types of informational norms at stake: Norms concerning appropriateness and norms concerning distribution. As she writes:

I posit two types of informational norms: norms of appropriateness, and norms of flow or distribution. Contextual integrity is maintained when both types of norms are upheld, and it is violated when either of the norms is violated. The… benchmark of privacy is contextual integrity; that in any given situation, a complaint that privacy has been violated is sound in the event that one or the other types of the informational norms has been transgressed {Nissenbaum: 2004uu} p.120

3.3.3.1 Norms of appropriateness

Concerning norms of appropriateness she writes that

norms of appropriateness dictate what information about persons is appropriate, or fitting, to reveal in a particular context. Generally, these norms circumscribe the type or nature of information about various individuals that, within a given context, is allowable, expected, or even demanded to be revealed. In medical contexts, it is appropriate to share details of our physical condition or, more specifically, the patient shares information about his or her physical condition with the physician but not vice versa; among friends we may pour over romantic

> entanglements (our own and those of others); to the bank or our creditors, we reveal financial information; {Nissenbaum:2004uu} p.120

And

> As important is what is not appropriate: we are not (at least in the United States) expected to share our religious affiliation with employers, financial standing with friends and acquaintances, performance at work with physicians, etc. As with other defining aspects of contexts and spheres, there can be great variability from one context to the next in terms of how restrictive, explicit, and complete the norms of appropriateness are. In the context of friendship, for example, norms are quite open-ended, less so in the context of, say, a classroom, and even less so in a courtroom, where norms of appropriateness regulate almost every piece of information presented to it. The point to note is that there is no place not governed by at least some informational norms. The notion that when individuals venture out in public—a street, a square, a park, a market, a football game—no norms are in operation, that "anything goes," is pure fiction. For example, even in the most public of places, it is not out of order for people to respond in word or thought, "none of your business," to a stranger asking their names {Nissenbaum:2004uu} p.121

She quotes the philosopher Ferdinand Schoeman to illustrate one way a person can violate norms of appropriateness by moving information between contexts

> "[p]eople have, and it is important that they maintain, different relationships with different people." [71-notes]

and

> [a] person can be active in the gay pride movement in San Francisco, but be private about her sexual preferences vis-à-vis her family and coworkers in Sacramento. A professor may be highly visible to other gays at the gay bar but discreet about sexual orientation at the university. Surely the streets and newspapers of San Francisco are public places as are the gay bars in the quiet university town. Does appearing in some public settings as a gay activist mean that the person concerned has waived her rights to civil inattention, to feeling violated if confronted in another setting?[72 —notes]

3.3.3.2 Norms of flow / distribution

Concerning norms of distribution she writes that her perspective is compatible with Michael Walzer's picture of different spheres of justice.

> According to Walzer, complex equality, the mark of justice, is achieved when social goods are distributed according to different standards of distribution in different spheres and the spheres are relatively autonomous. Thus, in Walzer's just society, we would see "different outcomes for different people in different spheres." Complex equality adds the idea of distributive principles or distributive criteria to the notion of contextual integrity. What matters is not only whether information is appropriate or inappropriate for a given context, but whether its distribution, or flow, respects contextual norms of information flow. [123]

And also that

> Free choice, discretion, and confidentiality, prominent among norms of flow in friendship, are not the only principles of information distribution. Others include need, entitlement, and obligation—a list that is probably open-ended. In a healthcare context, for example, when a patient shares with her physician details of her current and past physical condition, the reigning norm is not discretion of the subject (that is, free choice of the patient) but is closer to being mandated by the physician who might reasonably condition treatment on a patient's readiness to share information that the physician deems necessary for competent diagnosis and treatment. Another difference from friendship is that in the healthcare context, the flow is not normally bidirectional. Confidentiality of patient health information is the subject of complex norms—in the United States, for example, a recent law stipulates when, and in what ways, a physician is bound by a patient's consent: for example, where it is directly pertinent to diagnosis and treatment, where it poses a public health risk, and where it is of commercial interest to drug companies.[124]

3.3.3.3 3 elements

There are 3 key elements which Nissenbaum thinks affect whether sharing a particular piece of information is appropriate:

1) The type of information

2) The actors involved

3) The way the transmission occurs

She writes

> … informational norms are defined by three key parameters: information types, actors, and transmission principles….Whether a particular flow, or transmission of information from one party to another is appropriate depends on these three parameters, namely, the type of information in question, about whom it is, by whom and to whom it is transmitted, and conditions or constraints under which this transmission takes place. Asserting that informational norms are context-relative, or context-specific, means that within the model of a differentiated social world, they cluster around and function according to coherent but distinct social contexts. The parameters, too, range over distinct clusters of variables defined, to a large extent, by respective social contexts. {Nissenbaum:2015ki} p.839

3.3.4 Her opponent

THIS SECTION IS TERRIBLE AND NEEDS SERIOUS REWRITING. I CONFUSE THE DISTINCT AND DISTINCTIVE CLAIMS SEVERAL TIMES IN WHAT FOLLOW.

To get a better sense of how her view works, it will help to give her an opponent. That way we can see where she thinks the opponent is going wrong. Here's what her opponent is claiming:

> Online privacy is a distinctive venue defined by the technology and protocols of the net for which a single set of privacy rules can / should be crafted.

The opponent is thus making two claims.

3.3.4.1 Distinct

First, her opponent is claiming that the online realm and thus online privacy is distinct from other social realms. It requires its own set of rules, which will be different from the rules which operate elsewhere.

It may be helpful here to think about sports.[20] Each sport is *distinct*. For one, it has its own set of rules. If you know the rules of cricket, you know nothing about the rules of baseball. The rules of cricket and baseball are *distinct*.

Similarly, if you invent a new sport, you need to invent a whole new set of rules that will be distinctive of that sport. When mixed martial arts became a sport, rather than just an organized brawl, it borrowed rules from boxing and judo. But because punching someone in a judo match or choking a boxer get you disqualified in those sports, the MMA's rules make it a distinct sport.

If you are trying to invent a new sport and come up with something that has all the same rules as an existing sport, you have failed to create a distinct sport.

3.3.4.2 Distinctive

Second, her opponent is claiming that online privacy will have a distinctive set of rule. Online activity has its own nature and presumably that nature is shared amongst everything online. It would be a mistake to mix up online and offline activities.

The rules of baseball are *distinctive* of baseball. They make an activity a game of baseball. Setting aside minor variations, if you are playing a game that doesn't involve hitting a ball in order to run bases, 3 strikes and you're out, 4 balls and you walk, you are not playing baseball.

3.3.4.3 Contra distinct

Regarding the claim that online activity forms something distinct from regular life, she claims that the online environment

> "is not a single social realm, but the totality of experience to be conducted via the Net, from specific websites to search engines...crisscrossing multiple realms." [38]

To see the evidence for this, think of all the things you did online yesterday. Maybe you watched some movies on Netflix, checked out a cat in a dinosaur costume on a roomba on youtube, submitted your journal for class, scrolled through the pictures from your favorite celebrities and

[20] Cross your fingers, Adam's going to try a ball-involving sports analogy….

most attractive friends on Instagram, placed a refill prescription at the pharmacy, checked your bank balance, and impatiently swiped left before a lingering swipe right in a search for love (or whatever one finds on tinder).

How are those all the same thing? Sure you did them all on your phone, tablet, or computer. But why does that matter. You also called your mother on[21] your phone and texted your friend. Are those also the same? You worked on your paper or wrote a python script to make your life easier on the computer. Are those the same?

It doesn't seem like the instrument with which you did all these things should be decisive. If you found one of those weird phones which connect via a cord to the wall and called your mom on that, is that different?

Indeed, if you're like me, you may be extremely puzzled by the questions above. What would it even mean to say that all of these actives are the same? It just seems like they are different things which just happen to be done through the same device / medium. That, I think, is Nissenbaum's point.

3.3.4.4 Contra distinctive

Against the opponent's claim that online activity has its own is distinctive set of rules Nissenbaum argues that online activities are deeply integrated into social life. She writes

> "Not only is life online integrated into social life, and hence not productively conceived as a discrete context, it is *radically heterogeneous*, comprising multiple social contexts, not just one, and certainly is not just a commercial context where protecting privacy amounts to protecting *consumer* privacy and commercial information." [38]

For one, some online activities like shopping are continuous with real life activities. You might browse and purchase an item online and pick it up in a building made of brick-and-mortar called a 'store'.

At the very least, activities online have power to affect IRL communications, transactions, interactions, and activities (and vice-versa). Insult your friend online and then act like nothing

[21] You did call your mother, right? No? Go do it now. She's probably worried.

happened when you see them in person. If they get pissed off, condescendingly tell them that you insulted them online so they shouldn't be mad at you in person. See how well that works. (Don't do this.)

Or, to use another helpful analogy, it seems like her opponent's reasoning is like someone who rants that all criminals should be executed. They insist on talking about 'crime' without recognizing that there are many kinds of crime (e.g., murder, arson, mayhem, larceny, and jaywalking)[22] .

3.3.5 Decision heuristic

Let's suppose we buy that there are already existing norms governing information flows which apply to real life contexts and that these contexts align with online contexts. How should we go about translating our in-person norms to online norms?

Nissenbaum proposes a four part approach for determining what to do with informational privacy within a novel area.

(1) Locate relevant existing contexts

(2) Explicate entrenched informational norms

(3) Identify disruptive flows for the online contexts

(4) Evaluate disruptive flows against norms based on general ethical and political principles as well as context-specific purposes and values.

3.3.5.1 Locate relevant existing contexts

The tasks embodied in the first two steps seems clear. Figure out what the relevant in-person contexts are. Then figure out what the norms around information flows are in those contexts. Though of course actually carrying out those tasks may be tricky.

[22] Thanks to Johannah Caliban for this example.

In some cases the in-person contexts that correspond are obvious —online banking and in-person banking. In others, this is more difficult. We'll come back to that in a moment.

3.3.5.2 Explicate existing norms

Similarly, figuring out what the relevant norms are can also be difficult. She wants us to start with things like laws and policies. Those are large scale reflections of commonly held norms.

It is worth noting that in a society that is diverse on religious, cultural, political, and other dimensions, going with the laws enacted by those in the political majority may be misleading. Indeed, different populations may have different norms around information. Think of how some families seems to share and fight about everything that's on anyone's minds whereas other families keep disagreements under wraps, save for the occasional passive aggressive comment over dinner. That may not be cultural in the relevant sense, but at least you get the idea.

That said, one strength of Nissenbaum's view is that it is flexible and can handle all sorts of messiness in actual norms. We'll come back to this later too.

3.3.5.3 Identify disruptive flows

The next step is to turn to the online context and determine how it is that information may be flowing across contextual boundaries.

3.3.5.4 Evaluate flows against norms

Finally, we are to "evaluate disruptive flows against norms based on general ethical and political principles as well as context-specific purposes and values."

Part of this makes sense. Once we know what the existing norms are and how online stuff may breach those norms, we evaluate what to do.

However, the other part, concerning ‘general ethical and political principles’ . That’s where things get interesting. But let’s put that aside for a moment.

3.3.6 Easier cases

Let’s see how this picture is supposed to work with a couple of easier cases.

3.3.6.1 Privacy in medicine

It’s easy to imagine (or, sadly, recall) situations where someone in a medical office fails to adequately protect patient information. For example, a doctor may fail to close the door when discussing test results; office staff leaving your chart sitting on the counter for any passerby to read.

Notice that this involves not just protecting information about the patient, but also the patient’s questions. Would you ask the same questions with an audience?

Thus if we turn to online medicine —google searches, web md, etc— presumably these norms still apply.

3.3.6.2 Privacy in banking

Suppose you’re a bank executive trying to decide whether to allow online advertising to your customers. The place to start (in addition to the law) is to think about customers expectations of privacy when they bank in person.

Most customers would be concerned if the next person in line stood right behind them as they talked to the teller. Similarly, a teller who shouts “Sir. You have $7.23 in your account. You cannot withdraw $20!” so everyone can hear would quite rightly be rebuked.

That tells us that customers have an expectation of privacy when it comes to exchanging information with bank employees. None of that changes when the transaction happens online. As she writes

> "Whether you transact with your bank online, on the phone, or person-to-person in a branch office, it is not unreasonable to expect that rules governing information will not vary according to medium" [39]

3.3.6.3 Netflix, youtube, etc

Nissenbaum argues that data about online video use should be governed by the same principles as privacy in video rental stores. In particular, she claims that the constraints binding West Coast Video based on the Video Privacy Protection Act of 1988 should apply to online videos as well. [2011; p. 39]

3.3.7 Harder cases

3.3.7.1 Analogs of social media?

However, when we turn to things like social media, it's harder to sort out the relevant analogous in-real-life practices.

What would be the relevant real life context be for a social media site facebook or instagram? The world's most awkward party where your relatives, friends, co-workers, exes, and random people you had a nice conversation with once all mingle?

What about Twitter? As far as I can tell, its nearest analog is an angry mob.

3.3.7.2 2 guidelines for locating contexts in hard cases

So how do we determine the contexts when it's not obvious? Nissenbaum writes

> “Where correspondences are less obvious, such as consulting a search engine to locate material online, we should consider close analogies based not so much on similarity of action but on similarity of function or purpose.” [43]

More broadly, she suggests two guidelines for determining the relevant context when it isn’t clear:

(1) Look at how the company presents itself to users

(2) Look to ends/purposes/values involved and work back to determine the relevant norms

3.3.7.2.1 How the company presents itself

The first guideline is fairly straightforward. We should follow how the online offerings present themselves to the consumer in determining which norms are relevant. Thus if a site presents itself as an online university, the relevant norms are those which govern universities. If a site presents itself as similar to a library, the norms are those which govern libraries

Interestingly, Nissenbaum wants this to apply even if the data has different uses to the consumer and investor facing sides. Thus if the data is collected from the consumer under the guise of being a medical site, it cannot be sold off as an asset.

However, we might wonder about situations where the company is completely upfront about collecting and selling your data. To construct a case that might be a problem, we would need the upfront company to be clashing with established norms from a context. For example, if the medical advice website tweeted all the search questions in real time so that was part of the appeal —you can go to the site for information about fibromyalgia or just to see all the weird stuff people are looking up.

Note also that this is similar to Sax. [todo]

[Sidenote] This might raise some questions about the data for our accounting majors: Set aside for a moment the issues about amortizing the data. [If I’m understanding this correctly] For something to be an asset, it has to be tied to the company's main business operations. If a company provides medical information online in exchange for collecting data from its consumers, which it then sells, is the data an asset?

3.3.7.2.2 Looking to relevant values or purposes

When there is no clear analogy to reason from, we should instead start from ends/purposes/ values and work back from there. This is tricky to sort out.

She's basically saying that if we can't tell what the relevant context is, we should think about what sort of things are at stake in the use of the site (etc) and then go off of the norms that we already have concerning those things. So, for something like Twitter, we might think about values like free expression, public accountability of governments / companies, and the importance of public disagreement in a democracy. From there we would consider what norms are associated with those values. Perhaps, a general presumption in favor of non-interference by governments (or more generally powerful entities) would apply here. This might be in line with 1st amendment law and policy in the US.

Note that the relevant norms can come from law and policy. But they can also be found in commonly held reasonable expectations, i.e., what would an average person expect about privacy in that situation.

There are a variety of ways of testing when something is a reasonable expectation. I often find it helpful to think in terms of complaints. People complain about all sorts of stuff all the time. Sometimes when you hear a complaint, you wonder about the complainer's sanity or what else might be going on in their life. Othertimes, you think they have a point. If the bartender gives you a beer that's 90% foam, no one would bat an eye at your complaint. But if the bartender leaves a tiny bit more foam than usual, people will look askance at you if you start complaining; the bartender would be right to dismiss your complaint as unreasonable.

3.3.7.2.2.1 Problem: does this abandon contexts?

In cases where we're looking to values, purposes, et cetera, we might worry that we've thrown the contextual integrity approach overboard. It seems like her advice in these cases is to give up on the context based framework? Isn't she just saying, yeah, if you can't tell, just give up and do ethics / political philosophy?

Not necessarily. In the worst case scenario for her picture, we're just acknowledging that some situations may in fact be brand new. That doesn't do anything to undermine the usefulness of her approach for many other cases.

Moreover, there's no principled inconsistency to saying 'Look for contexts and use those; if none exist, you're going to have to figure it out based on other tools we already have.' This is still consistent with her overall approach that we should deal with informational privacy using existing tools / norms / expectations, rather than thinking we need something brand new.

3.3.8 Objections

Like any good author. Nissenbaum spends some time answering objections she imagines her view will face. Let's go over a few.

3.3.8.1 Overly conservative

She acknowledges that her view is in tension with the optimism often associated with technological change and 'progress'. As she writes

> One is that by putting forward existing informational norms as benchmarks for privacy protection, we appear to endorse entrenched flows that might be deleterious even in the face of technological means to make things better. Put another way, contextual integrity is conservative in possibly detrimental ways [125]

Note that this isn't (necessarily) 'conservative' in the US political sense associated with the republican party. Instead, the concern is that our laws and policies don't change fast enough to keep up with the pace of technology and new uses of data.

Think of it this way. Suppose someone invents gunpowder for use in fireworks. Laws and policies around gunpowder get made to ensure that fireworks are safe. Then someone realizes that you can put the gunpowder in a metal tube with some rocks on top and point it at people you don't like.[23] Relying on the existing laws and policies around gunpowder will not help you in managing this new use.

Her response [ToDo]

3.3.8.2 Contexts as cement shoes

Another worry is that if our regular life practices around privacy change for the worse, applying contextual integrity to the online world entails dragging down online privacy too.

A second worry is that contextual integrity, being so tied to practice and convention, loses prescriptive value or moral authority. In this era of rapid transformations due to computing and information technologies, changes are thrust upon people and societies frequently without the possibility of careful deliberation over potential harms and benefits, over whether we want or need them. Practices shift almost imperceptibly but, over time, quite dramatically, and in turn bring about shifts in conventional expectations. These changes have influenced outcomes in a number of important cases, such as determining that the Fourth Amendment was not breached when police discovered marijuana plants in a suspect's yard by flying over in a surveillance plane. [126]

3.3.8.3 Commercial nature of the web

[23] By the way, I based this example on a common but completely false history of gunpowder and guns. This source doesn't matter for point of the example. But since this might sound familiar, let me take this opportunity to set things straight. The story goes that the Chinese invented gunpowder but used it for fireworks, never realizing that it could be a potent weapon of war. Europeans eventually learned of gunpowder from the Chinese and had this insight.

This is bullshit. (And, somewhat racist bullshit at that; it has a ring of "oh those poor simple asians weren't smart enough to realize what they had their hands on.") The Chinese made bombs and gun-like things from the start. The reason guns became important to warfare in Europe first was an improvement in the mix of the powder (more potassium nitrate), which probably came about during the transmission to Europe by Arab traders, and somewhat better metallurgy (to make the guns).

She also acknowledges that the commercial nature of most online activity pushes in a very different direction than her contextual integrity approach. As she points out

- Private payment is the overwhelming means of supporting online activity
- The physical and computing infrastructure is privately owned
- Most websites are supported by payment for advertising
- The Net is almost completely privately owned by private, for profit entities

Given these considerations, it would seem that the norms of the competitive, free marketplace are the place to look for a regulatory approach.

Indeed, this is already a live alternative. When the FTC and other regulators approach privacy online, they tend to do so from the usual consumer protection mindset. In the US this often means protecting consumers from unfair business practices and subsuming the protection of personal information as a form of protecting the integrity of commercial transactions.

Nissenbaum's response to this is first to emphasize that the web is not entirely commercial. This is part of her claim about the radical heterogeneity of online activity. While many parts of the web are supported by advertising or subscriptions, that does not mean that the activity people engage in is purely commercial. Sure, some people post to Instagram with the aim of becoming a paid influencer. But many of us just want to share pictures of our friends, pets, or dinner.

More importantly, she borrows from Elizabeth Anderson to point out that many functions in society straddle boundaries between the commercial and noncommercial. The fact that private payment is involved does not require total concession to marketplace norms.

We expect things like education, health care, religion, telecommunication, or transportation to measure up to ideals. The fact that people pay for them is not decisive.

Consider, let’s see, now what might be an example that everyone is sick of from me. Oh. How about higher education. No one serious thinks that higher education is a fully commercial enterprise. While a university requires money and some of that money comes from its students, there are other values at stake. If students were just paying for diplomas, then we would succeed if every student gets a diploma, regardless of what they learn along the way. (Obviously, we want students to graduate and graduation rates are part of how we should assess the effectiveness of college; the point is that there are other considerations at stake too.

The same is true for any professional (doctors, lawyers, athletes). Profit is important. But we expect more from professionals than just doing the minimum of what they are paid for. An emergency room doctor who maximizes the number of patients she sees with no regard for whether they survive is a very very very bad doctor, regardless of how much she contributes to her hospital’s bottom line.

Therefore, she claims, the fact that most online activity is supported by commercial activity does not establish that informational privacy should be handled as a commercial matter.

3.3.8.4 Basing on the real world better not mean the real world

[Todo]

There’s a danger here that we should keep in mind when trying to use Nissenbaum’s theory in a cosmopolitan society where some groups have historically been more powerful than others.

4 Q2 Harms of informational privacy violation

VERSION v.0.1.1

4.1 Intro

We've discussed when an ethical company may profit from the use of personal data. Let's now turn to when things go wrong. Obviously, wrongness comes in degrees. We thus need to know how to distinguish between problematic misuses of personal data and cases where things have gone so wrong that lawmakers, regulators, industry groups, or others are justified in taking action against companies:

(Q2) When is a misuse of personal data significant enough to warrant moral condemnation, regulation, criminalization, or other forms of coercion to prevent?

4.2 Test cases

As we work through discussions of harm and blame, it will help to keep in mind some made up cases in which things go wrong with personal data.

Your mileage may vary. Some of these may seem fairly innocuous. Some may seem creepy but not necessarily harmful. Others may seem straight-up despicable. Take a minute as you go through them to try to put your finger on what is wrong about the ones that seem wrong. More importantly, pay attention to tweaks to the details that flip your reaction from 'meh' to 'oh hell no'. Those are exactly the things we're going to try to uncover in understanding how uses of personal data can harm.

4.2.1 Data hoarder

Data hoarder: Red is a data hoarder. She has an intense and obsessive desire to possess data. However, she doesn't actually care about the contents of the data. She just likes to know that she possesses it. Suppose —contrary to the real world — that she keeps her data in a perfectly secure database which only she can access. She doesn't look in the database; she just cares about having data, not doing anything with it. Indeed, we know that she'll never use it in any way which affects people's choices.

4.2.2 Argument tweeter

Argument Tweeter: A new app which runs on the various 'smart' devices in your home —your Alexa, iPhone, et cetera — detects any time there is an argument in its vicinity. The recording is immediately tweeted out.

Variations

There's nothing identifiable about the clip

The twitter account is unpopular, no one ever hears the clip

Instead of being tweeted, it is stored in a database

Sex rather than arguments

Instead of tweeting them out it stores the audio files in a database [non-identifyably tagged; identifiably tagged]

It creates a hash (irreversibly encryption) of the audio files and blinks out the resulting hashes in morse code on an LED in the middle of the desert.

Note

This raises issues of when the actual violation of privacy occurs. Is it when the clips are accessed? Or when they are collected.

What if it was set up in a way that you know your argumentative style will never be picked up

4.2.3 Internet isolation

Internet Isolation: Indigo connects with friends and family via the social media platform SnapFace. Due to an unforeseen confluence of parameters, the site's algorithm stops showing her posts to others. She doesn't know this. But she does notice a steep slide in likes / comments down to zero. At first it doesn't bother her. But, as time goes on and no one responds to even her best material —her most clever observations, her cutest outfits, the most adorable kitten/puppy/baby pictures— the perceived isolation and shunning start eating at her. She becomes anxious and depressed ..

4.2.4 People Ratings Agencies

People Ratings: Panopticon LLC creates a web of partnerships and data-sharing agreements with every major data-collection company. Through these connections, Panopticon LLC has access to nearly everything each American has ever said or done online, as well as in the vicinity of a camera, and every purchase they've made.

Here are just a few of the revolutionary services they provide:

Predictive Policing: Panopticon offers a service to law enforcement which allows seamless access to the identities of every person within a few blocks of the scene of a crime. This list of people is served up along with a proprietary rating of the likelihood that each person would create a crime.

Employee Screening: Panopticon offers a service to employers with custom ratings of prospective employees in terms of who is likely to be insubordinate, who will not complain about menial work, and other work-related dispositions.

Date Screening: Through a new partnership with Tinder, Grinder, Bumble, and all the other major dating apps, Panopticon offers ratings of prospective partners on everything from conscientiousness, cleanliness, likelihood of cheating, expected number of dates-before-sex, financial status, and other metrics. By offering users the ability to provide post-date/post-breakup ratings, it also provides ratings on kinkiness and sexual performance. [I'll stop here; this is creepy enough]

If these services do not seem problematic to you, feel free to add new details until they strike you as worrisome. All set? Good. Keep that in mind in what follows.

4.3 Methodology

Thus far I've tried to avoid discussing things in terms of privacy. But thinking through our cases, that's no longer going to be avoidable. When we ask what's going wrong in these cases, the irresistible answer is that people's privacy is being violated.

Hopefully at the point in class you're expecting us to dive into getting clear about what privacy is before applying the concept. If so, yay! This is usually a great strategy. Too many confusions and pointless disagreements arise when people think they're arguing about the same thing but are actually talking about different things.

4.3.1 The problem with starting from privacy

But I'm not sure that starting by clarifying the concept of privacy is the best approach here. That's in part because the issues we're dealing with are likely to some extent genuinely new. Thus existing accounts of privacy may or may not be up to the job of dealing with them.[24]

More importantly, our task here is a bit different from what we've done elsewhere. Previously, we dealt with cases where we weren't really sure what's right and wrong. Think back to Holley. We spent a lot of time trying to come up with some principle that could help us determine when it's okay for a salesperson not to divulge information. A lot of the cases we discussed left a lot of room for reasonable disagreement about what's right and wrong.

[24] Obviously, to really establish that we'd need to work through all of the existing accounts and carefully consider their strengths and weaknesses for dealing with our concerns. That's a Ph.D. dissertation; we're not going to do that here.

The way I've constructed our task here is different. We're starting with cases which nearly everyone will agree involve unacceptable activities. If the cases I've come up with don't elicit those intuitions for you, I'd bet that we could tweak them until you'd agree that they involve something wrong. I wouldn't have made that bet in the borderline sales cases from before.

Thus if we're talking about cases which we agree involve something wrong and a concept of privacy can't capture the wrongness, that's a strike against the account of privacy. It doesn't show us how to understand what's wrong in all the cases. Put a different way, no account of privacy is likely to change our minds about the wrongness of these cases. That's why I want to avoid talking directly about the concept of privacy —it's not likely to help us in this task.

4.3.2 Focus on harmful violations of privacy

What's the alternative? Notice that sometimes we don't care very much when privacy is violated. Think of someone accidentally walking in on you in a store changing room; not great, but usually not a huge deal. But we do care a lot when people suffer harms from their privacy is violated.

Thus the alternative is to start by thinking about how to explain how privacy violations can be harms. Approaching in this way, will also give us some resources to draw on. In particular, we can frame the discussion in terms of a well-worked out concept of harm .

If this approach still seems suspect, here's another analogy. Legal scholars sometimes discuss what are called *harmless trespasses*. If you own some land and I take a shortcut across it, leaving no trace, I have still violated your legal right to control who enters your property. But I think most people would agree that prosecuting someone for a harmless trespass should normally be at the very bottom of a prosecutors' priorities. We think the prosecutor should focus first on trespasses that involve harms (e.g., where I chop down your trees and knock down your fence dragging them away).

4.4 The concept of harm

We are going to start with the pretty clear intuition that people are harmed in our cases. This allows us to get some leverage on our problems since harm is a much better worked out concept, at least in some areas of philosophy.

I'm thus going to draw on an influential account of harm from philosophy of law that comes from work by Judy Thompson and Joel Feinberg. Like every view it has problems, some of which I'll mention as we go, but we need not focus on those here.

A bit of warning in advance, I'm going to go into a lot of detail about the concept of harm in what follows. We will need those details when we turn to applying it concept. But don't worry if you don't follow all of them on the first read of this. They are more here for reference so that you can refer back to them when you need them.

4.4.1 Harm principle

To begin, I'm going to stipulate that, on our use, harms and acts which cause harm are always grounds for public action or complaint. If you've ever heard of the 'harm principle' which comes up in discussions of criminal law, that's what's in the background here. The harm principle says something like:

(HP) An activity may be criminalized only if it causes harm to others

For those of you who have heard of the harm principle, that's probably because of the fact that this is a necessary condition. If we accept HP, it follows that actions which do not harm other people cannot be criminalized. Thus, at least in spirit, it comes up often in discussions of drug policy, prostitution, and any given internet list of dumb laws.[25]

Thus our starting point is that , by definition, all harms are bad; it is always wrong to harm others. If this seems trivial to you, just wait. As we will see, it will mean that some things which look like harms actually are not —in some cases, you can kill someone without harming them.

4.4.2 Definition of harm

With that background in place, what is the relevant notion of harm? For the view loosely based on Thompson and Feinberg, I will say that

(H) x is a harm if and only iff x is a wrongful setback of interest.

[25] Note that this doesn't automatically show that any of these things shouldn't be illegal. Drug use often negatively affects those around the user; prostitution is often bad for the prostitutes. It just reframes the question to whether these are harms in the relevant sense and gives us a means of answering that question.

That means there are two parts to a harm. There has to be (1) a setback of interest and (2) it has to be done in a wrongful way. This give us a framework for understanding when a harm is present. We first identify the setback. Then we ask whether it was wrong. If we can do both, there was a harm. If we can do one but not the other, there is no harm.

But before we can do that, we need to get clear on what counts as a setback of interest and the relevant notion of wrongfulness.

4.4.2.1 Setback of interest

Let's start with the notion of a setback of interest. Obviously, we first need to know what interests are.

4.4.2.1.1 Interests

An interest in this sense is anything you have a stake in. If you invest in a company, you now have an interest in it. There are a bunch of ways of thinking about this, but I'm just going to say that:

(I) S has an interest in x only if changes in x can affect S's well-being.

Notice I didn't say something like 'the changes can affect you in a negative/positive way'. That's because of the harm principle lurking in the background. Remember, if something turns out to be a harm, we are allowed to criminalize it (or at least morally condemn it). If I give you a cookie, that positively affects you. If the temperature of the room is a bit colder than you'd like, that negatively affects you. But if we say that the room being a bit chilly is a harm, given the harm principle, that means we may be okay with using the full might of the criminal law to enforce HVAC settings.

4.4.2.1.1.1 Well-being

Well-being is usually something which is more spread out over time. It's the sort of thing we have in mind when we talk about living a happy life. A happy life doesn't mean that you were happy at every single moment. If it did, your happiness hangs by a thin thread; one stubbed toe and your life is ruined. The tricky thing with thinking in terms of well-being is to pick the right time interval for assessing it.

You don't want it to be too long. You don't want it to be assessed over your entire life —the sort of thing you consider on your deathbed in old age when you think back over your entire life and wonder whether it was worth anything.[26] That's because lots of important events will get lost. That breakup in college was really painful and tough; that first promotion was amazing. But from the perspective of your entire life, these probably won't even register.

It also better not be so short that having an unusually bad cup of coffee from the student center on Tuesday morning damages your well-being. The idea is that when we think about how things are going for us, all the minor good and minor bad should normally drop out.

Thus for our purposes, I'm going to stipulate (without argument) that the right time-frame for assessing well-being is about a week:

(WB) x affects S's well-being only if x affects the quality of S's week

That is, things affect your well being when they affect the quality of your week. Imagine that on Friday you grab a drink with a friend who you haven't seen since last Friday. When you ask 'how have you been?' You are expecting to hear about the great date they had, the exam that they aced, the major fight they got into at work. If they start going on about how the cup of coffee they had Tuesday morning was terrible, you'd be pretty surprised or confused.[27] That sort of thing just usually isn't a big enough deal to actually affect how good or bad their week was; it doesn't usually affect their well-being.

Before moving on, let's summarize how we've made progress on our definition of harm. Now that we have a grip on how to assess well-being, we can rewrite (I) as

(I′) S has an <u>interest</u> in x only if changes in x can affect the quality of S's week.

Therefore, we can also rewrite our definition of harm

(H′) x is a harm if and only x is a wrongful setback of the quality of a person's week.

[26] [ToDo] Discuss ancient Greek philosopher's line "say of no man that he's happy until he's dead" to illustrate this?

[27] Assuming that your friend is not a serious coffee snob. I had a good friend in graduate school who took his coffee super seriously (he taught me how to roast green coffee beans). I could imagine one bad coffee really sticking in his craw so much that it would still be on his mind several days later.

Hopefully it seems like we're getting closer to something that we can actually apply to real cases.

4.4.2.1.2 Setbacks

We now know that the relevant interests are things which can affect the quality of your week. Now we need to talk about when an interest is set back. I know. I can see your eyes rolling already. Do philosophers really have to (over)analyze everything? Yes. It's what we do. This is actually interesting, I promise….

I'll start with the definition I'm headed towards, then tell you why we need it:

(S) x is a setback of S's interests if and only if S is worse off than she was or would've been if x hadn't occurred.

Why not just say 'worse off than she was'? That covers most cases. You steal my bike, I am worse off than I was before you stole it.

The answer is death. That's right, death.

Ok, not just death. There are other cases. If I steal your winning lottery ticket and replace it with a loser before you realize it is a winner, you are no worse off than before I stole it. Adding the counterfactual condition 'or would've been' allows us to still say the theft was a harm.

If you'd like to join me for a brief, though relevant, tangent about death and its relevance for the counterfactual, read on, otherwise, you can skip ahead.

4.4.2.1.2.1 Death

We are now to one of my absolute favorite puzzles in all of philosophy: How can your death be bad for you?

This goes all the way back to Epicures in ancient Greece, who used this to argue that it is irrational to fear death. But I'm getting ahead of myself. Here's the puzzle, paraphrasing Epicures

Where I am, death is not; where death is, I am not.

Do you see it? Let's walk through the puzzle. When you are dead, you stop existing. There is no more you.[28] If you still exist, you are only mostly dead (send your friends to Miracle Max, ASAP![29]). So, when you are dead, there is no one there for death to be bad for. Therefore, death cannot be bad for the person who dies.

Reactions to this puzzle differ. Epicures thought it was liberating —you have no reason to fear death. Others, myself included, see it as posing a problem to be solved. You might take a middle ground and say something like, yeah, it's irrational to fear death itself, but perfectly rational to fear dying. However, I think Nagel was right when he wrote in a footnote that "I think I should not fear dying if it weren't followed by death"[30]

4.4.2.1.2.2 Other problems

There are other problems for this whole picture of setbacks of interest. Seana Shiffrin, for example, gives the example of a rich asshole whose hobby is going around and dropping bricks of gold on people's toes and then running off. She calls this jerkwad Richie Rich.

If you've never broken a toe, consider yourself lucky. I've had more than my share (thanks, judo).

[28] NB, if you are religious and believe in an afterlife or if you are a big fan of Pharrell (http://nooneeverreallydies.com/), then in this sense, you believe that no one ever really dies. You would agree with Epicures that it is irrational to fear death; it might of course be rational to fear what happens after death.

[29] Then have fun storming the castle.

[30] [reference] In the very unlikely chance that you were wondering what my favorite footnote in all of philosophy is, now you know.

It is a particular kind of misery. You can still basically walk and do stuff. But you live in constant fear bumping into furniture, stepping wrong on stairs, having an ant crawl over your sock, et cetera. It sucks.[31]

However, at least for me, the misery of a smashed toe would be substantially offset by the $500k which came in the form of the shiny new 400oz gold bar resting on my foot. My going rate for allowing you to smash a (non-big) toe is, conveniently, about $250k.

Thus if Richie Rich drops the bar on my toe, I am not worse off than I was. The bar is worth more than my smashed-toe-rate. And I'm also not worse off than I would've been if he had dropped the bar on someone else's toe. Sure I could walk without wincing. But I'd much rather limp with a half a million. So, on this picture, there would be no setback of interests and therefore no harm.

Fixing our account of setbacks of interest in response to Shiffrin's attacks is very difficult. Trust me, I've tried. So, let's just note that it has difficulties which need fixing and keep moving.

4.4.2.2 Wrongfulness

The second component of harm is that it has to be wrongful. That makes sense when we remember that we are starting from the assumption that every harm is something that we can punish or coercively prevent people from inflicting.

There are many ways to make sense of the wrongfulness criteria. For the most part, any major moral theory can be slotted in here. However, we are going to follow Thomson and Feinberg in understanding the wrongfulness component as a rights violation. Therefore, a harm is a rights violation which makes the victim worse off than they were or would've been. More carefully, we can again rewrite our definition:

(H'') x is a harm if and only x is a rights violation which makes the quality of a person's week worse than it was or would've been.

[31] The terror of coughing /sneezing when you have broken ribs during cold season is far worse. Just FYI

4.4.2.3 Method

Now that we have a solid definition of harm, we can work through the cases we are concerned with to try to determine when people are actually harmed.

Remember, on our notion of harm, there can be bad things which happen to people which do not count as harms. There's both a minimal threshold of severity covering the setback of interests and the requirement that we be able to locate a right that's violated. Therefore for each alleged harm from informational privacy violation, we will ask what interests are setback and which right is violated.

Thus, spelt out in ridiculous detail, here's our method (for negative answers, we stop):

1) Find a seemingly problematic uses of personal information.

2) Determine how this use may negatively affect people (i.e., determine what interests may be affected).

3) If a use could negatively affect people, determine whether it could affect the quality of a person's week (i.e., determine whether it is a an interest in our sense)

4) If a use could affect the quality of a week, check whether it actually does make people worse off than they were or would've been.

5) If we do have a setback of interests in our sense, consider what right might be violated in the setback.

6) If there is a rights violation in the setback, we've found a harm. The company needs to knock it off; we are justified in stopping companies from doing it.

4.4.2.4 Odd cases

It’s worth emphasizing that, on our definition, to have a harm there has to be both the setback and the rights violation. If you only have one but not the other, there is no harm in the relevant sense.

In many cases this is straightforward. If you steal my bike, I am worse off than I were before (now I have to drive; I have to buy a new bike). The theft violated my property right to control what happens to and with my stuff. Thus you’ve harmed me. Easy peasy.

However, in some cases, applying this concept of harm leads to some counterintuitive results. Let’s work through a couple of cases; doing so will let us see how the view works in action.

4.4.2.4.1 Harmless rights violations

Suppose you are sitting at a coffee shop and have just finished reading a newspaper.[32] When your back is turned, I sneak up and steal your newspaper. Is this a harm?

Well, it seems like I violated one of your property rights. You bought the newspaper. Sure, you were intending to toss it in the recycling or leave it for the next patron. But I took it before you did that. So, it looks like what I did was wrongful.

Are you worse off than you were or would’ve been? Probably not. In fact, I might of just saved you a trip all the way to the recycling bin, so you might be better off. Of course, we might think up ways in which your well-being was affected —perhaps you now constantly feel unsafe or your confidence in humanity has been shaken to the core. But at the very least, only some people are going to respond that way. And for most people these won’t be setbacks that rise to the level of affecting your well-being. Therefore, it looks like there was no harm.

4.4.2.4.2 Hurting without harming (self-defense)

[32] Newspapers were these things that everyone used to read. Kind of like the internet but on paper. Just bear with me…..

Suppose that Scarlet kills Violet in genuine self-defense. Does Scarlet harm Violet?

Before we can get to that, I need to clarify what I mean by genuine self-defense since the self-defense defense has been horrifically perverted by legislation in places like Florida.[33] Genuine self-defense means that the defender faced a lethal threat (they were being attacked by someone trying to kill them), the means were proportional (you can't shoot someone threatening to punch you), and they had no escape (if you can retreat you must[34]).

Okay, back to our case. Does Scarlet harm Violet? In the regular English sense of the word 'harm', it's obvious. She kills her. How could that not be a harm? But remember, the question is whether she harms her in our sense; whether she wrongfully setback some of Violets' interests. Let's work through it.

Does Scarlet setback Violet's interests? Yep. Which ones? Plausibly all of them (ignoring the complication around death mentioned above). It's sometimes useful to summarize this by saying that Scarlet hurts Violet.

Does Scarlet violate one or more of Violet's rights? If so, which ones? This gets complicated and there are several possibilities. I'll just mention two ways this can go. You can skip this if you're not interested.

[33] There are plenty of legitimate arguments we can have over exactly what should count as self-defense. My philosophy of law class spends literally half the semester working through them; and I don't want to pretend I know what the answers are. However, I will say the subjective perceptions of the killer cannot be determinative in anything like a just society. (IIRC, the NRA and other backers of 'stand your ground' laws agree and characterize the Florida law/cases as mistakes)

[34] Some jurisdictions carve out the 'castle doctrine' such that you have no duty to retreat if you are attacked in your own home.

If the relevant right is a right to life, then it looks like there has been a rights violation. However, that means there has been a harm. Given the harm principle, that means we would be justified in punishing people who kill in genuine self-defense.[35] Violet teams up with the state to force Scarlet to choose between dying and going to jail. That does not seem plausible.

One common response is to invoke what's called the forfeiture theory. When someone unjustly tries to kill another person, they (temporarily) give up their right to life during the attack. Thus the attacker does not have a right to life that could be violated. Therefore, when Scarlet kills Violet, there's no harm.

There are several problems with the forfeiture theory. It creates complications with third-party assistance (whether the attacker forfeits their right overall or just to the victim affects how we deal with cases where a third person mistakenly thinks someone is being attacked and kills the attacker on their behalf). If also leads to a weird picture of rights. Remember that rights are supposed to be absolute prohibitions on people doing things to you. To see this, consider a wrinkle in how self-defense works:[36] Suppose that when Scarlet pulls out a gun, Violet stops trying to attack. But Scarlet is all pissed off and intends to kill Violet anyway, now Violet can kill Scarlet in genuine self-defense. If we are thinking about what's going on in terms the forfeiture theory of rights, it all looks weird: first Violet has no absolute protection against being killed (which Scarlet has), then Scarlet's right disappears and Violet's reappears. Absolute no longer means absolute.

An alternative is just to be more careful about what the relevant right actually is. You could say that a 'right to life' is just a convenient shorthand. What we really have is a right not to be killed unjustly. If the relevant right is a right to not be killed unjustly, then we ask when is it justifiable to kill. The answer will include what I said above in describing genuine self-defense. In other words, on this version of the right, self-defense is one of the built in exceptions to the prohibition on killing. Thus on this picture too, there will no rights violation when Scarlet kills Violet and therefore no harm.

4.5 Informational harms -- setbacks

Let's consider some of the ways personal data can be used to setback people's interests. It's worth

[35] Actually, this doesn't have to be true. There are other kinds of defenses under the criminal law that could get her off the hook. But that's too far afield.

[36] Yes, there actually have been cases like this.

going into some detail about these so that we can fully understand what the threats out there actually are, and what future threats may involve. We'll then turn to whether these setbacks involve rights violations.

Before we jump in, note that we should keep in mind that many of these practices on their own are not sufficient to setback interests, but they may combine in ways which do. See, for example, AT&T's reported digital strategy: https://www.theverge.com/2019/5/22/18635674/att-location-ad-tracking-data-collection-privacy-nightmare

4.5.1 Exploiting weaknesses

We all have bad habits and psychological weaknesses. Sometimes, we try really hard to hide them from others. The fact that these habits and vulnerabilities may be present in personal data about us or inferred from such data creates a significant opportunity for others to exploit them in ways which may significantly set back our interests.

Barocas and Nissenbaum point to some dimensions of this in discussing the way that targeted advertising online

> might not only lock individuals into past habitual choices from which they would like to escape, but may open them to manipulation and illegitimate control by others. If someone can identify your weaknesses and vulnerabilities by closely monitoring past behaviors and dispositions, that person may be able to shape your choices, actions, transactions, and purchasing decisions in ways that do not accord with principles and purposes to which you are committed. Even if you succeed, in your deliberate actions, to stay true to these purposes and principles, others may have their own reasons for targeting your weaknesses, prejudices, or vulnerabilities, and, thereby undermining your autonomy. {Barocas:2009ws}

Now, it will matter how significant these weaknesses are and what behaviors they can get you to do. Remember, for something to be a harm, the setback has to affect your well-being. Take a minor bad habits like spending more time than you'd like on sites dedicated to celebrity gossip. If an online advertiser cunningly lures you back to such sites, that does set back your interests. But if it's just a few more minutes a day, the setback may not rise to the appropriate level for it to genuinely count as a harm.

That said, we should be sensitive to cumulative effects. We can imagine cases in which each manipulative advertiser succeeds in stealing a few minutes a week from you. On its own, not enough for a harm. But together, you are wasting a couple of hours a week on sites which you don't really want to be visiting. That would potentially be a harm. (If you were stuck for 2 hours at the DMV last week, your friend asking you about your week would not be surprised to hear about it).

[ToDo] We might also be concerned about ads which manipulate by targeting emotions.[37]

4.5.2 Location

The list of interests potentially threatened by public availability of information about your location is likely long.

In some very extreme cases, the interests at stake can be the most important: life, security of person, freedom from fear. It should not be a surprise that the easy availability of personal location data enables stalking and other execrable behavior. The actress Rebecca Schaeffer was murdered in 1989 by a killer who got her address from the DMV.[38] Home addresses can be fairly easy to find online. Homes have doors and locks. Your current location is far more sensitive information.

The good news is that, for an individual, it is very difficult to find someone else's current location if they do not want to to be found. But the knowledge companies have of your current location is a different story.

Sometimes, if we know the purposes for which a company is tracking location, it is easy to identify the interests at stake. Suppose you have 7 good friends who each have a birthday party

[37] https://www.adweek.com/digital/why-media-buyers-are-mixed-on-publishers-selling-ads-based-on-emotion/

[38] See Margan v. Niles, 250 F. Supp. 2d 63, 68 (N.D.N.Y. 2003). Passage of the Driver's
Privacy Protection Act (DPPA), 18 U.S.C. §§ 2721–2725 (2000), followed shortly thereafter in 1994. Margan, 250 F. Supp. 2d at 68–69.
[Refs from Nissenbaum]

at a bar one night, but all in the same week. If you knew your employer or life insurance company would see that you've gone to a bar every night, you might hesitate to celebrate with your friends (even if you don't drink). Presumably, your interest in spending time with your friends is threatened here, as well as your interest in keeping your job, and your interest in being able to care for your family if you die. Those all seem pretty significant.

It's thus worth spending a bit of time on the details of how and when our current location can be tracked by companies. In what follows, I will mostly just describe how some of this works. Thus I want you to think through exactly what ways a kind of tracking might threaten a person's interests in each case. When you locate an interest, then try to think about whether it would affect well-being (i.e., the quality of someone's week). I started this section with some obvious harms. Your harm-antennae are probably raised and very sensitive right now. Thus let me remind you that there are lots of setbacks of interests which don't affect well-being.

Harm-antennae adjusted? Let's talk about how companies find your location and what they might do with it.

For companies, the difficulty of finding your location if you do not want to be found varies. There probably aren't too many companies who can access any given person's current location; probably only your cell-carrier is able to this in real time.[39] Though historical data about a person's movements is normally just as good; we are far more creatures of habit than we realize.

How do they get it? There is a bit of facial recognition floating around out there, but it's not usually a big source of location data. Your credit card company knows what you've bought and where. This also usually isn't a big source of data.

[39] There have been reports of cell-phone carriers selling location data to bounty hunters; the backlash led to the companies vowing to tighten up their systems. https://www.vice.com/en_us/article/nepxbz/i-gave-a-bounty-hunter-300-dollars-located-phone-microbilt-zumigo-tmobile https://www.theverge.com/2019/2/6/18214667/att-t-mobile-sprint-location-tracking-data-bounty-hunters

Usually, it's our devices.

4.5.2.1 Car

Your car's current location is probably in the database (or can be inferred from what's in there) of a vendor which provides Automated License Plate Readers to police and repo-men.[40]

Interestingly, many cars leak their location in unexpected ways. The wheels on modern cars report tire pressure via uniquely identifiable bluetooth signals. This matters since researchers have identified bluetooth signal harvesting devices on roads.[41] Here's the abstract from Grant Bugher's presentation 'Detecting Bluetooth Surveillance Systems' at DefCon22

> Departments of Transportation around the United States have deployed "little white boxes" -- Bluetooth detectors used to monitor traffic speeds and activity. While they're supposedly anonymous, they detect a nearly-unique ID from every car, phone, and PC that passes by. In this presentation, I explore the documentation on these surveillance systems and their capabilities, then build a Bluetooth detector, analyzer, and spoofer with less than $200 of open-source hardware and software. Finally, I turn my own surveillance system on the DOT's and try to detect and map the detectors.[42]

4.5.2.2 Phone

Let's start with your cell phone itself. Obviously, your cell carrier always knows where your phone is located, otherwise it wouldn't be able to direct calls/texts to you. As long as 3 towers can get your phone's signal, your location can be pretty exactly determined.

If your phone pairs with public wifi signals, your movements can be tracked with a great deal of precision, likely even your position inside a store, by whoever is operating the service.

[40] https://www.eff.org/pages/automated-license-plate-readers-alpr

[41] https://link.springer.com/chapter/10.1007/978-3-662-45317-9_3
https://www.schneier.com/blog/archives/2008/04/tracking_vehicl.html

[42] https://www.youtube.com/watch?v=85uwy0ACJJw

4.5.2.2.1 Apps

The apps you install on your phone are a different story. While phone makers have gotten much stricter about requiring the user to explicitly authorize an app to access their location, there are still a lot of ways apps reveal your location to companies.

Actually, before reading on, get out your phone and update your settings to make sure that whenever possible you only give apps permission to access your location while you are using the app.

Settings updated? Okay. Keep reading, you'll thank me.

[ToDo]

Summary of location tracking apps

https://www.vox.com/the-goods/2018/12/11/18136361/location-tracking-data-ad-targeting-facebook-google-amazon

New York Times feature: "Your apps know where you were last night, and they're not keeping it secret"

https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html

- 'Spouseware'

- Strava run data gives away locations of US military bases

https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

4.5.2.3 Social media

Many of us frequently divulge our current location through social media posts. And, even when we do not directly post to tell the world that we are at the Starbucks on Reseda right now, we often reveal locations accidentally. Many phones and cameras automatically tag the picture with the GPS coordinates at which it was taken. Some platforms like Facebook now remove much of this metadata when photos are uploaded. But this is not universally the case.

Given tools like Google's reverse image search, it is fairly trivial to figure out where a picture was taken.[43] You upload or provide a link to the image and the service finds other pictures likely taken in the same location. Check those out and you'll likely find someone who has included metadata or helpfully explained where it was taken.

Indeed, innocuous things like Google Street View (at least now that they blur out people[44], dogs, and cows[45]) can be used in determining location. Recently, a Japanese actress was attacked after a

[43] https://support.google.com/websearch/answer/1325808?p=ws_images_searchbyimagetooltip&visit_id=637096219113096997-2863741639&rd=1

[44] https://www.cnet.com/news/google-begins-blurring-faces-in-street-view/

[45] https://slate.com/technology/2016/09/google-street-view-respects-cows-privacy-and-blurs-its-face.html
https://www.express.co.uk/travel/articles/1085416/google-maps-street-view-identity-dog-blur-funny-photo

stalker found her location by examining high resolution pictures she posted on social media. The resolution was good enough that he used the reflections in her pupils and Google Street View to find where she lived.[46]

4.5.3 Threats to autonomy

At this point in class, you're (hopefully) sensitive to concerns about autonomy. Thus the role privacy plays in providing the conditions for autonomy leaps fairly quickly to mind. Indeed, this connection lurks in the background for several other ways invasions of privacy can set back your interests.

Nissenbaum helpfully summarizes some of these connections

> freedom from scrutiny and zones of "relative insularity" are necessary conditions for formulating goals, values, conceptions of self, and principles of action because they provide venues in which people are free to experiment, act, and decide without giving account to others or being fearful of retribution. Uninhibited by what others might say, how they will react, and how they will judge, unhindered by the constraints and expectations of tradition and convention, people are freer to formulate for themselves the reasons behind significant life choices, preferences, and commitments….. autonomy touches many dimensions of peoples' lives, including tastes, behaviors, beliefs, preferences, moral commitments, associations, decisions, and choices that define who we are. [From nissenbaum 2004 p.130] a

Thus invasions of privacy can set back a person's interests in being autonomous. Those are deeply important interests indeed since they underpin all sorts of other things which are important. Mess with my autonomy, you mess with my week. Thus this looks like a basis for potential harms.

4.5.4 Relationships

One theme throughout our discussion of personal data concerns how the information we share with others is deeply intertwined with our relationship to them. Nissenbaum summarizes some of these connections

[46] https://www.usatoday.com/story/news/world/2019/10/11/japan-man-arrested-stalking-pop-star-using-photos-eyes-pupils/3942667002/

> Information is a key factor in the relationships we have and form with others.… controlling who has access to personal information about ourselves is a necessary condition for friendship, intimacy, and trust….[D]istinctive relationships, for example individual to spouse, boss, friend, colleague, priest, teacher, therapist, hairdresser, and so on, are partially defined by distinctive patterns of information [From nissenbaum 2004 p.130]

Thus our interests in maintaining the integrity of our relationships may be set back by violations of privacy.

We will come back to this later when we discuss which rights are violated since on some accounts our privacy rights are founded in our ability to form and maintain relationships.

4.5.5 Democracy

Insofar as democracy is valuable and people have interests tied up with living in a political system which is based in and responsive to the needs of its citizens, invasions of privacy can set back these interests. Nissenbaum summarizes some of these concerns

> privacy is essential to nourishing and promoting the values of a liberal, democratic, political, and social order by arguing that the vitality of democracy depends not…on the concrete protection against public scrutiny of certain spheres of decision-making, including but not limited to the voting booth. Privacy is a necessary condition for construction of… "social personae," which serves not only to alleviate complex role demands on individuals, but to facilitate a smoother transactional space for the many routine interactions that contribute to social welfare. Similar arguments…[defend] robust protections of medical information on grounds that individuals would then be more likely both to seek medical care and agree to participate in medical research. In turn, this would improve overall public health as well as social welfare [From nissenbaum 2004 p.132]

4.5.6 Surveillance capitalism

[ToDo]

‘Surveillance capitalism’ coined by Shoshana Zuboff

https://promarket.org/road-to-digital-serfdom-surveillance-capitalism-visible-hand/

https://www.theverge.com/2019/5/22/18635674/att-location-ad-tracking-data-collection-privacy-nightmare

4.6 Informational harms -- rights violations

Some uses of personal information may threaten a wide range of interests: life, health, employment, property, freedom from embarrassment, autonomy, equal treatment, and many others. In many of these cases, the setback may be significant enough that it would make sense for someone to bring it up in response to the question “How was your week?” Therefore,there are at least some cases in which the use of people’s personal information leads to setbacks of interest in the sense used by our understanding of harm.

We don’t yet know whether these cases involve harms. For that, we need to know what rights are violated.

4.6.1 Which rights?

To determine what rights could be at stake, let’s distinguish between rights which are specific to privacy and rights that apply to other phenomena.

4.6.1.1 Non privacy rights

The rights violated might not have anything to do directly with informational privacy. Many of the rights threatened by ordinary crimes might be at stake in our cases.

As we've seen, some personal data could be used to enable stalking, intimidation, or other crimes. Those cases pretty clearly threaten a person's right to not be killed unjustly along with other rights to security of person.

When your personal data is used to fraudulently or coercively take away your money or your property, this looks like a straightforward violation of your property rights. For example, if I pay you off so you won't reveal my Netflix history to my pretentious art-film-snob friends[47], you have violated my property right to the blackmail money.

4.6.1.2 Right to protected sphere

In some cases, we won't be able to find one of these commonplace rights violations. For example, widespread surveillance of internet traffic probably won't violate any of your property rights or threaten your personal security.[48] Thus we need to look for rights which have to do directly with privacy.

If the idea of rights specific to privacy strike you as implausible, you're in good company. Some writers such as Judy Thompson maintain that all alleged privacy rights are derivative from more familiar rights. That is, when you look closely at an alleged right to privacy, you just find clusters of rights to life, security, property, et cetera.[49]

However, it's worth noting that in American law we do recognize legal privacy rights on their own. Under the common law, there is commonly (hah!) held to be a right to informational privacy. This is sometimes referred to as tort privacy. There is a tort of privacy invasion which applies when the defendant has intruded 'into [plantiff's] private affairs' or '[p]ublic disclosure of embarrassing private facts about the plaintiff.'[50]

[47] Actually, go right ahead. They already know me.

[48] Note this assertion depends on who you are. If you are a member of a group which the government has a history of targeting, such harms may be more likely. Still, it seems likely that the probability that any specific person in the group will be negatively affected by the surveillance is low.

[49] {Thomson:1975dt}. For criticisms see
{Inness:1992wq}, {Scanlon:hv}

[50] William L. Prosser, Privacy, 48 CAL. L. REV. 383, 389 (1960).

There is a ton of controversy over the details of such rights and over what they cover. We will talk about what 3 families of view which attempt to explain privacy rights and what they cover in what follows.

4.6.2 Accounts of right to protected sphere

Let's set aside the difficult of identifying the precise rights at stake. Pretty much any view is going to suppose a protected sphere of information covered by a right to privacy.

Before we jump in, we can simplify things a bit by noting that they all entail the existence of a protected sphere of information. We see this all the way back in the famous paper which kicked off the discussion of privacy. Warren and Brandeis were explicitly concerned with protecting information about the 'private life, habits, acts, and relations of an individual'[51] This means that people have a right to determine how and when information within that sphere is communicated to others.[52] That is,

(PSI) If x is information within S's protected sphere, S has the moral right to determine :
(a) who knows about x; and,
(b) how, when, where, and with whom x is shared

Therefore, if someone not authorized by S (intentionally) learns of x, they violate her rights. Same thing if they transmit x to others. I will sometimes call these rights violations trespassing the sphere. When such trespasses setback interests, there will be a harm.

Of course, I haven't yet said anything about what forms of information lie within the protected sphere. We can make some progress by considering how to justify the necessity of such a sphere. Let's explore 3 competing accounts which give different justifications of the protected sphere:

[51] {Warren:1890ct} 216.

[52] Muller, p.14

accounts based in liberal concerns about non-interference, accounts based in republican concerns about non-domination, and accounts that expand beyond the individual to consider concerns around relationships.

4.6.2.1 Negative Liberty account

Negative liberty accounts are concerned with freedom from interference by others. Thus very roughly, information will be in the protected sphere if it's public availability would enable or constitute interference by others with her ability to live the life she wants to lead.

To get a better sense of what this means, lets back up a bit and do a little political philosophy.

Why is this called 'negative liberty'? The distinction between between negative and positive conceptions of liberty goes back to a famous paper by Isaiah Berlin. Very roughly, if we ask 'When is a person free?', on the negative liberty account we get the answer: When others do not interfere with her choices. On the positive liberty account, we get: When she is good.

Here's what Berlin says about negative liberty

> I am normally said to be free to the degree to which no man or body of men interferes with my activity. Political liberty in this sense is simply the area within which a man can act unobstructed by others. If I am prevented by others from doing what I could otherwise do, I am to that degree unfree; and if this area is contracted by other men beyond a certain minimum, I can be described as being coerced, or, it may be, enslaved. Coercion is not, however, a term that covers every form of inability. If I say that I am unable to jump more than ten feet in the air, or cannot read because I am blind…it would be eccentric to say that I am to that degree enslaved or coerced. Coercion implies the deliberate interference of other human beings within the area in which I could otherwise act. You lack political liberty or freedom only if you are prevented from attaining a goal by other human beings (Berlin, 1969: 122).

I will sometimes refer to this account as the liberal account. This is a standard category of view in political philosophy.[53] It is not necessarily 'liberal' as the term gets used in American political discourse. Where there are connections between this sort of liberal political philosophy, they exist because some members of the American Democratic party are attracted to these views; not the other way around.

Let's spell out this picture in more detail and see how it fits with the idea of a protected sphere of information. Since the negative liberty picture likely strikes most American readers as obvious, we'll start with positive liberty to make it easier to see what the negative liberty picture implies.

4.6.2.1.1 Positive liberty

On the positive liberty account, the idea of freedom is perfectionist[54]. To be truly free, you need to have certain kinds of personal development and probably certain beliefs. Oftentimes, you need to be 'awakened' or 'enlightened' about the social forces acting on you before you can be free.

One place, indeed the original place for Berlin, this can be found is a broadly Marxist picture. For Marx, in a capitalist society workers are told by all sorts of social forces that the owners of the factory are supposed to be the owners and they are supposed to be just a cog; their labor belongs to those who pay for it. The Marxist thought is that you have to be freed from this false conception of the world before you can be truly free. Something like this is built into the positive liberty conception.

Alternatively, suppose you grow up in a society with a heavy caste system. The lives open to you are circumscribed from birth. You can be a tanner, a nightsoil collector, or a butcher. Those are your only choices. You can choose freely among them; no one will make you be a butcher when you want to be a tanner. You might think your choice to be a tanner is completely free. But it's

[53] https://plato.stanford.edu/entries/liberalism/

NB, some writers take liberalism as the name of a family of views which *includes* both the negative liberty and neo-republican accounts.

[54] Perfectionism in ethics is concerned with things like achievement and excellence. For example, on other views, an athlete setting a world record might be praiseworthy because it makes fans excited. But the perfectionist will claim that achieving someone no one else ever has is in itself valuable.

not. There are other careers that you could've pursued —actor, lawyer, fishmonger— but for the caste system. Your choice isn't truly free unless you recognize that the society has trapped you into those choices (and once you've recognized this, you probably need to overthrow the caste system to be fully free).

This picture means that we can be really heavy-handed in breaking you out of your false beliefs. Forceable reeducation, on this picture, would be freedom-enhancing; even if we have to torture you into enlightenment. For these reasons, Berlin believes we should reject systems which aim at positive liberty

> It is one thing to say that I may be coerced for my own good which I am too blind to see: this may, on occasion, be for my benefit … . [But] it is another to say that if it is my good, then I am not being coerced, for I have willed it, whether I know this or not, and am free (or 'truly' free) even when my poor earthly body and foolish mind bitterly reject it, and struggle against those who seek however benevolently to impose it … . (1969, 134)

That said, I mention this mainly for completeness and to illustrate the negative account. We won't talk much about positive liberty in relation to privacy. Though it should be clear that there's not going to be much presumption in favor of privacy, for those who haven't been properly situated / trained and don't have the conditions of positive liberty.

4.6.2.1.2 Negative liberty

The negative liberty view doesn't require anything of the individual's beliefs or self-conception. It just requires that others don't prevent her from doing things that she wants to do.

This isn't saying that all restrictions on action are wrong. When we fully flesh out a negative liberty approach, we will specify that some interferences do not diminish freedom. A fuller characterization of freedom on this view might be something like:

> You are free when you are able to live the sort of life that seems worth living to you compatible with equal liberty for all.

Thus if what you really want to do is murder children, you won't be able to complain that your freedom is restricted since the murdered children would be deprived of the liberty to live their lives. But where your actions do not diminish the liberty of others, preventing you from doing what you want is morally problematic.

4.6.2.1.3 Right to protected sphere

Thinking about privacy from the negative liberty perspective puts an emphasis on the ability to exclude others from accessing certain kinds of information because such access interferes with your ability to live the life you want. To be free, you must be able to, among other things,

- Keep other people from interfering with your choices
- Preserve available options for your choices
- Choose in accordance with your genuine preferences
- Retain control over how you present yourself to others
- Control much others know about our private lives

In general, if the privacy of some piece of information affects whether you can make choices on the basis of what you genuinely care about, that information will be within your protected sphere. Thus

(RPS-Lib) If the possession or use of information x about S by V interferes or has a substantial chance of interfering with S's choices, then x is in the protected sphere.

In some cases, the misuse of personal information directly interferes with someone's choices. These will be the central cases the liberal view is concerned with. For example, think of a gay person living in a homophobic society. Their ability to conduct their romantic and sexual life in accordance with their values will be restricted if someone 'outs' them. Thus disclosing someone else's sexual preferences in that sort of society will trespass the protected sphere. Since our romantic and sexual lives are very important, that would setback an important interest. Therefore, they would be harmed.

The 'substantial chance of interfering' is going to be trickier since the negative liberty view is very much focused on *actual* interference. Still, as long as the chance of interference is high enough, this will probably be okay. For example, in the law we often make certain things crimes when they are very likely to lead to a harm. Think of possessing burglary tools. We don't want the police to have to wait until the burglary has occurred to arrest someone sneaking around a house at night. Here's part of the relevant California statute dealing with, ahem, lock picks[55]

> California Penal Code - §s 466-469. Burglarious And Larcenous Instruments And Deadly Weapons. § 466. Every person having upon him or her...a picklock...or other instrument or tool with intent feloniously to break or enter into any building...or who shall knowingly make or alter...any key or other instrument...so that the same will fit or open the lock of a building...without being requested to do so by some person having the right to open the same, or who shall make...any instrument or thing...to be used in committing a misdemeanor or felony, is guilty of a misdemeanor.

Note that the 'substantial probability' in our RPS-Lib is built into this example, since, generally speaking crimes like possessing burglary tools require evidence of intent. In most cases possession of a crowbar is perfectly legal. If you are sneaking around a building at night with a crowbar, you may be charged with possessing burglary tools.

4.6.2.1.4 Objections

Let's consider some problems for applying the liberal account to the protected sphere of information. These will all be fairly high-level/ abstract concerns; we shouldn't treat them as in any way decisive, at least until we've considered exactly how they would play out in specific cases.

4.6.2.1.4.1 Self-censorship

Things get tricky for the liberal when the means by which your choices are interfered with involve no overt actions by others, but rather the perception of a threat that they will interfere. In many cases, the interference can turn heavily on your own self-censorship. If you know that all of your internet browsing history will be publicly available, you probably will not use the internet the way you would if others weren't watching. You might not watch the frivolous videos that help you relax after work. You might not google for embarrassing health information. You might not use the internet to help you explore your sexual identity. All of these would be significant interferences with your ability to live the kind of life you want to live, though no one else actively and intentionally prevented you from doing these.

[55] For all state laws concerning lockpicks see https://toool.us/laws.html

There are moves available for the liberal to reconcile this tension. At the least, there seems to be a big difference between situations where others try to get you to self-censor and situations where you just do it on your own. But that distinction is going to be pretty hard to pin down in a meaningful way. If a public official tells people they they 'need to be careful with their searches lest they land under scrutiny', then, sure, the self-censorship that promotes plausibly trespasses the right to the protected sphere.

But the more likely cases involve people being fearful of disclosure and scrutiny. Once we allow people's subjective perceptions of threats of scrutiny to count, we will be in trouble. We won't want to to say that weird paranoias (e.g., "The crows are corporate spies, so I will never go to a doctor's office with crows nearby") should count. But drawing a line will be very difficult. Even if we distinguish between topics of unreasonable paranoia and reasonable concerns (spy crows are unreasonable; being watched online is reasonable), within a topic there will be lots of variation. I have a relative who is incredibly paranoid about her information being stolen online (which, given her age, is good). But while she retains a visceral fear of online banking (probably the safest thing you can do online[56]), she will happily give over information to a random person calling her from the "[local cable company] computer repair service".

4.6.2.1.4.2 Focus on interference

Since the liberal account focuses on actual interference with people's choices, its resources to handle trespasses into the protected sphere which do not cause interference are limited. In general, as long as the data is not used to *interfere* with individual liberties, it is hard to explain what's worrisome.[57]

Indeed, on this approach, we lose sight of anything about the data itself. How concerned we should be about collection of some data doesn't depend on what the data is per se, it depends on how likely it is that it will be used to interfere with people's liberties.

For understanding how this view differs from the others, let me emphasize that the focus here is

[56] Assuming you aren't using public wifi. If you absolutely must bank while connected at starbucks, please use a VPN (or, at least, TOR).

[57] Much of the discussion here follows Stahl [refs]

on actual non-interference. If people could interfere with your choices but do not, there's no problem for your liberty. Again, if your knowledge of their ability to interfere inhibits you from acting, you're still having your liberty restricted on this view.

Consider our example of the Data hoarder. Given the assumption that the database is perfectly secure and that she will never have the urge to look at your data within it, her collecting the data doesn't seem to interfere with your choices. Thus the emphasis on actual interference for the negative liberty account seems to lack the resources to say that her collecting and storing all your information actually trespasses into your protected sphere of information. It's only if she looks, that your right to a protected sphere of information would be violated.

Let's try to bring this tension out further by imagining how it would handle government surveillance of the internet.

Many people are concerned that government programs run by 3-letter agencies (e.g., NSA, CIA, FBI, DIA) which suck up internet traffic violate privacy. But suppose, contrary to fact, that the laws and regulations around this practice prevent the information gathered from being used against any law-abiding person. If it really was true —use your imagination here— that this practice only affected the liberties of those plotting or perpetrating crimes, it looks like the liberal view wouldn't provide any basis for complaint.

Now the liberal has some moves available. If we stick to the abstract, the 'chilling effect', where the mere knowledge that you're being watched affects what you do, may help show there's something to complain about.

Or, if we move out of the abstract to the real world, we will remove all the assumptions that this surveillance will only ever be used to target actual criminals. History gives us no reason to think that this is or will be the case. It flies in the face of how these programs have been used in the last couple of decades. For just two examples, consider the suspicionless surveillance of American

Muslim communities[58] and the interoffice sharing of recorded phone sex sessions between American servicemembers stationed abroad and their loved ones stateside[59].

4.6.2.1.4.3 Intrinsic wrongness of privacy invasion

Another concern about the liberal account is that it cannot account for the intrinsic wrongness of privacy invasion. Many people have the intuition that our Data hoarder is doing something wrong.

To be clear, I'm not talking about criticisms of her character or her obsession per se. The claim isn't that she would be a wierdo. The claim is that she does something wrong, just by acquiring people's personal data. When we talk about 'intrinsic wrongness', we mean that the action is wrong completely independently of its effects or anything else.[60]

Again, because of the focus on actual (or likely) interference, the liberal account has trouble explaining what this wrongness consists in.

4.6.2.2 Republican account

Where the negative liberty view focuses on the danger interference with one's choices poses for autonomy, the republican view focuses on the existence of certain power relations.[61]

As Petit, one of the main modern proponents writes:

> The contrary of the *liber*, or free, person in Roman, republican usage was the *servus*, or slave, and up to at least the beginning of the last century, the dominant connotation of freedom, emphasized in the long republican tradition, was not

[58] [ToDo] [Refs]

[59] [ToDo] [Refs]

[60] If you find the notion of intrinsic wrongness slippery and hard to get the hang of, I'm right there with you. And I wrote a dissertation on it —I wanted to understand intrinsic badness by focusing on the most obvious case, pain.

[61] https://plato.stanford.edu/entries/republicanism/

> having to live in servitude to another: not being subject to the arbitrary power of another. (Pettit, 1996: 576)

For the republican, the mere existence of a domination relationship is objectionable. Where the negative liberty view has trouble explaining what is problematic in cases where no actual interference is likely, the republican can complain that as long as the option exists, it is unacceptable because people are not on an equal footing.

Two quick notes about the name. First, this view often gets referred to as neo-republicanism. That's because O.G. Republicanism has some unfavorable associations. From the Stanford Encyclopedia of Philosophy entry on republicanism

> One reason many people remain skeptical has to do with the fact that the classical republican writings often express views that are decidedly elitist, patriarchal, and militaristic. How could the basis for an appealing contemporary political program be found in such writings (Goldsmith 2000; Maddox 2002; Goodin 2003; McCormick 2003)?

Since we haven't talked about the classical views, I assume it won't confuse folks if I save letters by leaving off the 'neo-'.

Second, this is not 'republican' in the sense of the American Republican party. There may be some ideological connections, but where they exist, it's because some members of the Republican party are attracted to this position in political philosophy, not the other way around.

Indeed, my sense is that American political alignments cut across liberal and republican political philosophies. You can find threads of liberal political philosophy deeply woven into the commitments of some American Republicans and threads of republican political philosophy deeply woven into the commitments of some American Democrats.

4.6.2.2.1 Domination

For the republican, the key concept is domination. Let's understand domination as:

> S is dominated by V if V has the option of interfering with S's choices, regardless of whether the option is taken or not

Thus for the republican, a person is free when no one dominates her. That means the republican has very strong grounds to object to others occupying positions of power above her. It also implies —though this gets less attention— that the republican should think it is wrong for her to place herself into a position of power over others; sauce for the goose is sauce for the gander and all that.[62]

It might help to say how freedom from domination contrasts with the non-interference on the liberal view. From the SEP:

> [On] the non-interference view of liberty….we are committed to saying that the slaves of our well-meaning master enjoy *greater* freedom than the slaves of an abusive master down the road. Of course, the former slaves are better off in some respect than the latter, but do we really want to say that they are *more free*? For another, consider the slave who, over time, comes to understand his master's psychological dispositions better and better. Taking advantage of this improved insight, he manages to keep on his master's good side, and is consequently interfered with less and less. Thus, on the non-interference view of liberty, we are committed to saying that his freedom is increasing over time. Again, while it is clear that the slave's greater psychological insight improves his well-being in some respect, do we really want to say that it increases his *freedom* specifically?

This provides in some ways a more pure conception of freedom —where the negative liberty view is concerned with actual interference, this view gives grounds for complaint whenever it is possible that someone interferes with her choices.

4.6.2.2.2 Right to protected sphere

Thus we can see that on the republican account, the right to a protected sphere of information is grounded in concerns about preventing the existence of domination relationships. We can say that:

[62] A gander is a male goose. A goose is a female goose. Seriously, English, you're drunk. Go home.

(RPS-Rep) Information x about S is within S’s protected sphere if the possession or use of x by another party V would constitute V dominating S.

Obviously, like everything else, domination comes in degrees. But within our framework, we can just say that every trespass into the protected sphere is a rights violation. The degree of harm can then be assessed by the importance of the interest set back and the degree to which it is set back.

The fact that a government or company holding personal information is in position to use it in ways which might harm people means that the government or company dominates them.

Thus it looks like the republican has more resources to complain about violations of privacy independent of their effects, which, we saw, was the weakness of the liberal emphasis on actual interference. Indeed, a serious strength of this picture is that concerns about invasion of protected sphere of information need not track probability of interference.

4.6.2.2.3 Objections

While the republican view is able to better handle the cases where we are concerned in the absence of interference, it does not escape the problems entirely.

4.6.2.2.3.1 Is collection always domination?

One complication for the republican view is that in many cases individual collections of your information don’t put a company / person in a position to dominate you.

That may be sometimes true, as when it concerns something that might embarrass you. But the challenge posed by machine learning is that in the aggregate, patterns can be extracted from pieces of seemingly innocuous data. Those patterns may establish a domination relationship.

What should the republican say about this? Is it that the individual collections do not violate privacy —trespass the protected sphere— only the aggregation of the data?

More generally, the high-power notion of freedom held by the republican poses the danger of going too far.[63] How could we tell what information potentially creates a dominance relationship? Presumably, the republican doesn't think all information collection trespasses the protected sphere. That would entail believing that it is okay to prevent virtually all uses of information technology —not just computers, dusty old pen and paper records would count too.[64]

4.6.2.2.3.2 Intrinsic wrongness of privacy invasion

The republican conception of domination doesn't completely escape the concerns Data hoarder posed for the liberal. If it really is true that our hoarder will never, ever, ever use the data, and no one else can access it, can we really say that she is in a position of domination? It still seems like the probability of use is still playing a decisive role in determining the wrongness.

The republican might point out that as long as it is possible for her to change her mind, she still dominates those whose data she possess. But what if she designs the database software such that it is impossible to ever read the data?[65] In that case, the republican joins the liberal in being unable to explain what's wrong with the collection of data.

[63] Here's an analogy that may help explain what I mean by the notion being dangerously high-powered: I acquired from my dad the destructive impulse to use a power tool whenever possible. I have ruined countless projects and hours of work by pulling out the grinder when I should've used a file.

[64] There are anarchists who reject all social institutions which can possibly exert power over others. This is a view that goes back to Bakunin (https://plato.stanford.edu/entries/anarchism/), among others. Let's just acknowledge that this is a possible view and set it aside. 'Burn all modern society down' is an answer to every political question; it isn't specific to the issues we're trying to sort out.

[65] Databases normally support 4 operations: CRUD —create, read, update, delete. My suggestion is that her's would only do CUD. That's probably technologically impossible (updating and deleting require the ability to locate a record), so really we'd be imagining a database in which you can only put stuff in.

4.6.2.3 Relational account

Where the liberal and republican accounts have been focused on the individual and the effects of intrusions (or the possibility of intrusions) into the protected sphere on individuals, for the relational theorist, this misses some important effects of privacy.

On these views, privacy is a necessary condition of being able to form / conduct intimate and other relationships. As we've already seen back in discussing Nissenbaum, probably all relationships have their own norms about privacy (or as she calls it, information flows). Thus we will also be concerned with protecting professional and other social relationships, even when those relationships might not matter very much to the individuals involved.

4.6.2.3.1 Right to protected sphere

There are many different kinds of valuable relationships. Thus the sorts of information which fall within the protected sphere will vary widely depending on the relationship in question. In general, we can say that

(RPS-Rel) If the relationship between S and V would be negatively affected by the disclosure of information x to a third party, x is within the protected sphere.

Obviously, this will require a lot of caveats and qualifications. Criminal conspiracies are not the sorts of relationships we want to protect; we probably don't want to say that an informant on a criminal conspiracy does something morally wrong.

Instead of getting bogged down trying to sort our exceptions, let's just go quickly through a few kinds of relationship to illustrate what forms of information will fall within the protected sphere.

4.6.2.3.1.1 Romantic relationships

Think of romantic relationships. Being able to keep secrets connects to feeling safe and able to be intimate and connected with someone. Indeed, for many purposes, courts treat communications between spouses as covered by the 5th amendment protection against self-incrimination. If you are suspected of a crime, your spouse usually cannot be compelled to provide information.[66]

4.6.2.3.1.2 Friendships

Think of friendship. While friendships can be just as complicated and varied as any other relationship, it's probably safe to say that the amount of trust you place in someone to keep something sensitive a secret often tracks how close you are. You tell the secrets and fears which you really don't want getting out to your close friends; acquaintances get funny but not overly embarrassing stories. Thus the information within the protected sphere will differ depending on whether we are talking about you and your best friend, or you and a random acquaintance.[67]

4.6.2.3.1.3 Social organizations

Groups need ways of defining themselves against non-group members. Sometimes this is just done by the nature of the group itself. Send an Astros fan to hang out with a bunch of Dodgers fans. Their non-membership will be uncomfortably obvious.

Othertimes, the definition is created more intentionally by having information that is shared only within the group. Within such clubs or other social organizations, having certain shared secrets is important for group cohesion/bonding. Think of shared lore and history, secret symbols, and other things only known to members of sororities or fraternities.[68]

4.6.2.3.1.4 Workplace

Think of information sharing in the workplace. Since you often need to work productively with people who have very different political or religious views, norms of professionalism normally require that people do not go into too much detail about these topics. The ready availability of this information in the workplace may damage people's ability to work together (imagine someone brings in a chart ranking the liberal/conservativeness of her co-workers' tweets).

[66] Though this is complicated and I'm not a lawyer, so don't take this as advice. Also, the history of this privilege cuts against the point I'm making a bit.

[67] Maybe you could say that the kinds of information goes n the sphere are the same for any friendship. It's just that for less close friendships, the friends put less into the sphere.

[68] My grandpa was a longtime member of the Elks Lodge. I used to laugh at how seriously he took the secrecy of its various symbols —e.g., why is the clock on the emblem indicating that time? Sorry, grandpa. Now I get it.

Similarly, organizational coherence often requires those in leadership to maintain some secrets from those below (or above) them. More importantly, and I've seen this a gazillion times, getting people to agree and be on the same page with a controversial decision often means allowing them to agree for different reasons. If all the differing reasons behind a controversial decision were public, it can be very difficult to convince those outside the decision-making group to go along since everyone can find a reason to reject it.

Finally, some presumption of confidentiality is often important for enabling honest discussion. When I attended the Management Development Program at Harvard —basically a business school crash course for aspiring administrators—we were sworn to confidentiality on the first day and reminded of that oath constantly. I, of course, found it amusing. But I was also frequently surprised how reminding people of that oath was all it took to get someone to talk freely about something happening in their university.[69]

4.6.2.3.1.5 Consciousness raising

An important part of activism and empowerment by feminists and members of marginalized groups is to create environments which exclude members of majority groups. A group of women is less likely to have honest conversations about how they feel treated in the workplace and thus less likely to recognize their common experiences when men are around.

Take 'mansplaining' —the tendency of men to disregard the expertise of women and condescendingly lecture to them. Virtually every woman in the workplace has experienced this; many mixed-gender discussions of the topics immediately lapse into demonstrations of the phenomena with women pushed out of the conversation. It's thus hard to imagine this common problem getting a name if women did not have space to themselves to discuss their experiences.

More generally, marginalized groups need privacy from dominant groups in order to recognize the legitimacy of their experiences, deal with internalized oppression, and prepare political

[69] I shouldn't have been surprised since it does make a lot of sense. The world of academic administrators is very small. It is very easy for some momentary bit of candor to a rando at a conference to come back and bite your at you job. You don't last long in those jobs without a strong instinct to keep your mouth shut (hence why I will never be in one of those jobs).

action. Thus the protected sphere of information for a group of marginalized persons may be quite large.

4.6.2.3.1.6 Civil inattention

Finally, consider what sometimes gets called civil inattention. As Nagel pointed out, a functioning civil society requires room for strangers to be strangers.

One thing I loved about living in New York, which visitors often misunderstand, is the complete lack of pressure to be chatty with everyone. Unlike Los Angeles, where people mostly drive and thus rarely see the same strangers regularly, in New York (at least Manhattan) people mostly walk and take the subway. Making small talk with everyone you see regularly would be exhausting. Thus you might walk past the same person every day for a year without either of you ever showing the slightest hint of recognition. Maybe, just maybe, you nod occasionally. When people are so tightly packed together, ignoring others needn't be rude; it grants them the breathing space necessary for living their own lives.

4.6.2.3.2 Objections

It's a bit difficult to respond to the relational account from what I've said so far. Ignoring strangers, keeping your lover's secrets, fraternity secret handshakes, and all the other cases are pretty varied. They don't have much in common, other than the idea that certain kinds of relationships are valuable and require a degree of privacy to exist/flourish.

That said, some of the concerns we raised for the liberal and republican accounts apply here too. In particular, we seem no closer to explaining the wrongness of privacy violation when it is separated off from all other effects (i.e., it's hard to explain what's wrong with the Data hoarder).

[ToDo: Rewrite this] Similarly, the concerns about invading the protected sphere turn on controlling information which needs to be private in order to promote the valuable relationships. If we wanted to explain the intuition that ubiquitous surveillance is always wrong, we will find a gap for any information which is not necessary for some form of valuable relationship.

4.7 Can machines violate privacy?

Let's close off our discussion of privacy-related harms by considering a question whose relevance will become clearer once we start discussing responsibility: Can a machine violate privacy?

For example, the original business model of Gmail was to serve users ads based on keywords located in emails they send and receive. There was no Google employee behind the scenes reading your emails and determining which ads to show you. The process was entirely algorithmic. An outdoor brand would pay Google to advertise to people interested in, say, hiking. Google would look for words like 'tent' or 'hiking boots' in emails and serve ads to people sending/receiving those emails. To people who felt this violated their privacy, the response (I'm paraphrasing) would be 'but you and the recipient are the only humans looking at the emails. How could that violate privacy?'

Let's try to think through whether there could be harms in these sorts of cases using the various theories we've discussed.

4.7.1 No. Understanding matters

Let's mention one very flat-footed reply: No machine can violate privacy because machines do not understand what they are 'reading'. The argument thus may go:

(1) A violates B's privacy regarding x only if A understands x.

(2) Machines cannot understand anything.

Therefore, (3) Machines cannot violate privacy.

This argument is probably vulnerable to a straightforward objection when we are thinking about cases in which the purpose of violating privacy is to enable some human action. For example, surveillance of internet traffic intended to help stop terrorist attacks. Assume that any human action requires understanding. Either the machine understands what it is analyzing or it doesn't. If it does understand what is analyzing, then privacy is breached. (Obviously, the proponent of this argument will not want to buy this options since she denies that machines can understand). But if the machine does not understand what it is analyzing, then its output is useless unless a human operator intervenes. Since humans understand things, privacy is violated.[70]

[70] [ToDo] This argument can be found in [Muller]

Of course, if we set aside the assumption that the privacy violation is in service of human action, this objection doesn't get off the ground. No one supposes that in the Gmail example, the algorithm which sells adds to advertisers based on keywords actually understands anything.

We could fight about premise (2). But that would take us on a long detour through philosophy of mind and philosophy of language. So we won't.

More importantly, we don't need to worry about (2) unless (1) is true since it makes the privacy violation depend on the existence of understanding. If (1) is false, then it may be that privacy can be violated even when there is no understanding.

Why would we believe (1)? If you start with the thought that what we care about is other people knowing stuff about us, it will seem plausible. Suppose that someone steals your bag containing the diary in which you've written all your deepest darkest secrets while you are traveling in a country where no one speaks English (and phones which can translate text in a picture don't exist). Are you worried about your privacy? If you are, this view would claim, it's only because you're worrying that someone who speaks English will come along and read it. If you are certain that won't ever happen, then you shouldn't be concerned. And, certainly, we wouldn't say that your privacy has already been violated as soon as the theft occurred.

But we've already seen that the harms we're concerned with do not turn on the question of whether anyone/anything understand what they read. On all of our accounts, the harm can arise from self-censorship. And, thinking of cases like the Argument Tweeter or People-Rating, should remind us that we care about the ways our information can be used regardless of whether there is an understanding agent in the loop.

4.7.2 Application of 3 approaches

[ToDo]

Liberal

Neo-republican

Relational

Q3: Responsibility
v.0.0.1

5.1 Test cases

As usual, lets get some test cases on the table before we start.

5.1.1 Discriminatory bank algorithm

Human beings are fallible. It is very hard for even well-intentioned people to consistently apply a set of criteria in judging a wide range of cases. Machines on the other hand are great at this. It's what they do.

Thus it seems like a great idea to take decisions which have historically been infected by biases like racism out of the hands of humans. The computer won't care if you are black, white, asian, or latinx. It will only care whether you are statistically likely to pay back your mortgage.

However, underwriting does require judgment. We can't just apply a set of clearly defined standards, there are a lot of tradeoffs and decisions about risk that need to be made on the basis of lots of other information.

This is where machine learning techniques can be helpful. By simply providing the algorithm with a huge dataset about past mortgage decisions and defaults, it can extrapolate its own rules which mimic those of a competent underwriter.

If the problem isn't yet obvious, here it is: Historically, mortgage decisions have been distorted by racism. We won't get non-racist decisions by using that data to train the system. Racism in; racism out.

5.1.2 Gatekeeper algorithms

Zeynep Tufekci writes that

> Algorithmic gatekeeping is the process by which such non-transparent algorithmic computational-tools dynamically filter, highlight, suppress, or otherwise play an editorial role—fully or partially—in determining: information flows through online platforms and similar media; human-resources processes (such as hiring and firing); flag potential terrorists; and more. {Anonymous:SbzjS1LM} pp.207-8

Let's call these gatekeeper algorithms.

5.2 Intro

Now that we have a grip on the sorts of harms that might be involved with the use of personal data, we can now turn our third question

(Q3) How should we assign blame when people are harmed by algorithmic uses of personal data?

If an identifiable individual caused the harm, this isn't terribly interesting or controversial.

The more difficult issues arise when we have corporate agents or actions undertaken by mixes of human and machine. Both the technology itself and the way that technology is created may enable novel difficulties in understanding moral responsibility. Doorn nicely summarizes some of the issues here and how they differ from traditional thinking about responsibility.

> The ethical literature…often assumes: (1) that it are individuals who act, (2) that the consequences of their actions are causally direct traceable, and (3) that these consequences are certain. None of these assumptions seem to apply to many of the ethical issues raised by modern technology and engineering. First, engineering and technology development typically take place in collective settings, in which a lot of different agents, apart from the engineers involved, eventually shape the technology developed and its social consequences. Second, engineering and technology development are complex processes, which are characterized by long causal chains between the actions of engineers and scientists and the eventual effects that raise ethical concern. Third, social consequences of technology are often hard to predict beforehand. " {Doorn: 2012ij} p.2

Still, it's hard to get a grip on what new problems might be arising for moral responsibility. Normally, when a corporation or other hierarchical organization misbehaves, we just push responsibly up to senior leadership or the board of directors. Why isn't that enough here?

Suppose our company is responsible for something bad. Pick your example. The plane crashed. Software was hacked. Data was lost.

From the victims' perspective, the company is responsible. They (try to) sue the company to redress their losses.

From the Board's perspective, this is straightforward. If it was bad enough, they fire the CEO. Job number one for their successor is demonstrating that the disaster will never happen again.

From the new CEO's perspective, this is straightforward: fire whichever Vice-President oversaw the department whence the mistake arose.

You're the new VP of that department. You want to demonstrate, both to the CEO and to your subordinates, that such errors are unacceptable. Whom do you fire? How do you spell out a policy which holds subordinates responsible for such errors?

If it's a single engineer's error — she neglected to check the size of a variable and so opened the software to a buffer overflow[71]— that's easy. But what of the team that did the code review? Also easy. If they should've caught the error, fired. If some procedural quirk allowed the bug to sneak through, we tighten up our reviews.

[71] This is a very dangerous form of hack which allows the attacker to run code on the user's machine. Here's a good explainer video from computerphile: https://www.youtube.com/watch?v=1S0aBV-Waeo

But most of the time, it's not going to be this easy. Engineers solve engineering problems. If the problem wasn't stated in the specs given to them, how were they to know to solve it? What about problems that arise through the unforeseen interaction of very different components, created by different engineers? Obviously, we want engineers to be conscious of their job beyond what's in the specs of what they are working on; this is a key theme in both security and safety —that security and safety have to be everyone's job. But how far does this go? If an engineer points out a problem and it gets passed up the chain, isn't her job done?

There are cases in which each of the decisions was perfectly reasonable in isolation; it's only when they are combined within an engineering process that the danger arises

5.3 Two questions

When things go wrong we want to identify and blame those responsible. There's a flip side to this too. If you're an engineer working on a project, what are you morally responsible for? Are you only responsible for preventing harms which come from the particular component you are designing? Or are you responsible for ensuring the whole system doesn't cause harm?

Let's separate two questions:

(A) When is a person morally responsible for a harm?

(B) What does it mean for a person to be morally responsible?

The first question is about identification; the second is about implications. It may help to think of how a criminal case works. There are two questions for the jury. Did the defendant commit the crime? And, if they did, how should the be punished. That's what I have in mind here.

5.4 When is someone morally responsible?

The question of when someone is morally responsible for something is controversial.[72] However, most answers will claim that:

(1) S is morally responsible for harm x only if S is causally connected to x

Obviously we don't want to blame people for things they didn't do. Thus there must be some sort of causal connection between the agent and the outcome for which she is morally responsible. However, usually we want to go further and require that S has some control over the outcome / events.

(2) S is morally responsible for harm x only if S knew, should have known, or could have known that her actions may cause x

If the causal sequence that led to the outcome was completely and totally beyond a person's knowledge and anything they could've predicted, then it is unlikely that they are morally responsible. We can only hold people to what they can do, thus we cannot blame them for failing to do the impossible.

(3) S is morally responsible for harm x only if S could have done otherwise.

This is the always intuitive and always difficult to pin down free choice requirement. If S had no ability to do otherwise, we generally would not hold her morally responsible. The exceptions are usually cases where she somehow culpably put herself in the position where she could not do otherwise. [C.f., duress]

All of these are tricky. Thus to give ourselves the most to work with, let's update (1) and (2) with some concepts from philosophy of law and jurisprudence. In particular, let's add the common components of a crime to the mix, since some of the extra tools we're going to need may have already been developed in those areas.

[72] See for example https://plato.stanford.edu/entries/computing-responsibility/

As we go, I will try to set out how these practices related to technology complicate things.

5.4.1 Causal connection: Actus reus

Let's understand the causal connection required in terms of the actus reus of a crime. Every crime requires you to do something —it involves either a wiggling of the body or a failure to wiggle. Often we supplement this with the outcome of the wiggling. For example, the actus reus of murder is homicide: causing the death of another. If you do all that you needed to do to kill someone but they fail to die, it is not murder (though it is a different crime like attempted murder or aggravated battery)

5.4.1.1 Omissions

Notice that we included both overt actions and failures to act. The latter are often called omissions. In US criminal law, there are very few crimes where the actus reus is an omission — I've heard that there are only 2, one of which is failing to file your taxes.[73]

In tort law and in ethics more broadly there are culpable failures to act. Many think that you would've done something wrong if you fail to save someone's life when you could've done so at little or no cost to yourself. The possibility of being responsible via omissions is pretty important when we get to the sorts of harms caused by engineering.

5.4.2 Mental state: Mens rea

If you've watched a lot of CSI or other police procedural TV shows, you may know that (almost) every crime requires you to have a corrupt mental state. [74] This is the mens rea of a crime.

[73] 'Good Samaritan' laws which require people to aid others in need or face punishment are extremely rare and hard to apply. They could involve omissions as the actus reus, but might actually require some sort of overt act (e.g., driving away).

[74] 'Almost' because there are some strict liability offenses where no mental state is required. This is more common with infractions (a different category from crimes), e.g., speeding tickets.

Thus instead of talking about whether a person knew that the harm would/could result, we can distinguish different kinds of culpable mental state. The distinctions often affect our judgments of the crime's severity.

Under the Model Penal Code there are basically 5 culpable mental states:

- Purpose: You are trying to do x

- Knowledge: You know that what you are doing is x (even if you aren't trying to do it)

- Recklessness: You are aware of a substantial risk of harm from doing x

- Negligence: A reasonable person would've known that x creates a substantial risk of harm.

- Wanton and Depraved Heart[75] (sometimes: Malignant and Abandoned Heart): Basically negligence, but what you did was so f-ed up that we want to punish you more severely.

Let's look quickly at how these different mental states affect culpability.

Homicide, killing another person, is the actus reus of several crimes. It is not on its own a crime. The crime of murder is committing a homicide with purpose or knowledge (or a wanton and depraved heart). If you know that what you are doing is killing someone or if you are trying to kill them, you are committing murder.

Manslaughter is reckless homicide. You aren't actually trying to kill the person, but you are doing something that you know has a substantial chance of killing them. There are a variety of

[75] Obviously, this is the best phrase in law. Soooo much better than 'A frolic of one's own'

gradations here. Many jurisidictions distinguish between voluntary and involuntary manslaughter; depending on the details it may be that voluntary manslaughter involves recklessness and involuntary manslaughter involves negligence.

Some jurisdictions have negligent homicide statutes. These get tricky. For example, (IIRC) New Jersey had a negligent homicide statute which tried to punish drug dealers when people they sold the drug to die. That was struck down as unconstitutional.

Finally, in many jurisdictions, if you do something so despicable that your negligence demonstrates a lack of concern for human life, you may be charged with murder. Someone who thinks it is a fun game to throw pieces of brick off a highway overpass between the cars underneath may be charged with murder when someone dies. Their mental state was basically negligence —the whole point of the game was to not hit the cars and they thought they were so good at it that they wouldn't actually hit a car— but statutes make it murder.

With this more sophisticated understanding of what's baked into responsibility, let's see if technology actually poses new problems.

5.4.3 Problems: Causal connection

The causal connection required for responsibility gets complicated in several ways by technology.

5.4.3.1 Problem of many-hands

Suppose again that you are the new VP overseeing the division which created the problem. You want to credibly promise to the CEO that this problem will never happen again. That means figuring out who is responsible and taking appropriate corrective action.

There is an epistemic problem here. It's hard to know who made which choice, especially from the outside. This is the original provenance of the problem of many-hands.[76] To see the problem, consider this, from the Stanford Encyclopedia of Philosophy

> One classic example of the problem of many hands in computing is the case of the malfunctioning radiation treatment machine Therac-25....During a two-year period in the 1980s the machine massively overdosed six patients, contributing to the eventual death of three of them. These incidents were the result of the combination of a number of factors, including software errors, inadequate testing and quality assurance, exaggerated claims about the reliability, bad interface design, overconfidence in software design, and inadequate investigation or follow-up on accident reports. Nevertheless…it is hard to place the blame on a single person. The actions or negligence of all those involved might not have proven fatal were it not for the other contributing events. This is not to say that there is no moral responsibility in this case as many actors could have acted differently, but it makes it difficult to retrospectively identify the appropriate person that can be called upon to answer and make amends for the outcome.[77]

Subsequent literature has focused on sorting out whether this is just an insider-outsider epistemic problem, or whether there is a more metaphysical problem as well.[78] I'm going to suggest that there is.

[76] The original term comes from

Thompson, D. F. (1980). Moral responsibility and public officials. American Political Science Review, 74, 905–916.

[77] https://plato.stanford.edu/entries/computing-responsibility/

Associated references:

(Leveson and Turner 1993; Leveson 1995) (Nissenbaum 1994; Gotterbarn 2001; Coeckelbergh 2012; Floridi 2013),

[78] See: van de Poel, I., Nihlén Fahlquist, J., Doorn, N., Zwart, S., & Royakkers, L. (2011). The Problem of Many Hands: Climate Change as an Example, *18*(1), 49–67. http://doi.org/10.1007/s11948-011-9276-0

It very well may be that every individual decision was perfectly reasonable on its own. It may be that the responsible entity is not one individual but the entire team. Moreover, the team's responsibility may only be understandable when we zoom out to the organizational structure in which it operates. The policy guidance, engineering requirements, and other institutional factors may be crucial for understanding what is responsible.

To be sure, this is very strange. We are not wired to ascribe responsibility to abstract things — teams within organizational context— where we cannot reduce the blameworthiness to the individuals comprising the group. Indeed, much of the ethical literature on responsibility in this context has focused on expanding our notions of responsibility to improve engineering practice. If we focus on ascribing blame —backwards looking responsibility— in the way I've discussed, we will not prevent future problems.

Another response to this claim is to push the metaphysics further. Perhaps we need to expand our understanding of what can count as a moral agent. As so often when we are suspicious that the moral issues raised by technology somehow require an extravagantly revisionist metaphysics, Luciano Floridi is here to help. We'll get to him in a bit.

5.4.3.2 Temporal and physical distance

Causation is also complicated by the temporal and physical distance between actions and events enabled by computing. Years may pass between the creation of the code and the result of the harm. This poses a significant epistemic difficult. Indeed, many of the things human moral psychology[79] has trouble dealing with —distance, bad things happening to faceless strangers, uncertainty— will be features of technology-enabled harms.

More importantly, the engineer often has no idea who the consumer of a product will be or how it may be used. But many products can be used in completely unintended ways. If the creator of the code/machine had no way of anticipating that some feature they included would be used in an unanticipated way which results in a harm, how are we to say that they are responsible?

[79] This shouldn't be a surprise. We evolved in relatively small groups in relative isolation. We are really good at moral evaluation of effects on people we know and can see; bad at those we can't. Obviously, that doesn't mean we can't find ways to overcome them.

5.4.3.3 Development practices

The ways technology is developed pose several challenges for responsibility.

5.4.3.3.1 Libraries

Software (and often hardware) development involves using libraries that others have developed. This is often a good thing. It both saves time and can promote security. Unlike when I first started programming, nowadays, if you are developing a web app, it would be irresponsible to try to write it all from scratch. Web developers use frameworks that others have created.[80] Because lots of people use and test the framework, its bugs and security flaws can be rapidly detected.

For a small example, in my grading app, the code I've written comprises about 10 MB. The PHP libraries comprise 175 MB and the JavaScript libraries comprise 338 MB.

However, this has the downside that developers are essentially stringing together black-boxes — they don't really know how their software is working. It is also problematic because you will only get the latest security updates if you update the libraries. But new versions of libraries come with all sorts of changes, some of which may break your system. Take it from me, the misery of trying to figure out how a necessary update has created a weird bug is quite exquisite.

This means that the people creating a program are not entirely creating the program.More importantly, they often are not in a position to understand how the program works. This is a completely standard practice, and on the whole, a huge benefit.[81] But it does pressure both the causal connection and the knowledge required for responsibility.

5.4.3.3.2 Methodologies

Software teams have a lot of different strategies for building products. Methodologies like scrum

[80] For example, in web development, I use Laravel [ref] for all the server-side operations and Vue to help build the part that people see.

[81] It also allows people trained in philosophy to do things that 20 years ago would require significant training in math or computer science.

and agile are very common. These involves breaking a system down into bite sized chunks so that each developer works on a series of small, often disconnected parts. The software can thus be continuously improved, allowing you to get to market faster.

In a mature organization, there will be a code review by other members of the team or managers to check over work before it goes into production. Though this is not always the case. Presumably, when there is a review process, the reviewers acquire some amount of responsibility. Similarly, an organizations decision not to have a review process may be important to assessing responsibility for harms.

5.4.4 Problems: Mental states

If responsibility requires awareness of what you're doing (or that a reasonable person would've been aware, as in negligence), the interaction of technology and people complicates responsibility

For one, end users of technology often have wildly mistaken ideas about how a system works. Modern technology is like magic. Most of us are not magicians. Let's discuss a few ways this complicates things.

5.4.4.1 Bugs

Every system has bugs. A good development and testing regime will ensure that obvious problems created by bugs don't arise for users. Though many companies skimp on testing in order to get to market quickly. More importantly, many bugs will involve completely unanticipated corner cases.[82] In some cases, those situations should've been anticipated. In others, there's really no way they could've been.

[82] From wikipedia: "a corner case (or pathological case) involves a problem or situation that occurs only outside of normal operating parameters —specifically one that manifests itself when multiple environmental variables or conditions are simultaneously at extreme levels, even though each parameter is within the specified range for that parameter." https://en.wikipedia.org/wiki/Corner_case

[ToDo: Examples]

Oftentimes, bugs which have consequences for end users will be completely hidden from users and developers by interfaces.

5.4.4.2 Biases

Another important issue for the required mens rea is that people are very often wrong about how well machines function. We often either overweight or underweight their accuracy; we may believe the machine readout over our own eyes; or fail to do so when we should believe the machine.

For some examples,

The opacity of many computer systems can get in the way of assessing the validity and relevance of the information and can prevent a user from making appropriate decisions. People have a tendency to either rely too much or not enough on the accuracy automated systems (Cummings 2004; Parasuraman & Riley 1997). A person's ability to act responsibly, for example, can suffer when she distrust the automation as result of a high rate of false alarms. In the Therac 25 case, one of the machine's operators testified that she had become used to the many cryptic error messages the machine gave and most did not involve patient safety (Leveson and Turner 1993, p.24). She tended ignore them and therefore failed to notice when the machine was set to overdose a patient. Too much reliance on automated systems can have equally disastrous consequences. In 1988 the missile cruiser U.S.S. Vincennes shot down an Iranian civilian jet airliner, killing all 290 passengers onboard, after it mistakenly identified the airliner as an attacking military aircraft (Gray 1997). The cruiser was equipped with an Aegis defensive system that could automatically track and target incoming missiles and enemy aircrafts. Analyses of the events leading up to incident showed that overconfidence in the abilities of the Aegis system prevented others from intervening when they could have. Two other warships nearby had correctly identified the aircraft as civilian. Yet, they did not dispute the Vincennes' identification of the aircraft as a military aircraft. In a later explanation Lt. Richard Thomas of one of the nearby ships stated, "We called her Robocruiser… she always seemed to have a picture… She always seemed to be telling everybody to get on or off the link as though her picture was better" (as quoted in Gray 1997, p. 34). The captains of both ships thought that the sophisticated Aegis system provided the crew of Vincennes with information they did not have.[83]

5.5 Implications of moral responsibility?

Turn now to what we mean by moral responsibility? That is, if someone is morally responsible for a harm, what does that entail for how we may relate to her? May we call her mean names? Shun her? What does it entail for how she should relate to herself? Should she feel guilty? Should she be mad at herself?

As it is a common notion which we use all the time, responsibility is hard to pin down. A useful strategy for getting started, is to invert the question. For example, here we ask what it means to say that someone's not responsible. Pretty clearly, it means they owe no apology, it would be wrong to punish them, and attitudes like blame would be inappropriate.

Thus we can say that moral responsibility means or implies at least three things

1) Reactive attitudes and evaluation are appropriate

2) Compensation / retribution / apology may be owed

3) Punishment may be appropriate

Now, it is true that some writers think that the we need to break the connection between responsibility and blameworthiness. That is a pretty radical suggestion which we'll get to below. First, let's take each of the three components

[83] https://plato.stanford.edu/entries/computing-responsibility/

5.5.1 Reactive attitudes and evaluation are appropriate

Reactive attitudes are mental states which we take towards someone's actions. These include states like praising, blaming, criticizing, punishing, being disappointed, being pleased, and the like. They matter for ethical evaluation because the appropriateness of experiencing them is governed by our moral norms. It is appropriate to feel grateful or praise toward someone who rescues children from a burning building. There is something wrong with a person that fails to feel angry or critical towards someone who gratuitously injures a child.

Thus when I say that moral responsibility involves reactive attitudes and evaluation being appropriate, I mean

(1) S is morally responsible for x only if it is appropriate for S to be the subject of reactive attitudes like blame for x

If Scarlet saves a bunch of children from a burning building while Violet just watches, it would be appropriate to praise Scarlet. It would not be appropriate to praise Violet.[84] That's because Scarlet is morally responsible for the children-saving. Violet is not.

This is a necessary condition because there may be other non-moral reasons for punishing / rewarding —e.g., rewarding the salesperson who sold the most last month. Though there may sill be some connection with moral responsibility because our sense of justice is tied to responsibility. If you actually sold more product but someone else gets the salesperson of the month award, you can legitimately complain that it is unfair.

5.5.1.1 Reactive attitudes

The reactive attitudes which are appropriate can get tricky. Generally speaking, a person deserves blame when they've done something wrong.

[84] NB, that doesn't mean that we necessarily should criticize Violet. Just the when we are making a list of heroes, Scarlet belongs on it, Violet doesn't.

[ToDo: other reactive attitudes]

However, there are also cases of blameless wrongdoing. In these cases, a person may owe an apology or otherwise be responsible for repairing what has been broken.

5.5.2 Compensation / retribution / apology may be owed

If someone is not responsible for a harm, it would not make sense (or even be unjust) to demand that they compensate the victim. However, these forms of repair are appropriate when someone is responsible for the harm.

(2) S owes compensation / retribution / apology to V for harm x only if S is morally responsible for x.

5.5.3 Punishment may be appropriate

It is unjust to punish someone who is not responsible for a wrong. Punishment here covers both legal punishment at the level of society —criminal sentences or punitive damages—and non-legal interpersonal forms —e.g., shunning.

(3) S may be punished for x only if S is morally responsible for x.

Note that this is a place where judgments of moral responsibility diverge from causal claims. This could also be true of someone who did the relevant act but was unable to be morally responsible. This is why children are not treated the same as adults. It's also why the insanity defense[85] is required for justice —if someone lacks the capacities for moral responsibility, it is wrong to punish them.

5.6 Floridi and gatekeepers

Let's turn now to a fairly extravagant claim: that some computer programs may need to be regarded as moral agents responsible for harms.

5.6.1 Moral patients

Before getting into Floridi's picture, let's start with the notions of moral agents and patients. These concern what sort of beings 'count' in moral considerations.

Suppose that, deep in the wilderness, you smash a rock with a hammer. Since the rock doesn't belong to anyone, you do nothing wrong. Indeed, it seems weird to even ask about whether it was right or wrong. Whereas, if you then hit me with a hammer, the question of right or wrong is very natural. To capture what is at stake, we will say that

x is a moral patient iff x's interests must be considered in moral decision-making

We do something wrong if we ignore the interests of a moral patient. Rocks are not moral patients. We do not need to consider their interests in deciding whether to throw them. Animals are moral patients. We do need to consider their interests before we do things to them. Throwing a rock at a dog is wrong because of what it may do to the dog, not the rock.

5.6.2 Moral agents

We can ask about what sort of beings can act in moral/immoral ways. Occasionally, a primate study will make breathless headlines with claims like 'Chimps can act morally' [ToDo: Add deWaal refs]. TV shows ask us to consider whether psychopaths are capable of morality. Mental illness and the insanity defense are common subjects of controversy. All of these are questions about when something counts as a moral agent.

[85] The insanity defense (as well as the defense of infancy —being a child) is different from other defenses in that it asserts that the defendant is not the sort of being which can be subject to the law.

Thus we will say that

x is a <u>moral agent</u> if and only if x's actions can be subject to moral assessment

By 'moral assessment', I mean that it is appropriate to use moral concepts in judgment of its actions. Normally, I would put this differently: moral agents can be morally responsible. However, as we'll see below, Floridi wants to say that moral agents may not be morally responsible. Thus we need the wider definition.

Volcanos are not moral agents. When a volcanic bomb —a rock blasted into the air— hits and injures a person, it makes no sense to ask if the volcano did something wrong. It's not the sort of thing which can do wrong.[86]

Similarly, most animals are not moral agents. Some species may be, the jury is still out. Often, when we seem to apply moral concepts to animal actions —we say that the cat playing with the doomed mouse is cruel— we don't really mean it, as is shown by our immediately inviting the 'cruel' beast into our home.[87] Humans are normally moral agents, with the exception of small children and those with severe mental disabilities.

Notice the asymmetry of what counts as a moral agent and what counts as a moral patient. My dog is a moral patient; he is not a moral agent. The same applies to very small children. It would be very wrong to ignore the welfare of a child. The (small) child cannot be blamed when they do something that would be wrong for an adult to do, like throw a rock at the dog since they do not yet have the cognitive abilities necessary for telling right from wrong.

5.6.3 Warm up

As warm up, Floridi notes that as humans have made moral progress, especially in the recent past, our understanding of who/what counts as a moral patient has enlarged. Many writers now include

[86] If you believe that a volcano is a god which requires human sacrifice to keep it from erupting, then you probably believe the volcano is a moral agent.

[87] We're probably using moral concepts as analogies. We are saying something like 'If a human behaved like that, they would be cruel'.

- Future people
- Subjects of posthumous harms (harms to people who are dead)
- The environment / natural world
- Animals

in the class of moral patients.

Including a natural environment (e.g., a pristine forest) as a moral patient means that it merits moral consideration on its own right. This goes beyond saying that humans enjoy hiking in pristine forests and therefore we have reason to maintain them. It goes beyond saying that pollution of waterways has adverse effects on ecosystems which cause harms to humans. It means that we are to evaluate effects on the forest's interests alongside human interests.[88]

Similarly, saying animals are moral patients implies that we have reason to not be cruel to animals because it is wrong to do so. This is in contrast with Kant who thought that it is wrong to be cruel to animals because cruelty to animals hardens a person and makes her more likely to be cruel to people.

With that in mind, Floridi asks us to consider that maybe it is now time to enlarge the scope of what/who counts as a moral agent. In particular, maybe it's time to to treat artificial agents like computer programs as moral agents too.

Time to get crazy.

[88] If you're confused by how we would actually do this in practice, I'm right there with you. Most suggestions involve things like appointing a human advocate for the forest or just being absolutist and saying the interests of natural environments can't be traded off against human interests.

5.6.4 Starting points

Floridi wants us to approach this question with an adequately open mind. He thus proposes some ground-rules. He builds a framework of levels of analysis to situate and formalize those rules.[89] We'll ignore that framework and just focus on the principles he wants us to start with.

5.6.4.1 No anthropocentrism

If we start from the assumption that everything within the moral realm is based on human beings, we've already closed the door to potentially important moral considerations coming from non-humans.[90]

Thus if we want to be fully open-minded about algorithms, we must start by purifying ourselves and setting aside any anthropocentric biases. He writes

> Limiting the ethical discourse to individual agents hinders the development of a satisfactory investigation of distributed morality, a macroscopic and growing phenomenon of global moral actions and collective responsibilities resulting from the 'invisible hand' of systemic interactions among several agents at a local level. Insisting on the necessarily human-based nature of the agent means undermining the possibility of understanding another major transformation in the ethical field, the appearance of artificial agents (AAs) sufficiently informed, 'smart', autonomous and able to perform morally relevant actions independently of the human engineers who created them, causing 'artificial good' and 'artificial evil'. Both constraints can be eliminated by fully revising the concept of 'moral agent'. {Floridi:uy} p.3

[89] When you see 'LoA' in his paper, that's what he's talking about.

[90] If you're seeing huge 'DANGER' signs and bright red blinking lights in your mind when you read this, I'm right there with you. To my mind, this move here is the original sin of the argument and the place I would/will direct my attacks. But we'll get to that…

If it helps you get in the swing of things, it may help to keep in mind that the human body, brain included, are (at least largely) machines. Doctors are glorified mechanics.

5.6.4.2 Observability

We're also going to conduct this discussion by defining everything in terms of things which can be observed. Since we can see everything that's going on in a machine, it wouldn't be fair to demand that it have something non-observable. For example, if you thought that humans are moral agents because they have an immaterial spook that resides in them but no one can see (because it's immaterial), then you'd be biasing the inquiry against other possible entirely material agents such as machines.

In other words, if we are committed to not being anthropocentric, we will want to require observability.

Notice that this may be a bit important since most human psychology is unobservable (behavior is; mental states aren't). Thus everything we said above about mens rea looks like its going to be right out the window.

5.6.4.3 Terminology

In order to avoid anthropocentrism in discussing agency, Floridi makes use of a picture loosely drawn from computer science. That entails some fancy terminology.

The systems we're interested in have an internal state which can be changed through various transition rules. In other words, they store information internally and have specific procedures for changing that information.

Let me illustrate this by a different version of the example he gives.[91] Consider a simple finite-state autonoma such as a vending machine or the gate at a (old school) parking garage which goes up when the appropriate amount of change has been put in.[92]

I'll write this out in (rough) Python. (Lines that begin with a '#' are comments for humans to explain what's going on and not part of the program; same for lines that are between two sets of 3 quotation marks: """This is a comment""")

First we define some variables that hold the information we need. This is the internal state of our machine.

```
# The amount of coins needed to raise the gate
required = 3

# The amount of coins that have been put in
paid = 0
```

Then we define some functions which actually do the work. These are the transition rules

```
def raise_gate():
  """Actually raises the gate"""
  # Turn on motor, etc, goes here

def coin_inserted(amount):
  """Function which gets called whenever someone
  puts a coin into the machine"""
```

[91] His example is MENACE the tic-tac-toe learning matchbox machine. {Floridi:uy} pp.8-10

[92] Here's a relatively accessible video explanation https://www.youtube.com/watch?v=vhiiia1_hC4

```
    # Increase the stored value of what's been paid
    # by the amount that just got put in.
    paid += amount

    # Now we check whether enough coins have been inserted
    # by comparing the paid variable to the required variable
    if paid >= required:
        # The customer has paid enough so
        # we call the function which raises the gate
        raise_gate()

    # If not enough coins have been entered,
    # nothing else happens. We wait for the function to be run
    # again when another coin is inserted.
```

What happens is that every time a coin is inserted into the machine, we run the `coin_inserted` function. It adds the number of coins to the stored value `paid` (it changes its internal state). Then it checks whether enough coins have been inserted and if that's true, it opens the gate.[93]

This doesn't have to be an electronic process. The old version of these machines operated mechanically.

5.6.5 Agents

Let's start with what it is to be an agent in general, we can then understand moral agents as a proper subset of agents. If we stick to criteria which must be observable and formulate them in ways which are not anthropocentric, Floridi claims we will find 3 characteristics of agents.[94]

[93] Note that if you put more than the required number of coins in, you are screwed. There's no give_money back function. (For those of you learning python: you would want to write this using exceptions, that way you can also handle things like non-coins being inserted)

[94] [p.7]

5.6.5.1 Interactivity

First, an agent is interactive in that it can be affected by changes in its environment and when it can affect its environment.

Interactivity means that the agent and its environment (can) act upon each other. Typical examples include input or output of a value, or simultaneous engagement of an action by both agent and patient —for example gravitational force between bodies {Floridi:uy} p.7

A rock can be affected by its environment. Hit it with a hammer and it changes. However, without some outside force acting upon it, the rock does not affect its environment.

Your computer is interactive. You type stuff in, it does stuff. Similarly, you and the Earth are interacting simultaneously through gravity. You are pulling up on the Earth just as hard as it is pulling down on you.[95]

A volcano doesn't really interact with its environment.The movement of magma which results in its eruption just is the thing that leads to the eruption. There's no intermediate step. If vulcanism was way different and volcanos erupted because their thirst for human sacrifice was unquenched this would be different, assuming that the volcano thinks "You know, I haven't had a human snack in a long time. I should get those villagers' attention…"

5.6.5.2 Independence

Second, an agent can change its internal state without direct response to interaction; it can perform internal transitions to change its state. Floridi calls this 'autonomy', but since we've been using that term in a very specific way, I'm going to call it independence.

[95] Yep. Thanks general relativity.

Autonomy means that the agent is able to change state without direct response to interaction: it can perform internal transitions to change its state. So an agent must have at least two states. This property imbues an agent with a certain degree of complexity and decoupled-ness from its environment. {Floridi:uy} p.7

Consider our friend the gate at a parking garage which goes up when the appropriate amount of change has been put in. It clearly isn't just responding to the input of a coin since it responds differently after the appropriate number of coins have been inserted. Thus it is independent in Floridi's sense.

5.6.5.3 Adaptability

Third, an agent is <u>adaptable</u> when it can modify the transition rules by which it changes its internal state.

Adaptability means that the agent's interactions (can) change the transition rules by which it changes state. This property ensures that an agent might be viewed… as learning its own mode of operation in a way which depends critically on its experience. Note that if an agent's transition rules are stored as part of its internal state then adaptability follows from the other two conditions. {Floridi:uy} p.7

You demonstrate adaptability when you learn new things or change your mind. Suppose you didn't know that affirming the consequent is a logical fallacy. Like all of us, you think the sentence 'If it has rained recently, the street is wet' is true. But when you look out the window upon a recently hosed sidewalk, you incorrectly conclude that it must have rained recently, and decide not to come to class. That's too bad. If you had come, you would've learned some logic and no longer been tempted to make bad inferences like that. You would see wet streets and look for other evidence that it has rained.

We leave our friend the parking garage gate behind with this criterion. Any time you've put in less than the required amount, it doesn't open; when you've put in the required amount, it opens. That never changes. On its own, it can never decide to hold your car hostage for 4 coins. It can never decide to always allow free exits to those who sympathetically portray garage gates in their teaching.[96]

[96] Obviously, I'm hedging my bets.

Adaptability is one of the hallmarks of some artificial intelligence systems involving machine learning. Think of a convolutional neural network that detects cats in pictures. You start off with a bunch of pictures of cats labeled (true) and tons of other stuff labeled (false). The system starts with transition rules that are random —it guesses whether a picture contains a cat. On the basis of whether it was correct, it updates the rules by which it determines if something is a cat. Eventually, it gets really good at cat recognition through updating its own transition rules.

5.6.6 Moral agents

Let's suppose that we've now captured what it is to be an agent. Presumably, moral agents are a proper subset[97] of agents. Which ones are they? The ones that can do moral actions, duh. Or, as Floridi puts it, some actions are <u>morally qualifiable</u>. Moral agents are agents which can do morally qualifiable actions.

When is an action morally qualifiable? Floridi claims that

> An action is morally qualifiable if and only if it can cause moral good or evil.

This definition is allegedly neutral between consequentialist and 'intentionalist' theories. The moral goods and evils could be found in states of the world —e.g., a world which contains 5 people suffering agony is worse than a world in which only 1 person suffers. Or they could be found in the motives / character of persons —Blue intends to torture 5 people.

From there we can define a <u>moral agent</u>

[97] That is, all moral agents are agents but not all agents are moral agents. Compare: diet coke is a proper subset of sodas.

Agent S is a moral agent iff S is capable of morally qualifiable action.

That's all there is to it. Anything which meets the criteria for being an agent and which can perform morally qualifiable actions, will be a moral agent.

Any resistance to this can only be baseless anthropocentrism. To help us see this, Floridi gives an example of two entities H and W which are able:

> i) to respond to environmental stimuli — e.g. the presence of a patient in a hospital bed — by updating their states (interactivity), e.g. by recording some chosen variables concerning the patient's health….
> ii) to change their states according to their own transition rules and in a self governed way, independently of environmental stimuli (autonomy), e.g. by taking flexible decisions based on past and new information….
> iii) to change according to the environment the transition rules by which their states are changed (adaptability), e.g. by modifying past procedures to take into account successful and unsuccessful treatments of patients.

It seems H and W qualify as agents on our definition.

> Suppose that H kills the patient and W cures her. Their actions are moral actions. They both acted interactively, responding to the new situation they were dealing with, on the basis of the information at their disposal. They both acted [independently]: they could have taken different courses of actions, and in fact we may assume that they changed their behaviour several times in the course of the action, on the basis of new available information. They both acted adaptably: they were not simply following orders or predetermined instructions; on the contrary, they both had the possibility of changing the general heuristics that led them to take the decisions they took, and we may assume that they did take advantage of the available opportunities to improve their general behaviour. The answer seems rather straightforward: yes, they are both moral agents. There is only one problem: one is a human being, the other is an AA [Artificial Agent] ….So can you tell the difference? If you cannot, you will agree with us that the class of moral agents must include AAs like webbots." [12-13]

Remember, it's no good protesting that one is a human and the other a robot. Your resistance to this conclusion is just baseless pro-human prejudice. After all, you agreed to Floridi's setup where all the relevant concepts 'agent', 'moral action', et cetera are defined to be neutral between humans and non-humans.

It should be no surprise that if you drain our human moral concepts of anything specific to humans, you'll get a result that applies to non-human things.

5.6.7 Are gatekeeper algorithms moral agents?

Let's see how Floridi's picture works with one place that artificial intelligence touches most of our lives by discussing what Zeynep Tufekci calls Gatekeeper algorithms. In particular, let's focus on the algorithms which decide what posts you see on social media.[98]

Are gatekeeper algorithms moral agents? Let's run down the list.

Are they interactive? That is, does it do things in response to its environment? Yes. People do things on Instagram. The algorithm makes decisions about whom to show those things to. Then it shows them those things.

Are they independent? That is, do they change their internal state? Yep. The algorithm adds the new post to its queue of posts to make decisions about. It makes a decision about whom to show the posts too. Then it shows them the posts.

Is it adaptable? That is, can it change its own internal rules by which it changes internal state? Yes. This is the 'learning' part of 'machine learning'. The system has a complicated statistical

[98] {Anonymous:SbzjS1LM}

model which predicts what posts will maximize engagement. It updates this model based on whether those predictions were correct.

Therefore, the gatekeeper algorithm is an agent.

Is it a moral agent? That is, can it cause moral goods and evils?

Well, there is a reasonable amount of research which suggests that people who heavily use social media suffer in psychological ways. They may have more depression and have lower self-esteem. Though the extent of the effect is pretty questionable.

There's no question that social media has contributed to violence[99] and other moral evils. Ideological violence seems to be socially contagious; seeing others of your ideological affinity commit violence increases the likelihood that you will do so. This was an integral part of the Islamic State's social media strategy.[100] Increasing the probability of innocent people being murdered seems pretty clearly a moral evil.

[99] [ToDo Add ref to Indian mob violence case]

[100] [ToDo ref]

But is the algorithm the relevant agent? Well, at least at Facebook, that's literally the official story.[101] For example, according to the company itself, Facebook's algorithm neglected to prevent a terrorist from live-streaming the murder of innocent people at a New Zealand mosque

> The members of Congress who gathered for a closed-door briefing had lots of questions for Brian Fishman, Facebook's policy director for counterterrorism. One of the biggest: Why didn't Facebook's counter-terror algorithms—which it rolled out nearly two years ago—take down the video as soon as it was up? Fishman's answer, according to a committee staffer in the room: The video was not "particularly gruesome." A second source briefed on the meeting added that Fishman said there was "not enough gore" in the video for the algorithm to catch it. Members pushed back against Fishman's defense. One member of Congress said the video was so violent, it looked like footage from Call of Duty. Another, Missouri Democrat Rep. Emanuel Cleaver, told The Daily Beast that Fishman's answer "triggered something inside me." "'You mean we have all this technology and we can't pick up gore?'" Cleaver said he told Fishman. "'How many heads must explode before they pick it up? Facebook didn't create darkness, but darkness does live in Facebook.'"[102]

Therefore, the Facebook gatekeeper algorithm is a moral agent.

[101] https://www.fastcompany.com/40475913/facebook-and-google-apologies-for-fake-news-ignore-the-system-itself
https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters
https://www.wired.com/story/facebook-can-absolutely-control-its-algorithm/
https://www.ibtimes.co.uk/facebook-blames-spam-algorithm-blocking-links-wikileaks-dnc-email-leaks-1572488

Refs from https://boingboing.net/2019/04/10/once-again-facebook-blames-an.html

[102] https://www.thedailybeast.com/facebook-tells-congress-new-zealand-shooting-video-wasnt-gruesome-enough-to-flag

Let's try to figure out what that means.

5.6.8 Moral agency without responsibility

Let's consider the implications of what Floridi has (allegedly) shown. He writes

> agents (including human agents) should be evaluated as moral if they do play the 'moral game'. Whether they mean to play it, or they know that they are playing it, is relevant only at a second stage, when what we want to know is whether they are morally responsible for their moral actions. [14]

Thus we should separate the question

> Is x a moral agent?

from

> Is x morally responsible for its actions?

The class of things which are moral agents is larger than the class of things which can be morally responsible for their actions.

He claims that prescriptive (normative?) discourse is larger than responsibility attribution. Indeed, we can morally evaluate actions in beings incapable of responsibility

> Good parents…engage in moral-evaluation practices when interacting with their children even at an age when the latter are not responsible agents, and this is not only perfectly acceptable but something to be expected. This means that they identify them as moral sources of moral action, although as moral agents they are not yet subject to the process of moral evaluation. [15]

Similarly, a search-and-rescue dog can save a person's life; my dog may decide to give me the snap and injure me. Those are moral goods and bads. So we can call them moral agents, without saying that they are morally responsible for their acts. Well, the author of a SEP article elaborates on something Floridi mentions in passing about dogs

> Dogs can be the cause of a morally charged action, like damaging property or helping to save a person's life, as in the case of search-and-rescue dogs. We can identify them as moral agents even though we generally do not hold them morally responsible, according to Floridi and Sanders: they are the source of a

moral action and can be held morally accountable by correcting or punishing them. [103]

That last part is key. A moral agent which is incapable of moral responsibility is still subject to moral accountability. What is that?

The whole conceptual vocabulary of 'responsibility' and its cognate terms is completely soaked with anthropocentrism. This is quite natural and understandable, but the fact can provide at most a heuristic hint, certainly not an argument. The anthropocentrism is justified by the fact that the vocabulary is geared to psychological and educational needs, when not to religious purposes. We praise and blame in view of behavioural purposes and perhaps a better life and afterlife. Yet this says nothing about whether or not an agent is the source of morally charged action. Consider the opposite case. Since AA [Artificial Agents] lack a psychological component, we do not blame AAs, for example, but, given the appropriate circumstances, we can rightly consider them sources of evils, and legitimately re-engineer them to make sure they no longer cause evil. We are not punishing them, anymore than one punishes a river when building higher banks to avoid a flood. But the fact that we do not 're-engineer' people does not say anything about the possibility of people acting in the same way as AAs [14]

The point raised by the objection is that agents are moral agents only if they are responsible in the sense of being prescriptively assessable in principle. An agent x is a moral agent only if x can in principle be put on trial. Now that this much has been clarified, the immediate impression is that the objection is merely confusing the identification of x as a moral agent with the evaluation of x as a morally responsible agent. Surely there is a difference between being able to say who or what is the moral source of the moral action in question and being able to evaluate prescriptively whether and how far the moral source so identified is also morally responsible for that action. Well, that immediate impression is indeed wrong. There is no confusion. Equating identification and evaluation is actually a shortcut. The objection is saying that identity (as a moral agent) without responsibility (as a moral agent) is empty, so we may as well save ourselves the bother of all these distinctions and speak only of morally

[103] https://plato.stanford.edu/entries/computing-responsibility

> responsible agents and moral agents as synonymous. And here is the real mistake, because now the objection has finally shown its fundamental presupposition: that we should reduce all prescriptive discourse to responsibility analysis. But this is an unacceptable assumption, a juridical fallacy. There is plenty of room for prescriptive discourse that is independent of responsibility-assignment and hence requires a clear identification of moral agents. [15]

What problem does this solve?

> Our more radical and extensive view is supported by the range of difficulties which in practice confronts the traditional view: software is largely constructed by teams; management decisions may be at least as important as programming decisions; requirements and specification documents play a large part in the resulting code; although the accuracy of code is dependent on those responsible for testing it, much software relies on 'off the shelf' components whose provenance and validity may be uncertain; moreover, working software is the result of maintenance over its lifetime and so not just of its originators…. Such complications may point to an organisation (perhaps itself an agent) being held accountable. But sometimes: automated tools are employed in construction of much software; the efficacy of software may depend on extra-functional features like its interface and even on system traffic; software running on a system can interact in unforeseeable ways; software may now be downloaded at the click of an icon in such a way that the user has no access to the code and its provenance with the resulting execution of anonymous software; software may be probabilistic; adaptive; or may be itself the result of a program (in the simplest case a compiler, but also genetic code). All these matters pose insurmountable difficulties for the traditional and now rather outdated view that a human can be found responsible for certain kinds of software and even hardware. Fortunately, the view of this paper offers a solution at the 'cost' of expanding the definition of morally-charged agent.

5.6.9 Objections

Now that we've got Floridi's picture on the table and seen what it's supposed to help with, we can turn to assessing it. Should we buy what Floridi is selling? I think not.

To begin, we really need to go back and think through his starting point. Remember, we were just supposed to assume a non-anthropocentric and entirely observable notion of moral agency. But what if there are crucial components to our ordinary conception of agency which this approach has just waived away? If that's the case, the foundations of his account are built on sand.

So what's missing? Think of an accountant who decides to go along with the CEO's demands and report false numbers to the Board. She gets caught and is punished for the fraud. What would we normally say about her? Probably that she made the wrong choice. That she paid too much attention to the CEO and not enough attention to her ethical obligations.

Now think way back to when we first talked about autonomy. Remember the example of the person who has a spasm and smacks the person sitting next to them in the face? Because the action was beyond her control and she didn't intend to hit the person, we said she wasn't responsible.

The point is that any thinking about human agency involves decision-making, i.e., the use of some form of reason. The unethical accountant is blameworthy because she made the wrong decision. The spasmodic smacker is blameless because she made no decision.

With that in mind, let's train our sights on Floridi's account of moral agency. I'm not going to go after the specifics of the view. I'm going to target the approach which supports it.

5.6.9.1 Contra observables

Already we have some tension with Floridi's requirement that agency be based on observables. Consider our corrupt accountant. What would we see as she commits her misdeeds? Well, she takes a call, hangs up the phone and sits at her desk for awhile staring at the wall. Then she types some stuff in Excel, makes a fantastic PowerPoint, and heads upstairs to the boardroom. Our spasmodic sits quietly on the train and then suddenly smacks the person next to her. The unobservable things going on inside their heads are the crucial factor in our judgments of their moral responsibility.

Floridi does consider an objection that runs along similar lines. He imagines his opponent claiming that

> To be a moral agent, the AA must relate itself to its actions in some more profound way, involving meaning, wishing or wanting to act in a certain way, and being epistemically aware of its behavior. [13]

His response, having set out an account of agency is straightforward

> "agents (including human agents) should be evaluated as moral if they do play the 'moral game'." [14]

That is, we've agreed that observability is a criterion of any account of moral responsibility, so the opponent loses.

But wait. Moral responsibility is a concept humans have had for a very long time. Every culture has a version of it. We use it unthinkingly in evaluating whether the accountant made a bad decision. If it has unobservable conditions, so what? I don't know your motives. They are hidden in your head. They still matter for determining what you are responsible for. Indeed, this was the whole point of discussing mens rea above. Morally defective mental states are a requirement of responsibility.

Thus let me bring out a few things which are very standardly pre-conditions of moral responsibility. For one, you need to be able to use reason. Reason has standards, namely logic.[104] To use logic, you need to have a concept of beliefs/claims being true or false. Apparently, children younger than 2[?] lack the concept of truth and falsity. They cannot apply the concept to their situation. If that's true, a very small child cannot be morally responsible since she does not have the relevant capacity for reason.

Does an algorithm have the capacity for reason? On the one hand, it is certainly built out of logic.

[104] That's not to say people need to be able to formulate these standards or explicitly use the rules you learn in logic class.

All non-quantum computers are at the end of the day just a lot of logic circuits. But being able to make logical decisions does not entail having the concept of logic or the concept of decisions. Arguably a computer does not have the concept of true or false;

In this vein, recall what Floridi's says about small children

> Good parents…engage in moral-evaluation practices when interacting with their children even at an age when the latter are not responsible agents, and this is not only perfectly acceptable but something to be expected. This means that they identify them as moral sources of moral action, although as moral agents they are not yet subject to the process of moral evaluation. [15]

Does it? I think not. Or, at least, I can think of two possible alternative explanations of what's going on that don't require the child to be a moral agent.

First, because they are human children they will eventually become sources of moral action. Thus we treat them as though they are moral agents to help them learn to become moral agents. Indeed, since the relevant psychological capacities likely come in gradually, treating them as though they are agents helps them 'grow into' actual agency.

Second, human brains are lazy. We are used to relating to other adult humans as moral agents. Thus our default way of relating to other creatures is as though they are moral agents too. We treat the child as an agent because we are used to treating people that way; if we stop and think, we catch ourselves making unreasonable judgments about infants.[105] Think also of how we find ourselves treating pets as agents. You might think this is what's going on if you've ever found yourself telling the cat, who just came in 30 seconds ago and is now demanding to go out, to make up its damn mind.

5.6.9.2 Contra anthropocentrism

The other ground rule which Floridi raises is the specter of anthropocentrism. He claims that if we set up the game to exclude non-human agents, we won't be able to take seriously the

[105] One of my friends told me about her infant "He keeps dropping his pacifier and then gets upset. I find myself thinking things like 'if you want it, why'd you drop it?' and then I think I'm crazy. He's a baby, what am I talking about."

possibility of artificial agents. That's fine. But it is quite another thing to drain moral responsibility of everything that makes it distinctive and a subject of concern.

Notice the way the paper ends with him trying to explain what it means for a machine to be a moral agent that creates a moral harm. We might, for example, do things to it. It may be reprogrammed or taken out of service morally responsible.

But there is nothing that the machine should do to make up for the moral harm it causes. The company which owns it may have obligations, but it doesn't even make sense to say that the gatekeeper algorithm in our Internet isolation case should do something different. All we can say is that it does something bad by not showing our victim's posts to others. But we have to say that the programmers are the ones who should revise it so that it no longer creates the harm.

Any form of actual moral agency identifies the agent as the wrongdoer (or good doer), sure. But it also means that that agent ought to act differently. It means that agent ought to feel bad for what they've done. It means that agent should make amends. In other words, there is more than one kinds of moral action that is tied to moral agency. If a non-human machine can only do one kind of action, it cannot be a full moral agent. If that's anthropocentrism, then so what? Ethics and morality are features of human life. We had it first. If the robots want it too, then need to be more like us; we needn't conform our concepts to them.[106]

5.6.9.3 The asymmetry

Finally, a point about this whole approach. Recall that to warm us up, Floridi pointed out that we've already expanded the scope of moral patients in order to make us more receptive to expanding the scope of moral agents. Now, this claim isn't crucial to his paper. He could give it up and still do everything he wants. But it's worth thinking about since approaching the history of moral progress in the way he suggests misses a lot.

Maybe there's a good reason we've been slower to expand the scope of moral agents than moral patients. We've expanded moral patient-hood to things whose interests we recognize need to be

[106] Note that what I'm saying here does not affect attempts to extend moral patient status to, say, animals. Those arguments are based on animals having interests.

weighed up alongside human interests. The future generations and animals have (or will have) interests like ours and we've recognized that they ought to be weighed alongside ours. The environment is a bit different; it is still rather controversial that it could have interests independent of human and animal interests.

Indeed, including other beings as moral patients does not entail granting them equal status as humans. It is wrong to torture a mouse for your amusement. But if you could only save a human child or a mouse from a fire, it would be wrong to choose the mouse.