Paternalism and the False Assertion of Privacy[†]

William R. Delise

Abstract: Feeling less private has a way of motivating us to want more privacy. But not all legitimate claims to privacy are associated with experiences of diminished privacy. As I argue, this is because the experience of privacy is related to the experience of autonomy or authenticity, and not all claims to privacy are causally related to experiences of diminished autonomy or authenticity. Moreover, the most valuable, most relevant claims to privacy in the contemporary world are those without the relevant motivating experiences. If I am right, paternalism ought to be prescribed. I identify the "pure" cases for intervention (fingerprinting by Recommender Systems) and "impure" cases for intervention (disclosure of inferential data on social kinds). These cases have a great deal of overlap, which overdetermines the justification for intervention. En route, I try to do two things. First, by introducing a thought experiment called "the entity," I try to demonstrate the rationale for "Group Privacy." Second, I modify Bentham's panopticon to show how commodity privacy is more dangerous than it prima facie appears. Not only are we sometimes unable to detect the violation of our privacy, but commodity privacy can create a false sense of security that has the effect of making us less private.

Introduction

As is widely known, today we face a great number of threats to our privacy. Among others, one thinks of surveillance states, Big Data, smart home devices, built-in AI tools, passive geolocation, social media algorithms, digital advertising, facial recognition, and data leaks. While we can and often do conceive of some of these vectors as threats, or even existential threats to the kinds of lives that we value living, it is rare that we find ourselves motivated to take action in securing our privacy. I would like to propose two reasons for this.

[†]Among others, I am grateful to Dr. Duncan MacIntosh for advising this project, to Michael Johnson-Cramer for listening to and commenting on an early version of the argument, and to Dr. Nicole Ramsoomair, Rosie Price-Digby, and Larissa Wilks for helpful discussion.

On the one hand, privacy today is a nuisance to choose. A precondition of the expedience we value in much of today's technology is access to information about our identities. Choosing to secure our own privacy often means choosing to give up these expediences. In some cases, we are unable to give up the aspects of services that rely on the disclosure of our identities in order to use only the more inert functions of the service.[1] And in the moment of disclosure, we are able to freely sign away our information by one or two movements of the thumb to agree to a privacy policy that we have not read, and are not expected to have read.

On the other hand, the times we ought to prize our privacy the most are often the times we fail to feel a need to secure our privacy. The present discussion will thematize this problem. Not all legitimate claims to privacy are associated with experiences of diminished privacy. This is especially true of the moment of disclosure of information. But while some unfelt disclosures can 'come back to bite us' later, not every delayed effect is felt to be causally connected to a particular or unspecified instance of disclosure. As I will argue, this is because the experience of privacy is related to the experience of autonomy or authenticity (Rössler 2005; Feinberg 1984b), and not all claims to privacy are causally related to experiences of diminished autonomy or authenticity.

If I am right on this point, we ought to endorse paternalism as our next move. If we are right to value privacy and if we are right to think that individuals are not always positioned to protect their own privacy, then there are cases where intervention may be legitimate. In particular, we ought to commit resources to the planning, creation, and maintenance of systems that control illegitimate access to personal information and inferential information about social kinds, or the construction of high powered models for predicting human behavior in a market setting. Moreover, mechanisms exist today that can not

[1]For instance, we can conceive of very useful types of social media, like old versions of Facebook or Instagram, or blogs, that volunteer pieces of content chronologically from a list of followed accounts. This kind of application does not require a Recommender System that has a profile of the user (other than the list of followers). Mainstream social media applications today do not permit one to opt-in into simpler modes of functionality. Plausibly this is because Recommender Systems maximize on utility measured through user attention, and user attention is what is valuable in a social media product to a social media company, so it is rational for the company to bake these systems in.

only extract information or influence us without our noticing, but which also rely on an increased (false) sense of privacy to condition an increased openness to disclosure or influence.

My argument will be helped by a thought experiment I will call 'The Entity'. The Entity is an ideal Bayesian rational maximizer whose goal is to get the best possible beliefs about consumer behavior in order to generate the maximum utility for itself by influencing consumer behavior in action guiding contexts or by building dispositions in consumers. The analogy between The Entity and an agent in the real world is not, I think, a tenuous one.[2] Because membership in a social kind will turn out to be one of the most important inferential facts for The Entity, I will try to show that the paternalist prescriptions generated fall into distinct "pure" and "impure" cases (Dworkin 1972). Pure cases for intervention are those in which the ones whose behaviors are limited by the intervention are the same as the persons who are protected (think limitations on fingerprinting in action guiding contexts).[3] Impure cases for intervention are those in which the ones whose behaviors are limited are not the same as the persons who are protected (think disclosures of facts that constitute a commons of inferential data about persons in general).[4]

---

[2]At the time of writing, this analogy is further vindicated by a recent testimony from a Whistleblower at Meta, Sarah Wynn-Williams. See especially the final ten minutes of the hearing, where the Wynn-Williams discusses the practice of profiling the emotional state of teenagers of thirteen to seventeen years old and of young mothers, to serve advertisements when self-esteem is detected to be low. (C-Span 2025)

[3]Fingerprinting is a common practice in investigative police work and is used analogically in forensic cybersecurity or data science. In the physical world, all humans have a distinct pattern on their fingers which can be recorded and used to recognize the owner at a later date. Fingerprinting in the digital world refers to the set of facts about an agent which can later be used to re-identify that agent, or in some cases, correlate similar agents given similarities in these identifying facts. Digital fingerprint is a great deal more inferentially powerful than its physical counterpart because the information that constitutes a digital fingerprint has inferential uses about its owner or similar fingerprinted individuals, while the swirl of skin on a person's finger tells us little if at all about their behaviors. Action guiding contexts are time slices in which an agent takes stock of relevant information, deliberates, and makes a choice. Relevant examples might be comparing items in online shopping, or deliberating whether to like or leave a comment on a piece of content.

[4]Some information about persons is a commons. A commons is a scarce resource that must be shared by two or more agents. When a toddler asks me "what is a teacher?", the information I supply in 'teaching' the toddler about the concept "teacher" provides inferential data for the toddler about teachers in the world. Suppose I am a teacher and want my child to know that teachers teach people, that they generally love learning, and that they make less money than they ought to. This is relatively innocent: I want my child to know my profession well. But suppose my child goes to high school and begins to make offhanded comments about their instructor's economic status. Here I gave the toddler access to information about a group I am part of in order to enable inferences about me, but gave up inferential data about others that led to some harms.

If we are right to worry about The Entity or to recommend intervention in impure cases, then we ought to support the thesis in contemporary philosophy of privacy that groups are a legitimate locus for informational privacy rights just because the information we are protecting about individuals can pose harms to other members of the relevant groups (Floridi 2016). More will be said on the topic of "Group Privacy" in the section on The Entity.

Privacy as a concept has a storied and at times contentious semantic ascension. It picks out a number of states of affairs that are contingently valuable. Plausibly humans want to be private at least some of the time, but I will work narrowly with the scope of 'persons in liberal societies.' Persons do and ought to value privacy because, as Rössler (2005) argues, privacy and autonomy are related, and we value our autonomy.[5] In particular, I will use the relatively broad definition of privacy as "the control over access by others," and as divided into the three vectors: informational, decisional, and local privacy (Rössler 2005, 43). Informational privacy represents control over access to information, decisional privacy represents control over access to our exercise of autonomous action, and local privacy represents control over access to physical spaces. I will be using G. Dworkin's definition of autonomy (Rössler 2005, 53) and Feinberg's (1984) definition of authenticity and offense to make sense of the experience of privacy and the distinct feeling of a diminishment in privacy.

I will begin with the necessary preliminaries: definitions of privacy and autonomy. Then I will try to show that in every experience of diminished privacy, there is an experience of diminished autonomy. Hence if there is a diminishment of privacy that is not accompanied by a felt diminishment of autonomy, the violation of privacy will not have a corresponding experience.[6] At this point, I introduce my 'The Entity' thought experiment which will be used throughout. I then point to the special quality of feeling

[5]As is well known, autonomy is the cornerstone of the ideal liberal society. One thinks of the Millian Harm Principle which proscribes all interventions except those which protect the autonomy of persons.

[6]The important quality of this corresponding experience is its motivating quality. I take it that if we value privacy, then feeling less private has a way of motivating us to want more privacy. Put in another way, a violation of our privacy can only count as a normative reason to shore up our privacy if we experienced the violation and can express it as a historical fact.

private that conditions our openness to disclosing private information and suggest that the commodification of privacy can have this effect.[7] If I am right, then we might experience ourselves as existing in the inversion of Bentham's panopticon. I will speak on this at length in the related section, but in essence we do not today experience ourselves as subordinated by a surveillance state. Rather, we sometimes get to enjoy the voyeurism of the surveiller, and in doing so, become easier to watch. I close by distinguishing which "pure" and "impure" cases for intervention I think are generated by the disconnect between experiences of privacy violations and real privacy violations, and I show how this bears on endorsements of "Group Privacy." It turns out that pure and impure cases have a great deal of overlap, which overdetermines the justification for paternalist policies about privacy by diversifying the kinds of reasons that count.

Autonomy and Privacy

In the present discussion I will be focusing on a right to privacy defined as a right to control over access by others. Then, privacy is a predicate that describes something this control is predicated over. My home is private if I can deny others access to my land, my phone is private if I can deny others access to its contents, and my body is private if I can deny others control over its functions.

Rössler defines privacy "as control over access by others and thus as protection against unwanted access from other people, this access being defined both as actual physical admission (to spaces) and as metaphorical access to one's personhood, in the sense of access to information on the one hand and in the sense of possibilities for intruding and intervening in a person's behavior on the other" (Rössler 2005, 44). This definition is particularly useful because it is broad enough to account for the several desiderata of

[7]By the commodification of privacy I just mean the process of making privacy appear desirable on its face, not for the goods we estimate it gets us. What I call commodity privacy occurs when a company markets a product as private because privacy is a fad and therefore labeling a service as 'private', whether or not this includes meaningful additions to protecting privacy, can be expected to increase sales. We should not understand the significantly large group of people who has fallen prey to paranoia in the digital age as victims of commodity privacy: if the service does get them this good, then there is no problem. Commodity privacy arises when privacy is desired for what it ought to do, but not what it does.

privacy as a concept. It can account for the type of privacy we refer to by a sign on an empty plot of land "PRIVATE PROPERTY," or the implicit rules governing certain spaces like confessionals. It can also account for the type of privacy we mean when we say that the contents of our phone are private, or when we express a desire to keep gossip shared at the workplace private. It accounts very well for the fact that we say, e.g., "whether I will carry my child to term is my private matter,"[8] or "whomever I choose to marry is a private affair."

These examples can be further divided into three categories of privacy protections, or what we might call vectors of privacy violations. When we map the extension of possible violations in these categories, we get what we might call the attack space of privacy. These three categories are: "privacy of place, privacy of information control, and privacy of decision or action" (Rössler 2005, 44). For short, I will refer to these categories as local privacy, informational privacy, and decisional privacy. When, for instance, a nosy neighbor enters my garden, they have committed a violation along the vector of local privacy. When, moreover, they see that I am growing illicit plants in my garden, they have committed a violation along the vector of informational privacy. When, finally, they choose to report my secret crime to the authorities and my crops are seized, they have committed a violation along the vector of decisional privacy.[9] Here we see that a particular violation in one vector (snooping in my garden) can constitute an accidental violation in another vector (spotting my illicit plants). There are many configurations of these vectors, including both cases of overlap and cases in which a vector is violated in a vacuum. Consider for instance, if I share a harmless secret with a poorly chosen confidant, who mindlessly spreads this fact to whomever they should stumble upon. Clearly my informational privacy has been limited (for I was

[8]Though it might not arguably settle the matter whether the unborn foetus has a privacy right in conflict with the mother's.

[9]I use a purposefully ambiguous example here to demonstrate that what may, in a pure sense, fall under the predicate "private" might not deserve protection in the larger scheme. It is up to local systems to determine limitations on our local, informational, or decisional privacy–but these limitations still constitute a diminishment of that privacy.

unable to control access as I intended), but my local privacy has not been limited, nor has my decisional privacy (because the secret was harmless).

I follow Rössler in thinking that the value we place in autonomy explains much of the value we place in privacy. On her view, "we value (certain forms of) privacy for people in liberal-democratic societies because without the protection of privacy it is not possible to make sense of the idea of individual freedom and autonomy that is basic and central to liberal democracies" (Rössler 2005, 44). There are two ways a lack of privacy can pose a problem on this view.

The first is what we might call the behavioral problem of privacy– or BPP. The thesis is that in every situation where we have to make a choice, there are certain explicit and implicit rules, and we use these rules to make expectations about how other persons will react. As a general trend, when we expect that our actions are exposed to more people, or to persons with less good will, we are more reticent in our actions. Obversely, when we expect that our actions are more secret, or exposed to persons with more good will, we are more open in our actions. So a significant shift in the amount of privacy we expect will necessitate a fairly major change in our everyday behavior. This is also the thesis of Bentham's Panopticon on Foucault's (2008) interpretation: the continual expectation of surveillance by an unspecified other has a powerful and self-perpetuating subordinating force.

The second way in which lack of privacy can pose a problem is what we might call the authenticity problem of privacy– or APP. This problem is closely related to BPP, but first I must define autonomy and authenticity in order to explain its significance. With Rössler, I will be using G. Dworkin's famous definition of autonomy:

> A person is autonomous if he identifies with his desires, goals, and values, and such identification is not itself influenced in ways which make the process of identification in some way alien to the individual. (Rössler 2005, 53)

The condition of autonomy is dependent on the individual's ability to identify their current and historical actions with their beliefs, attitudes, and desires. In other words, autonomy depends on a close relationship between 'who I am' and 'who I think I ought to be.' This identity condition which constitutes the first half of Dworkin's definition we might call the individual's authenticity. An individual is authentic if and only if she identifies with her own desires, goals, and values. The second half of Dworkin's definition determines whether, following the condition of authenticity, the individual's acts are autonomously or heteronomously determined.

There are a number of ways in which the process of self-identification can be made alien to an individual. The most salient is the application of explicit or implicit coercive forces on their decision making process. We might plausibly define a coercive force as a fact in the world that creates the belief that a certain desirable action would result in sanction. Say for instance that I live in a very impermissive society and to cope, I desire to make jokes about political figures. In fact, say my identity comes to depend a great deal on the joy both I and my audiences get from these jokes. Now consider that even though I have been putting on comedy shows in secret so far, a journalist has come to my latest show and announces they will record and publish the event. Reasonably, I can expect sanction were I to go on with my usual repertoire. The journalist, presumably, has a right to be here and so I cannot restrict her access to the space or to the jokes I will make. Hence I choose to limit my behavior; I become more reticent; my jokes are tame and everyone leaves relatively disappointed.

Here my behavior been limited by a diminishment of my privacy. This is an example of what I will call the Behavioral Problem of Privacy – BPP. But additionally, there is something distinctively inauthentic about how I was coerced to respond to changes in the context's conventions. I am not pleased that I had to censor myself and were this process to repeat at length, we would be unable to declare my comedy shows displays of my autonomy. This is what we might call the Authenticity Problem of Privacy

– APP. So long as autonomy, or more minimally, so long as authenticity is something we value, we require the ability to place reasonable limits on access to persons, places, and information.

A less obvious example from Feinberg (1984a) is informative. Feinberg draws a connection between privacy and the impact of 'offense' to autonomy. The paradigm case goes like this. You are sitting on the bus, engaged in your daily commute from work to home. You look up from your book and survey the cabin around you. To your shock, you find a man not much older than yourself sitting in the nude, completely unbothered. There is something in this case that is distinctly experienced as a violation of privacy. There are two salient aspects of the man's action. On the one hand, he is nude–totally conventionally inappropriate for a bus. On the other hand, he seems unable to cognize the relevant convention about nudity–he is unbothered. Locally speaking, while the man *qua* man is permitted to be in the bus cabin, the man *qua* nude is totally prohibited. Informationally speaking, it is the nude man's privacy that is unwittingly violated. This must be understood as arising from a normative misunderstanding. The audience feels that the action (nudity) ought to be protected against prying eyes. But the putative victim does not feel victimized at all. This leads to the decisional violation: you feel embarrassed *on their behalf*, and what is more, the radical misunderstanding about norms between the two of you is shockingly alienating. You feel offended and this makes authenticity hard in the present moment. This speaks to the complexity of possible experiences of diminishments in privacy and should be taken to demonstrate a less obvious aspect of the relationship between privacy and autonomy.

So I have argued that privacy is control over access by others. I have also argued that we value privacy because we value autonomy. There are two threats that a lack of privacy poses to our autonomy: the behavioral problem and the authenticity problem. If I am right, then we can use the characteristics of these problems to predict which kinds of privacy violations have corresponding and saliently motivating experiences. Obversely, if we can point to diminishments of privacy (failures to limit access) that are not accounted for in experiences of changes in expectations (BPP) or alienation (APP), then we can expect

that these diminishments of privacy will not have corresponding experiences at all. This point will see its full demonstration in the next section, but one example before I continue.

Recall the nosy neighbor entering my garden. Suppose that my neighbor enters my private garden and I catch him in the act of discovering my illicit plants. In this case, I have discovered that he failed to obey restrictions on his access to my garden, as well as restrictions on his access to the information about what I grow in my garden. I now tell him off: I threaten him not to inform the authorities about what I am hiding in my garden. If he chooses to respect my wishes, there has still been a sort of violation of my decisional privacy just because I was forced to respond to an alienating situation (the neighbor's invasion of my garden). Likewise, if he chooses to report me anyway, then not only will the authorities intervene on my decisional privacy, but there is a sense in which the neighbor orchestrated that invasion. But all of this counts only just in case I catch my neighbor's incursion. Suppose instead that my neighbor enters my garden several times over the course of a year and I never discover him. Meanwhile, he has been accumulating evidence of my illicit gardening project. At some terminal date, he chooses to massively disclose the accumulated evidence to the authorities, who now impose serious sanctions on me. Clearly every incursion into my garden was a violation of my privacy, but there was no related experience of its diminishment. Likewise, the taking of photographic evidence, or snipping some samples of the crop was never discovered by me, and therefore never experienced by me. No less, my decisional privacy was ultimately violated and I experienced this as a very serious infraction. I had no recourse to secure my privacy while the danger was present and I might even now be unaware just how the authorities became informed that they ought to intervene.

While everyday people often have reasons to worry about their privacy for the sake of interpersonal relationships, it seems that the really consequential kinds of violations of our privacy we ought to worry about are the ones that are of a similar form to this example. They have the characteristic of being secret until it is too late. This is a characteristic of how modern technology has evolved and is

not, I think, a problem that can easily be reconciled with tools for the individual to protect themselves with. We are often "nominally empowered but actually unable" to control access to our privacy in many of the important contexts posed by today's technology (Milano, Taddeo, and Floridi 2020, 8). We ought to look seriously into what, if any, paternalistic measures ought to be taken. To that end, I now move to introduce 'The Entity' in order to distinguish the "pure" and "impure" cases for intervention.

The Entity

Imagine there is an ideal Bayesian rational maximizer with a considerable amount of capital at its disposal: call it 'The Entity'.[10] The Entity will try to get the best possible beliefs in order to make the best possible predictions and get the highest possible utility for itself. Let us imagine The Entity exists in a modern, capitalist framework where the relevant data are consumer behaviors at large. When the entity gets sophisticated enough at predicting consumer behavior, it becomes possible for the entity to influence consumer behavior in reliable ways. Suppose that when the predictive power reaches a high enough degree of sophistication (call this an asymmetry), it becomes possible for the entity to reliably influence consumer behavior without detection.[11]

For ease of understanding we might limit the entity's area of influence to Recommender Systems – RS.[12] We can define an RS to be "a class of algorithms that address the *recommendation problem* using a content-based or collaborative filtering approach, or a combination thereof" (Milano, Taddeo, and Floridi 2020, 4). The recommendation problem is just the task of finding good items for the user. Recommender Systems have their own ethical challenges but we might charitably conceive of the average RSs as well-

[10]According to instrumental rationality, an agent is rational if and only if it chooses to perform the action that its beliefs suggest will get it the most utility with respect to the agent's preference structure. The Bayesian rational agent makes use of conditional probability to form beliefs about expected utility under lottery: cf. Bayes' Theorem.

[11]A reasonable analogy might be made between the entity and any number of agents in the world. One might think of governments or corporations. I note that the important feature of the entity is it is an ideal maximizer: real world agents might not be, or need not be, though sophisticated computer guided agents can be.

[12]With the caveat that a real world analogy would involve a significantly more sophisticated attack space than a single RS, or a combination of multiple RSs.

meaning utilitarians. They operate by profiling a user, applying an algorithm to predict utility, and posing content based on those inferences, which is then evaluated based on user interaction with the content to update the profile or algorithm. Those of us who use social media interact with RSs every day, but RSs are present in many other applications. For instance, travel booking web applications, online shopping, and subscription television services.

Now suppose that while the average RS is a well-meaning utilitarian, the entity will use the RS as a space in which to generate utility for itself in addition to or in exclusion of utility for the user. If a user detects that their RS is either not directly serving them, or worse seems to be doing something malicious, it seems reasonable to suppose that the user will remove themselves from that attack vector. On the present framework we can account for this as having experienced a violation of decisional privacy because the RS generated a sequence of recommendations that, when understood as 'intended good recommendations' created a sense of inauthenticity.[13] Or, the RS generated a sequence of recommendations that were so specific to an aspect of the user's personality that it cannot be understood except by inference from that characteristic, which creates an experience of violated informational privacy. Likewise, if we were to indicate to an RS through our actions or explicitly through a survey that we enjoy a certain activity, say, cycling, and later detected a drastic increase in cycling advertisements outside of the RS, we might suppose our data was sold and choose to cut off use of the RS to preserve our informational privacy. Supposing that The Entity has reasonably good beliefs about consumer behavior, we can expect that it will, where possible, avoid reducing the population of consumer targets from which to extract utility. I think it is then rational for The Entity to try to generate an asymmetry in predicting

[13]For example, say you use a certain social media application to get some quick and easy enjoyment via short form videos, but instead of the usual sugary gruel, you are met with a sequence of overt advertisements. This sequence repeats at length and we feel as if our ability to choose the content we consume has been eliminated. Or otherwise consider what if the RS begins recommending what one might consider offensive content, like gore or hardcore pornography: in these cases, the thinking that the RS is representing your personal identity but inability to identify with it poses the authenticity problem of privacy.

power (i.e., to be able to influence without being detected) before beginning to meaningfully weigh its own utility over the user's utility.

Plausibly, if we are consumers, then we do not want The Entity to develop an asymmetry in predicting power. Where RSs (or things that look like RSs) are supposed to be well-meaning utilitarians, a smart and malicious third party is able to use the content recommendation context to do a number of things like nudge us ethically, or recommend a product, or hide certain facts from us, or adjust our attitudes towards salience more broadly.[14] This looks a great deal like a violation of our autonomy, which is something we ought to care about.

The question is then how to create the asymmetry in predicting power and how to prevent its creation. This question is statistical in nature: all that is needed is a relatively large pool of data and relatively sophisticated inductive logics to do the inferential work. The literature sometimes calls this "collaborative filtering" which can be defined as "the system [constructing] a model of the user based on the data it has gathered on other user's interactions" (Milano, Taddeo, and Floridi 2020, 9). I am thinking here of the strong inferential power of membership in groups or social kinds. I will take a group to mean just a set of individuals designed at a level of abstraction – LoA (Floridi 2016, 1). Designing a group at an LoA just means choosing a piece, or several pieces of inferentially rich information and constructing a set of the tokens the relevant fact is true of.[15] Let me give a very low level example to illustrate.

Eight hikers enjoy promenading in a nearby forest. This forest in particular is in a national park and the agency that maintains the land has installed surveillance cameras. As the eight hikers are walking,

[14]Many persons today, likely wrongheadedly, use social media to get their news. This reliability of the social media application to provide factual news some of the time might trickle down implicitly into other content.

[15]This process of design, interestingly, does away with some type errors. For instance, if I am interested in doing some research about who has been involved in the construction of churches in Lodi, Italy, then there are at least two groups I ought to construct: Churches, and Church Builders. In the Church group, it might contain *"la chiesa di San Lorenzo Martire"*, and "*la chiesa di Sant'Agnese"*, but would contain neither *"Ostinelli Patrizio"* (an eyewear shop), nor "the fresco by Callisto Piazza in the apse of San Lorenzo's church". Patrizio's shop is of a similar LoA (buildings) but of a different kind than is thematized (churches), while the fresco is at a different LoA (frescoes, buildings). Likewise, if "Callisto Piazza" can be considered a Church Builder (for painting the fresco), then "Callisto Piazza's hand" would not fit in the Church Builders group–the LoA differs.

a number of facts can be discovered about them. Three of the hikers are wearing very expensive backpacks with built in water storages. One of these hikers has a rather expensive GPS watch on his wrist. Two more of the hikers are wearing cheap day bags, and the other three are all wearing blue matching bags with a logo. The surveillance cameras are also recording audio and pick up some further facts. The man with the GPS watch is looking to purchase a new mountain bike. Meanwhile the other two with nice backpacks are listening to him explain how cool the newest bikes are. The three with matching bags have been heatedly debating which Italian grape makes the best bottle of wine under thirty dollars.

Now suppose that there are several touch points where items could be sold to these hikers at a later date. We do not know their names, but we know some facts that are correlated with attitudes towards certain products. If we know in the first place that we are working with a hiker, we might recommend a backpack. If we know what backpack is already owned by the hiker, we might gain further information about their attitudes towards e.g., bikes or wine. Furthermore, if we know that a given hiker has both a GPS watch and a nice backpack, we might suppose that (within our closed sample) this hiker must be the bike enthusiast. For application to The Entity, scale this model up to the size of the human population on earth. Every fact about a person and every correlation between facts is inferential data. At such a large population as can be collected via collaborative filtering, it is relatively trivial to build up a behavioral model, and to fingerprint an individual against the behavioral model.[16]

This threat has led to a movement in the contemporary philosophy of privacy that endorses groups as the locus of privacy rights. Call this "Group Privacy." Floridi's own example does an excellent job of illustrating the premise:

> Yet even from a nominalist perspective, we should acknowledge that both friendly and hostile users of big data may not care about Alice at all, but only about the fact whether Alice, whoever she is, belongs to the group that regularly goes to the church, or mosque, or

[16]Collaborative filtering is content filtering that is compared between users. It goes off the assumption that there are common markers of utility between users which are discoverable by comparison. The rendered profile can be further divided into the social kinds with greater or lesser commonality, which can later be used for behavioral predictions.

> synagogue, uses Grindr, or has gone to a hospital licensed to carry out abortions, or indeed shares a feature of your choice. In military technology, Alice is hardly ever a High Value Target, like a special and unique building. She is usually part of a High Pay-off Target, like a tank in a column of tanks. It is the column that matters. (Floridi 2016, 19)

We might choose to value Group Privacy over Individual Privacy because individuals are relatively uninteresting except in rare cases, but groups tend to have at least one or two interesting parties as members. Put in another way, it is a great deal like putting smaller targets on a larger target in a game of darts. The bullseye is the individual, but if you are dealing with groups, then you can get some yield by hitting anywhere in one of the smaller targets. I think the column of tanks example is excellent, but one further example Floridi gives is informative. "Imagine a small departmental library. We need to move it from one building to another. […] We try to lift the pile and notice that it has now acquired a property that none of the books has: it is too heavy to be moved by a single person, despite the fact that each book in it is reasonably small and light" (Floridi 2016, 9).

Recall the definition of privacy as control over access by others. We can reasonably conceive of cases for each vector of privacy. A group collectively owns a factory and restricts access to its interior without express permission. A village deliberates and democratically elects to move further away from the volcano despite attempts to sanction such an emigration by the local authorities. A group of similar looking men in one sector of a totalitarian regime agree not to disclose the fact that their community belongs to a persecuted religious denomination.[17] Notably in this last example, if even one of these men were to disclose this fact, it would incriminate the rest of the community.

Let me prefigure my conclusion here before introducing one final point and example. Feeling less private has a way of motivating us to want more privacy. But as I have shown, not all legitimate claims to

[17]Importantly, the kind of Group Privacy Floridi advocates places the right on the group itself not the individuals taken severally. I do not think this is in tension with the idea that individuals taken severally are owed strong privacy rights. The book stack example illustrates this well: while we call the 'stack' of books heavy, really if you placed the books on my arms one by one, the books thought severally would become quite heavy even as they become the stack. The books do not combine into something more than the sum of their parts.

privacy are associated with experiences of diminished privacy. This cuts out the motivating experience and seems to prescribe a turn towards paternalism. My closing project will be to map the "pure" and "impure" cases for intervention. Pure cases are those in which the list of persons restricted by paternalist means is the same list of the persons who are supposed to benefit from those means. Impure cases are those in which the lists are not identical (Dworkin 1972, 68). Pure cases, I will try to show, are points where an individual is fingerprinted by an RS. Impure cases tend to overlap with pure cases, but are constituted by contexts in which individuals disclose personal information with inferential value about other individuals, which means the pure and impure reasons do some work to overdetermine paternalist prescriptions.

The Inverted Panopticon

I would like to make a digression about the assertion of privacy, or conversely the non assertion of privacy. Here we may think of two very opposite places: the confessional and the panopticon. In his essay *The Rhetoric of Privacy*, K.J. Peters advances the claim that "[p]rivacy is and always has been a rhetorical organization of space," moreover that "[p]rivacy is not a location or demarcation that lies in wait for a speaker and a listener. Privacy is a rhetorical assertion, and in the hands of a physician, psychologist, or journalist, privacy assists the pursuit of knowledge" (Peters 1998, 346–47).[18] The essay provides a number of fruitful examples but the confessional is the most informative for my purposes.

He writes, "[t]o investigate the hidden soul and leverage its closely held contents from the penitent, the Catholic Church, in 1215, asserted and formalized the private place of confession" (Peters 1998, 353). For Peters, the significance of the sense of security provided by privacy is not *as* sense of security. Rather privacy is and can be asserted in a space in order to engender a sense of security and

[18]Literally, for the physician, psychologist, or journalist, the supply of testimony is determined in large part by the degree of privacy experienced by the one testifying. Think of parents listening in to a doctor ask about a child's sexual or drug practices, a psychologist asking about trauma in front of the abuser, or a journalist requiring that a whistleblower's name be disclosed with the testimony rather than be kept anonymous.

promote the disclosure of information. A feeling of local or decisional privacy in a context can change our attitudes towards how much informational privacy we need to secure for ourselves. Peters continues, "the intimacy and solemnity of confession eased hesitation and moved the penitent to 'humility and true contrition of the heart' thereby assisting in securing a good and complete confession" (Peters 1998, 355). The correlation between the sense of privacy within the confessional and the true humility and honesty of the confession is only preserved if the priest reliably maintains the boundaries of privacy asserted by the confessional. Were the confession to be later violated and the confessor to discover themselves to have been harmed along some vector of their privacy, future confessions would become either rare or untrustworthy.[19]

The point I would like to make here is that along with the behavioral problem of privacy's reticence dimension, there is also an openness dimension. Contexts in which privacy has been reliably asserted are contexts in which we are accustomed to acting less carefully about our privacy because we expect that our desired safeguards are already in place. Now let me introduce, with Foucault, the concept of Bentham's panopticon before returning to the significance of this point. I quote here in full:

> Bentham's *Panopticon* is the architectural figure of this composition. We know the principle on which it was based: at the periphery, an annular building; at the centre, a tower; this tower is pierced with wide windows that open onto the inner side of the ring; the peripheric building is divided into cells, each of which extends the whole width of the building; they have two windows, one on the inside, corresponding to the windows of the tower; the other, on the outside, allows the light to cross the cell from one end to the other. All that is needed, then, is to place a supervisor in the central tower and to shut up in each cell a madman, a patient, a condemned man, a worker or schoolboy. By the effect of backlighting, one can observe from the tower, standing out precisely against the light, the small captive shadows in the cells of the periphery. (Foucault 2008, 5)

[19]Much being at stake in an honest confession as it is prerequisite for forgiveness of sin.

In short, the Panopticon is a prison that asserts its totalitarian control by means of continual surveillance by an unspecified other. The prisoners are made to alter their behavior into the space of actions thought to be permissible by the watcher from the tower.

> The major effect of the Panopticon: to induce in the inmate a state of conscious and permanent visibility that assures the automatic functioning of power [… and] to achieve this, it is at once too much and too little that the prisoner should be constantly observed by an inspector: too little, for what matters is that he knows himself to be observed; too much, because he has no need in fact of being so. (Foucault 2008, 6)

To echo the language of privacy as a rhetorical assertion, the Panopticon asserts a rhetorical space of surveillance. The inmate need not be surveilled so long as she perceive herself to be watched: the panopticon makes explicit the implicit structure of a totalitarian surveillance state.[20] Much of the fear-mongering about privacy today echoes the idea of the panoptic, Orwellian police state. Notice that even in the classical formulation of the Panopticon, the interesting subject is the individual, and in fact every individual. While a panoptic society is increasingly possible given technological advancements, we by no means live within a panoptic society as conceived by Foucault. But as things stand now, the individual's behavior is insufficiently interesting.[21] In Floridi's words: "There are very few Moby-Dicks. Most of us are sardines. The individual sardine may believe that the encircling net is trying to catch it. It is not. It is trying to catch the whole shoal. It is therefore the shoal that needs to be protected, if the sardine is to be saved" (Floridi 2016, 20).

How should we interpret the task of catching the whole shoal? I argue that today the strategy is very different to the one invented by Bentham and discussed by Foucault. Today we live in the very inversion of the Panopticon. Recall that with respect to the behavioral problem of privacy, there are two

[20]This function can be echoed in a more everyday way by the installation of, e.g., video cameras that do not actually record, in stores or on front doors, in order to dissuade burglary.

[21]Let alone that we are not positioned to be coerced into basically a capitalist prison where the watcher owns the means of production and asserts surveillance as means to maximize production. The assertion of a Panopticon needs to entail a strong enough force not to be broken while it is asserting itself. Remember that the point of the system is to perpetuate its dominating force even if it is not at a given moment exerting that force–the force must exert itself. If sufficient momentum is not picked up, then the system does not work.

directions: reticence or openness. The Panopticon asserts a context of surveillance on the inmates in order to create dispositions of reticence. Moreover, it is reticence that is sometimes unnecessary because it is only the feeling of surveillance that is perpetuated, but not the surveillance itself. The Inverted Panopticon works in directly the opposite way: it asserts a context of privacy on the inmates in order to create dispositions of openness. Just like the original Panopticon, this privacy need not be reflected in fact, so long as it is a felt privacy.

In the Inverted Panopticon, the goal is to facilitate the disclosure of information by creating a sense of intimacy. Furthermore, and I will develop this shortly, the inmate does not need to perceive themselves as an inmate. In fact, the Inverted Panopticon does its work best when the inmate perceives themselves as the watcher. Taken even further, I might suggest that a degree of voyeurism is naturally exciting to humans: individuals are interesting to other individuals, not just groups. Let me give a recent example.

The social media application, Instagram, has a feature for watching short form videos called "Reels."[22] Users can watch the video and have the option to interact with it in largely three ways. First, they can like the video by double-tapping the screen or clicking a button.[23] Second, they can leave a comment on the video, which can be accessed by others (and liked or commented on by others), by accessing another button on the side. Third, they can share the video with another user of the application, or send an online link to someone via, e.g., a messaging application or email.

A recent feature builds on Reels and introduces a different way to proceed through videos. Where customarily an RS will build a profile (colloquially called "The Algorithm") and recommend videos to the user based on that personalization, the new feature allows individuals to scroll through a list of videos

[22]Social Media technologies iterate rapidly. The following is true as of April 17, 2025, and version 376.0.0 of Instagram for iOS 18.1.

[23]Liking a video increases a 'like' counter on the video and leaves a small red heart with the username, visible to other users who follow that account.

liked or commented on by about ten to twenty individuals in whom the RS thinks the user is interested.[24] I think it is plausible that this feature creates a rhetorically private space. This is just to say that where formerly, leaving a like on a video was done more or less emptily to signal enjoyment, now more complex signaling is able to be done because one can expect their friends to notice their likes. For instance, one can leave a like on something political to endorse the content, or on a questionable piece of content to express something subversive. Moreover, it is not uncommon practice to take up these subversive acts with friends (by sending a photo of the screen to the friend). In other words, while using this feature the user gets the sense of being put in a room where content is shared between individuals they think are interesting and, vice versa, they think are interested in them.

This point is to say nothing of the fact that using social media and interacting with content very rarely, if ever, generates immediate experiences of limitations in our autonomy. Liking a video is relatively innocent so long as the content of the video is correspondingly innocent. Likewise, sending a video to a friend is a totally everyday and innocent thing to do. But every instance of interaction with a piece of content strengthens a user profile and, moreover, strengthens "collaborative filtering" models.

The experience of scrolling through videos on a phone is also not unlike flipping through surveillance cameras, or modifying one's view from the watch tower of the Panopticon. Here, however, individuals are choosing which facts (arguably fictions) about their lives to disclose to the hungry watcher. So I think it is plausible that when we use social media, or other services with a kind of reciprocal interpersonal interaction via RSs, we find ourselves in the Inverted Panopticon–really in the position of the inmate, but with all the enjoyment of the watcher. Social Media is not the only context that can assert this rhetorical space, either. For instance, we are very often sold privacy on the label of services. This can be accompanied by fear-mongering. Say for instance a given email provider advertises itself as private. What it means, in some cases, is that emails are sent with End-to-End Encryption. This safeguards against

[24]Usually these are friends or individuals to whom videos are sent by the user.

what are called Man-in-the-Middle Attacks (analogous with someone snooping your mail while it travels to its destination), but it might not mean that the email provider has an obligation not to sell certain information from the inbox, like subject headers. This data is inferentially rich and therefore we ought to care about its disclosure. But not only do we not experience this disclosure, we are (plausibly) told that we are more private than ever in order to prime us to be open to disclosure. In feeling more private, we are rendered less private than ever.

Privacy Paternalism

Feeling less private has a way of motivating us to want more privacy. But as I have shown, not all legitimate claims to privacy are associated with experiences of diminished privacy. Furthermore, sometimes we even feel more private than ever when we are in the process of having our privacy violated. This cuts out the motivating experience and seems to prescribe a turn towards paternalism.[25] In closing, I will try to do some of the work to map out the topography of relevant interventions. In particular, I distinguish between "pure" cases for intervention and "impure" cases for intervention. I think this distinction is important for getting a solid grasp of what a privacy protection really means in terms of dedicating political capital to create legislation. Pure cases are those in which the list of persons restricted by paternalist means is the same list of the persons who are supposed to benefit from those means. Impure cases are those in which the lists are not identical (Dworkin 1972, 68).

Pure cases are relatively straightforward: if someone wishes to walk off a bridge, we might choose to intervene in order to protect their life. Hard cases might generate more discourse over whether a particular intervention is justified: smoking kills, but many people value smoking–should we go out of

[25]A further notable division is that between interventions that contradict what we believe is best for us and interventions that adhere to what we believe is best for us but which we are not situated to choose. In other words, sometimes we have the right desires but lack the true beliefs or hold false beliefs that fail to motivate us to take action on our own behalves. In many cases the interventions here advocated will be of the kind easier to justify, where we simply fail to notice the harmful situation. The more complex contexts are the ones in which we really do desire to give up our individual privacy in order to receive benefits like access to (monetarily) free email, or social media.

our way to stop smokers from smoking? Maybe new smokers, or young smokers–but what about long-term habitual smokers?

Impure cases can be justified more narrowly via the Millian harm principle. As is well known, the principle states that "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others" (Dworkin 1972, 64). Let me give an example. Two hikers enjoy spending time on a suspension bridge, while another enjoys sitting by the edge of the cliff and watching them frolic. Along comes yet another hiker, who takes great pleasure in cutting down suspension bridges. This sadist begins their work with great zeal, but the harm principle enjoins the sitting hiker to intervene, and by fisticuffs (the sitter is twice the cutter's size), tragedy is avoided. The cutter's autonomous exercise brought threat against the suspension-bridger's and was (rightly) abrupted by the sitter.

I will focus here on violations of informational privacy. Pure cases for intervention are generated when the person who is threatened is the same as the person whose autonomy must be limited. Hence we might point to instances where a user is directly interacting with an RS and is permitting themselves to be profiled. Narrowly construed, we might want to protect the individual in this case because the RS (hypothetically, in the case of The Entity) might secretly influence the behavior of the individual (intentionally or not). This might be relatively mundane in causing the user to buy this product rather than that product. Or it might be more significant, like shifting the user's disposition towards certain ideologies over the long-term. We also must attend to the fact that an RS, generally, is successful if it reliably recommends content that provides utility to the user. So there is tension between the negative influence, if it exists, and the positive influence, which is existentially necessary for the RS.

Impure cases for intervention are generated when the person(s) who are threatened are not identical to the person(s) whose autonomy must be limited. We can find a great deal of overlap with pure cases because, as I have argued, individual profiling is also useful for collaborative filtering. Plausibly, all

information that is interesting about an individual and is true of two or more individuals is inferentially valuable information. The interesting ethical question is then where the limit lies. Plausibly, we ought to prioritize harm reduction for many people over benefiting a single person, hence we might argue that no individual ought to be permitted to give up a swath of inferential data about persons in general, because the utility the RS can generate for that one person is significantly less than the threat it can pose to the rest of the population. But this view might be too risk averse. There are also RS eudaemonic views that argue we ought to get the best possible RSs (evaluated based on their ability to generate utility) and propagate them in order to deputize (a slice of) our flourishing to a machine (Milano, Taddeo, and Floridi 2020, 6). Certainly big data in medicine has catapulted the ability to research rare diseases, to develop treatments, and to diagnose patients. This is an undebatable benefit, but it comes with the tradeoff, e.g., of possible data leaks which can either fuel targeted attacks or feed malicious RSs.

The sorts of impure cases that do not overlap with pure cases mostly take the form of the proxy secret keeper. Take the confessional example, for instance. It is an obligation of the priest to keep the confession secret, but only for the benefit of the confessor.[26] At a larger scale, we might think of big data in medicine, or any service that requires the storage of mass data. Intuitively, persons whose data is stored in a massive data center has an interest in keeping that data safe within that storage. But as I have argued, this also holds for any person the data bears inferentially on. These cases where data or secrets are stored by proxy take the form of impure cases without overlap with pure cases.

The upshot of this preliminary mapping should be the recognition that because pure and impure cases often have overlap, and because these cases are justified for different reasons, the overlap cases are somewhat overdetermined in comparison. Meanwhile, the proxy secret keeper cases are already intuitively valuable targets and receive increasing protection. The exact question of how, when, and on

[26]Arguably the sanction from God that the priest would receive if he made a disclosure is the kind of sanction we use to enforce a paternalistic measure, not what is protected by the paternalism.

whom to implement paternalistic measures remains hanging and ought to be thematized by applied ethicists or other agents in the privacy advocate community.

This discussion began with the fact that feeling less private has a way of motivating us to want more privacy. But, as I have shown, not all legitimate claims to privacy are associated with experiences of diminished privacy. Interestingly this is because the experience of privacy is related to the experience of autonomy and authenticity, but not all violations of our privacy are immediately or obviously violations of our autonomy–especially along the vector of informational privacy. We are often put in a position from which we are unable to detect violations in our privacy, or from which we are unable to stop them. In another's words, we are often "nominally empowered but actually unable" to control access to our privacy in some of the contexts posed by today's technology (Milano, Taddeo, and Floridi 2020, 8). There might be malicious actors which can seamlessly blend in among otherwise innocuous (and helpful!) Recommender Systems. Meanwhile, we are often sold privacy both in name and in feeling, but less often actually made private from the sorts of violations most of us ought to really care about. Via social media, in a fashion not at all dramatic like the original, we experience ourselves as precisely in the inversion of Bentham's panopticon. Not only are we sometimes unable to detect the violation of our privacy, but commodity privacy can create a false sense of security that has the effect of making us less private. Hence, I espouse a more thorough paternalism.

References

Meta Whistleblower Testifies on Facebook Practices. United States Senate, 2025. https://www.c-span.org/program/senate-committee/meta-whistleblower-testifies-on-facebook-practices/658354.

Dworkin, Gerald. 1972. “Paternalism.” Edited by Sherwood J. B. Sugden. *Monist* 56 (1): 64–84. https://doi.org/10.5840/monist197256119.

Feinberg, Joel. 1984a. *Mediating the Offense Principle*. New York: Oxford University Press.

———. 1984b. *The Moral Limits of the Criminal Law*. New York: Oxford University Press.

Floridi, Luciano. 2016. “Group Privacy: A Defence and an Interpretation.” In *Group Privacy*, edited by Bart van der Sloot, Luciano Floridi, and Linnet Taylor. Springer Verlag.

Foucault, Michel. 2008. “‘Panopticism’ from ‘Discipline & Punish: The Birth of the Prison.’” *Race/Ethnicity : Multidisciplinary Global Contexts* 2 (1): 1–12.

Milano, Silvia, Mariarosaria Taddeo, and Luciano Floridi. 2020. “Recommender Systems and Their Ethical Challenges.” *AI and Society*, no. 4, 957–67.

Peters, K. J. 1998. “The Rhetoric of Privacy.” *The Dalhousie Review* 78 (3): 345–63.

Quine, W. V. O. 1951. “Main Trends in Recent Philosophy: Two Dogmas of Empiricism.” *The Philosophical Review* 60 (1): 20–43.

Rössler, Beate. 2005. *The Value of Privacy*. Cambridge: Polity Press.