20 Manipulation, Real-Time Profiling, and their Wrongs[1]

Jiahong Chen and Lucas Miotto

1 Introduction

Claims denouncing manipulation in the online sphere are pervasive and familiar (Susser, Roessler, and Nissenbaum 2019; Bradshaw and Howard 2018). Most writers, in academic circles and beyond, seem to agree that wrongful forms of online manipulation occur because of – or are at least facilitated by – new digital technologies. Despite such agreement, little has been done to pin down the exact wrongs that come from online manipulation. Like other contributions to this volume, this chapter aims to add to this discussion. We argue that paying closer attention to what we call "real-time profiling" allows us to identify the wrong-making features of an important subset of online manipulative practices; practices which are taken to be "manipulation's future" (Spencer 2020).

The chapter is divided into four sections. The first is where we characterise real-time profiling in socio-technical terms. In the second, we show that real-time profiling is analogous to some forms of interpersonal manipulation and that, for that reason, there is a presumption in favour of seeing real-time profiling as an example of manipulation. The third is where we propose our account of what makes real-time profiling wrong. Contrary to some extant accounts of the wrongness of manipulation (both in the online and offline spheres), our proposed account does not link the wrongness of real-time profiling to covertness, deception, harm, autonomy, or to bypassing individual's rational capacity. As we will argue, real-time profiling is wrong both because it involves what we call "psychological hijacking" and because it works as a gateway to further wrongs. In the final section we explore some implications of our account for the legal regulation of online manipulative technologies. We argue that existing legal frameworks are not fine-grained enough to deal with the wrongs associated with real-time profiling and related forms of online manipulation.

2 The Rise of Real-Time Profiling

Profiling is anything but new in the online world. As an important part of the digital advertising ecosystem, profiling has been a technique widely

DOI: 10.4324/9781003205425-24

employed by different categories of actors across the online marketing sector (Chen 2021, ch. 1). Traditional profiling systems are said to be "interest-based" since they are designed to make inferences about what might interest the user and what demographics the user is likely to fit in. The operation of such systems depends on the ability to observe users' online traces over time, so as to gain an increasingly precise understanding of their preferences. The predictions may change as more data accumulates, but generally, traditional interest-based profiling is mainly about grouping users according to stable and longitudinal features.

In recent years, and largely due to an array of socio-economic factors,[2] the capabilities of profiling techniques have developed beyond some of the technical constraints as seen in earlier days of online advertising. Profiling broke through the boundaries of browser-based tracking, as well as accuracy and speed constraints and culminated to what we call "real-time profiling". The practice of real-time profiling, as we describe it, has two main steps:

1. A private or public agent collects information about an individual's *present status*. This can cover an array of aspects: from the individual's current health status to how hungry, stressed, or annoyed she is. Once the information is evaluated, a profile of the individual's present status is built.
2. The private or public agent then attempts to influence the individual's actions, choices, or preferences in the immediate future based on the constructed profile.

As we can infer from these steps, real-time profiling differs from traditional interest-based profiling in that it is designed to track certain transient and dynamic characteristics of a user and to adjust interactive strategies in real time. Real-time profiling may or may not rely on the long-term construction of a user's profile; the goal is to work out the instant status "on the spot" rather than a relatively static aspect about the user. To illustrate the phenomenon, consider the following case involving Uber (Lindsay 2019; Mahwadi 2018):

> (Uber) It has been revealed that vehicle-for-hire app Uber could implement a technology that enables it to assess users' level of inebriation and battery level. The technology could be used to get users who are in a more vulnerable position to pay more for their standard rides or to push these users to choose a premium ride.

Similar examples of real-time profiling abound. We know of gambling firms that can influence users based on their location (e.g., at sports events, Rudgard 2018), or of smart devices that can influence individuals on the basis of their current level of stress or heart rate (Brown 2018; Shapiro 2016; Alvarez 2017; Charara 2020), or even of eye tracking technology that can be used to influence individuals based on what they are paying attention to at a particular moment (Metz 2016; Valliappan et al. 2020). In principle, as a set of targeting strategies, real-time profiling can be implemented on a variety of technical settings, theoretically including any human–machine interactions. These may include, for example, how online services present information, how smart devices change the ambience of a space, how robots adjust actions towards humans, or even how the urban infrastructures configure resources (e.g., "smart cities" or "smart transport" initiatives). More importantly, different strategies can be combined to make the assessment more accurate and to find the "optimal" way to interact with and influence the profiled individual (Sax 2021).

Each real-time profiling strategy may exhibit its own challenges, which requires specific discussion. Be this as it may, a general inquiry raised by the practice of real-time profiling concerns its moral status. Real-time profiling seems to be the sort of practice that calls for justification. And the reason for it is that, at face value, it resembles instances of wrongful manipulation in relevant respects. In what follows we briefly elaborate real-time profiling's resemblance to manipulation.

3 Is Real-Time Profiling Manipulative?

We need not assume that all instances of real-time profiling are instances of wrongful manipulation.[3] Our goal is simply to show that some cases of real-time profiling and typical cases of wrongful manipulation are analogous in relevant respects. To begin with, consider the following case of interpersonal manipulation:

> (Phil) Phil and Claire, a married couple, earn about the same salary. Phil plans to get Claire to pay for a much larger share of their household's expenses so that he can save up for a trip with his friends. He knows that Claire is much more receptive to his proposals when she is relaxed and after enjoying a good meal. Phil makes a plan: he gives Claire a spa-day voucher and spends the day cooking Claire's favourite meal while she is away. When back, Claire finds Phil at the dinner table, meal ready. After a pleasant dinner, Phil suggests that she pay a larger fraction of the household's expenses. As predicted by Phil, Claire accepts.

Phil got his own way with Claire not by persuading or reasoning with her. Nor did he get his own way by coercing, pressuring, blackmailing, deceiving, or lying to her. His act is – intuitively at least – manipulative. Some features are key to that assessment. First, Phil *attempted to influence* Claire's decision. Second, the *primary motive* for Phil's behaviour was the benefit he would get from influencing Claire (i.e., saving money). Finally, and more importantly, Phil's attempt to influence was specially tailored to take advantage of an aspect of Claire's *deliberative capacity*, namely the fact that she would be more receptive to Phil's proposals at a particular time. We find variations of these features in some definitions of manipulation.[4] However, here we take them not as necessary or sufficient features of manipulation but merely as features which are often salient in core cases of wrongful manipulation; features that would give us a *presumption* that a case is an example of manipulation in the absence of stronger countervailing considerations.

Now, when we compare Phil's influence with an example of real-time profiling, we can see that they share the aforementioned features, which gives us *some* reason to presume that at least some instances of real-time profiling are examples of manipulation. To make this point clearer, consider a hypothetical case of real-time profiling:

> (MoodX) A social media company, MoodX, develops an algorithm that predicts its users' current mood with high accuracy. With the help of the algorithm, MoodX advertises products tailored to users' current mood. Sales of advertised products skyrocket as a result.[5]

MoodX clearly attempted to influence users with its algorithm and advertisement strategy. And we can see that the way MoodX chose to attempt influencing users was part of MoodX's *unilateral* plan: a plan primarily motivated by its benefit to MoodX.[6] We can also see that MoodX's influence was tailored to take advantage of a particular aspect of users' deliberative capacity,[7] namely the fact that their moods affect their buying choices. In virtue of exhibiting these features, we can say that MoodX – and other examples of real-time profiling – can be presumed to be an example of manipulation in the absence of countervailing considerations.

While Phil's and MoodX's influences can be seen as (or presumed to be) examples of manipulation, it is worth drawing attention to one way in which these examples may deviate from some typical cases of manipulation (e.g., doing small favours for others to feel obliged, placing more expensive products at eye-level). Both Phil and MoodX exploited their target's *transient and dynamic features*. The content, message, action, or conditions set out by Phil and MoodX are not simply tailored to their targets on a personal level but more importantly, to a precise point of time when the internal or environmental circumstances have changed such that the impact of their move is maximised. The kind of manipulative influence that we are focusing on should thus be seen as a distinct subset of manipulative influence: one where the manipulator is sensitive not just to *who* to target but also to *when* and *where* to target someone.

Now we may wonder whether, despite their similarities, cases of real-time profiling would be in some sense distinct from interpersonal manipulative influences like Phil's. We believe that there is no difference in *kind*. The obvious difference, when there is one, has to do with the intensity and the scope of the real-time profiler's influence. Real-time profiling happens in the online environment and the profiler is typically either a corporation or a public agent who has information and technological resources that enable constant observation of the target's online and offline activities. In a sense, we can say that the profiler is *always around* (at least when the target is next to the right gadgets) and usually has information about the target which is unfeasible to obtain via everyday interpersonal interactions (in fact, profilers might be more insightful of targets' current status than the targets themselves). The profiler, therefore, typically does not face temporal, spatial, and access restrictions that interpersonal manipulators (like Phil) do. As such, their scope for interference with the target is much wider.

As mentioned earlier, some writers have suggested that this form of real-time interference is "manipulation's future" in the online domain and is bound to become more common (Spencer 2020, n2). Being the future of online manipulation or not, this form of manipulation raises ethical concerns and calls for justification. It is thus worth examining what makes real-time profiling (and related forms of manipulative practices) wrong when they are wrong. That is the task we take up next.

4 Why Is Real-Time Profiling Wrong?

As we have seen, there is a presumption in favour of seeing some cases of real-time profiling as instances of manipulation. Whether *all* instances of real-time profiling fall under a properly regimented concept of wrongful manipulation is not something that interests us. Instead, we are interested in what makes some cases[8] of real-time profiling pro tanto wrong when they are wrong.[9] In this section, we analyse the features of real-time profiling that make it wrong.

Let us return to MoodX. At face value, MoodX did something *pro tanto* wrongful. A few elements in this case can help explain why. From the description, and from what we have said about it in the previous section, we can infer that MoodX had a unilateral plan to profit from the sale of products and took steps towards making it successful. What seems to make MoodX's action wrong, however, is not simply that it had a unilateral plan and acted on it, but the *way* in which MoodX implemented its plan matters.

Recall that in the previous section we said that MoodX's influence was tailored to take advantage of a particular aspect of users' deliberative capacity. This can be fleshed out in more specific terms. What seems to be the case is that MoodX *hijacked* users' psychology; it worked out a way whereby users' own psychological states – that is, their moods – served MoodX's unilateral plan. And the act of hijacking someone's psychology, we submit, is an essential part of the explanation of what makes real-time profiling wrong. To explain why, we must be more precise about what *psychological hijacking*, as we call it, involves.

As we understand it, psychological hijacking is a means by which one attempts to implement one's unilateral plan. Hence, for it to take place, the hijacker must at least:

(a) Have a unilateral plan *P*.
(b) Intend that *P* is successful.[10]

(c) Believe that an action or a series of related actions, ϕ, is a means to the success of P.
(d) Perform ϕ.

Conditions (a)–(d) suggest that psychological hijacking can take place even when the hijacker's plan is unsuccessful. However, even though psychological hijacking does not depend on the success of the hijacker's *plan*, it can only be said that an individual has psychologically hijacked another if the action (or series of actions) performed by the hijacker – that is, ϕ – succeeds in generating a particular effect. Namely, by ϕ-ing the hijacker must:

(e) Make some of the target's psychological states[11] *subservient* to the hijacker's intention that P succeeds.

Two qualifications about (e) are in order. The first concerns the hijacker's intention. As per condition (b), the hijacker must intend to see his unilateral plan through. But one may think that the hijacker must also intend the specific effects mentioned in (e). Such a requirement, we submit, would make our account unnecessarily under-inclusive. Most evidently, it would rule out the possibility of one being engaged in psychological hijacking (and in real-time profiling) without realising it. For example, it is possible (though perhaps unlikely) that MoodX's executive board were unaware of how the new algorithm worked and decided to implement it solely based on the incomplete information that its implementation would maximise sales. Insofar as we can reduce MoodX's decisions to the decisions of its executive board, we could say that MoodX did not intend that users' moods become subservient to their plan to increase sales. But it would still be correct to say that users were psychologically hijacked (and we could even imagine MoodX's executive board making a public apology highlighting the fact that they would not have implemented the algorithm had they known how it worked).[12]

The second qualification concerns "subservient". By saying that the hijacker makes the target's psychological states subservient to the hijacker's intentions, we mean that the hijacker's influence establishes a hierarchy between the hijacker and the target. The hierarchy in question can be construed as a hierarchy between the target's psychological states and the hijacker's intentions. The hijacker behaves as if the target's psychological states (including the target's plans, intentions, and preferences) are *less valuable* than his own intentions. No, or little, regard is given to the target's standing to demand that her own psychological states are not placed at the service of the hijacker. The target's psychological states are treated as mere means to the success of the hijacker's unilateral plan.

The notion of subservience is, therefore, key to understanding why psychological hijacking features in the explanation of real-time profiling wrongness. Of course, similar forms of subservience are sometimes justified. For example, a social media company that intends to prevent users from engaging in self-harmful behaviour could use information about users' current moods or stress levels to induce them to seek professional help. In a way, we could say that the company also created a hierarchical and instrumental relationship whereby users' intentions, desires, or preferences were treated as less important than the company's intentions. But, contrary to MoodX's case, here the company's influence seems morally acceptable. And it seems so because the company's plan took the *interests* of individuals into account. But notice that *even* in this case, the company would not get off the hook with ease. Their interference, even if in the name of users' interests, would still stand in need of justification.[13] And that is so precisely because there seems to be something (pro tanto) wrong with creating hierarchical and instrumental relationships.[14]

Note that the wrongness of subservience is closely tied to its need and to the availability of alternative options. When an individual has other reasonable means to achieve their intended plan, doing so while making someone subservient is particularly condemnable. For example, MoodX had alternative ways to profit and to advertise its products. Sure, perhaps the alternatives would have been less effective, but choosing effectiveness over treating their users in a non-hierarchical and instrumental way would itself be a form of being reckless about morality (Chen 2021, 72–73).

What is striking about psychological hijacking – in the context of the cases of real-time profiling that we have been considering – is that there are often alternative ways to get targets to adopt the profiler's plan that show some regard for the target's consent, standing, or interests. But despite there being alternative ways to influence targets, profilers still choose to engage in a form of influence that gives rise to subservience. That is one of the reasons why this form of online manipulation often strikes many as deeply problematic.

Now, because psychological hijacking makes some aspect of the target's psychology subservient to the hijacker's intentions, it might be thought that psychological hijacking is a form of *domination*. Whether the hijacker–target relationship amounts to a relation of domination is debatable. One reason against seeing it as such concerns the scope of the subjection which is constitutive of relations of domination. In typical relations of domination (e.g., the slave–slaveowner relation), the dominated's "normative reasons to do what the [dominator] proposes constitutively track considerations that are dependent on the power-facts" (Vrousalis 2019, 8). Contrarily, because the hijacker influences the target by meddling with pre-existing reasons or other psychological features, we cannot say that the target's reasons (motivation, disposition, etc.) that are subservient to the hijacker's plan *constitutively* track considerations which are dependent on power facts.

Another reason against seeing the hijacker–target relation as one of domination concerns the transactional and transient – as opposed to structural and persistent – nature of their interaction. At least in the set of cases of psychological hijacking that interests us (i.e., cases of real-time profiling), the hijacker's influence over the target's psychology is episodic and dependent on the hijacker's *actual* exercise of his power over the target. By contrast, paradigmatic cases of domination are cases where we find an institutionally stabilised and enduring power relation whereby the mere subjection to the dominator's power, and not its exercise, is what calls for justification. As Dorothea Gädeke puts it, "conceptualizing both opportunistic and robust capacities to interfere as forms of domination risks losing sight of what is distinctive of non-domination as opposed to non- interference" and "[risks] misconstruing domination as an anomaly perpetrated by individual wrongdoers instead of as a feature that pervades society" (Gädeke 2020, 199). This is, of course, not to say that psychological hijacking can never constitute a relation of domination in the online domain.[15] The point is that we need not see psychological hijacking as necessarily constituting relations of domination to explain what is wrong with it.

A further clarification concerns the equation of psychological hijacking with bypassing rational capacity. The two forms of influence should not be conflated. In fact, sometimes psychological hijacking occurs only if the hijacked properly exercises their rational capacities. Consider another example of real-time profiling:

> (Election) To promote chaos and polarisation during elections, a search engine changes its algorithm to condition the information that users are exposed to according to their real-time online behaviour. Robust evidence favouring one's preferred candidate is presented at a calculated time when the user is believed to be less emotional and more likely to take evidence-based decisions. Polarisation rises as a result.

In (Election), the search engine did not bypass individuals' rational capacities – at least not in the sense of suppressing individuals' rational deliberation or disengaging individuals' "system 2", to borrow Kahneman's terminology (Kahneman 2011).[16] Raising one's confidence in a proposition based on stronger evidence is simply what should be expected from individuals who properly exercise their rational capacities.[17] The search engine's act, therefore, did not bypass users' rational capacities. In fact, in this case the exercise of their rational capacity was necessary for the search engine to achieve its plan. Be this as it may, the search engine is still engaged in psychological hijacking and its action still seems wrong for the reasons we have discussed earlier.

The idea that psychological hijacking creates a hierarchical and instrumental relation where there was none and where there need not be one is an important part of the explanation of what makes real-time profile wrong – and it might even help us explain why unsuccessful or benign instances of real-time profiling still raise moral concerns.

With all that said, it would be mistaken to assume that psychological hijacking alone provides the full explanation for what makes real-time profiling wrong. Another element should be included. To find it, let us once again return to (MoodX). A further fact that we can infer from this case is that the implementation of the algorithm *transformed* user's moods into vulnerabilities. While vulnerabilities are often associated with intrinsic characteristics, such as age, gender, or disability status, this is not the case in real-time profiling.

In cases of real-time profiling, like in (MoodX), the transient and dynamic status of profiled individuals gives the profiler a unique opportunity to exercise influence that is specific to the context (the current circumstances that the profiled subject is undergoing) and the relation (the personal, commercial, or political relationship between the profiler and the profiled subject). We can say, therefore, that the profiler's increased knowledge and his influence in the online environment work together as an *enabling* condition: they remove an obstacle for individuals to be wronged in different ways. For example, by figuring out how to influence individuals on the basis of their moods and by making this influence possible, MoodX *is now able to* get individuals to do more than buying products. In principle MoodX could rely on its ability to influence users on the basis of their moods to exploit, abuse, harm, or discriminate them. Just as it is said that gateway drugs prime or prepare someone's organism for heavier substances by removing some natural inhibitors, we can say that real-time profiling is a *gateway wrong*: By transforming some psychological features into vulnerabilities, the profiler removes obstacles and creates opportunities for individuals to be wronged in different ways. We take it that removing obstacles and creating opportunities for individuals to be wronged in different ways without a strong justification for doing so is itself pro tanto wrong. After all, this amounts to subjecting individuals to unnecessary risks.

Now, one may object by submitting that something similar often happens in interpersonal or online interactions without giving rise to moral concerns. For example, by befriending someone we might remove some obstacles to wrong the person. We might, for example, make the person more vulnerable to emotional blackmail or to abuse of trust. So why do we not say that befriending someone is also a gateway wrong? The reason is that despite making each other more vulnerable to some wrongs, when we form genuine friendships, these concerns are mitigated by the fact that we treat each other as equals, non-instrumentally, and by the fact that making each other vulnerable is not constitutive of the relation but simply an inevitable by-product. Contrarily, in cases of real-time profiling, creating a vulnerability on the profiled subject is an integral part of the way in which the profiler chooses to exert its influence. That is why it is worth highlighting that real-time profiling – but not befriending – is a gateway wrong.

Notice, however, that the fact that real-time profiling is a gateway wrong should not be seen as a de facto harm-based explanation. Cases where the profiled subject fails to adopt the profiler's unilateral plan can help us clarify the point. Arguably, no harm occurred to an individual who resisted the temptation to make a bet at a sports event despite being induced to doing so by their phone's real-time profiling. But the attempted influence did work as a gateway wrong. It removed an obstacle and created an opportunity for the individual to use his money against his own interests and subjected him to the risk of having his information about being at a sports event used for other detrimental purposes.

The claim that the profiler's interference works as a gateway wrong because it enables further wrongs should also not be conflated with a claim about the wrong-making features associated with the enabling conditions of real-time profiling. The occurrence of individual instances of real-time profiling typically depends on the satisfaction of a series of background conditions. In (MoodX), for example, we presupposed that the company had the relevant information about users' moods. But, in all likelihood, the company would not have been able to influence users had it not possessed such information. The possession of information then, in this case, works as an *enabling condition* for real-time profiling. Despite enabling conditions varying from case to case, we might say by way of generalisation that they are conditions the satisfaction of which places the hijacker in a position of power; in a position where he can, intentionally or not, interfere with the target's psychological state so as to cause the target to favour his plan.

Once we draw attention to the enabling conditions of real-time profiling, it is not difficult to see that an agent may wrong others by merely satisfying them. For example, it could be argued that MoodX has wronged users even before interfering with the online environment and users' moods. The mere acquisition of information about users' moods was (arguably) pro tanto wrongful because it gained access to intimate details about the users without a sound justification. Though such wrongs raise serious concerns, they should not be conflated with the wrongs *of* real-time profiling (one of them being that real-time profiling itself enables further wrongs). As such, they are less important for our purposes.

As per our account of real-time profiling, the profiler not only makes the profiled subject subservient, in some specific sense, to the profiler's unilateral plan but also does so whilst simultaneously enabling further wrongs. Therefore, when wrong, real-time profiling is wrong both because it involves psychological hijacking and because it works as a gateway wrong. These two aspects represent the key normative characteristics of real-time profiling but not exclusively to it, since we can observe the same characteristics in interpersonal counterparts of real-time profiling (e.g., Phil's case).

Having identified the wrong-making features of real-time profiling (and analogous interpersonal manipulative practices), we now move on to discuss what such normative reflections mean for regulatory initiatives.

5 Regulatory Implications

First, we must acknowledge that just because something is morally problematic it does not necessitate regulatory interventions. Other considerations, such as the scale of the impact, the costs of regulation, and the possibility of correction by less invasive mechanisms, may affect the policy outcome.

As much as we believe that at least the most blatant forms of real-time profiling should be regulated, the appropriate scope and venue of regulation will depend on further research. Nevertheless, we see the need to explain the regulatory implications of our theoretical findings, especially because ongoing public debates are taking place around the world about regulating online manipulative practices. Assuming that online manipulation is something that calls for legal regulation and that policymakers have good reasons to proceed with legal interventions, this section will briefly explore how our reflection on real-time profiling may help highlight the flaws in the current regulatory frameworks and perhaps more importantly, point towards a more promising direction. We have chosen the European Union (EU) regime for our analysis, but there is no obvious reason why the implications discussed in this section do not apply to other jurisdictions. We focus on two of the most relevant areas in the EU legal order, consumer protection and data protection law, before moving onto further comments on the latest developments in the proposed regulation on digital services and artificial intelligence (AI).

5.1 Consumer Protection Law

It is probably not difficult to think of the relevance of consumer protection law in addressing at least some of the challenges arising from real-time profiling. When it comes to *commercial* targeting (but not *political* targeting), individuals are usually protected as consumers. The EU's Unfair Commercial Practices Directive (UCPD),[18] for example, prohibits misleading, aggressive, and otherwise unfair commercial practices (European Commission 2005, art 5(3), para 28 Annex I).

Our theoretical discussions about real-time profiling could raise (and partly answer) the question as to whether the current consumer protection legal framework can fully address manipulative marketing practices.

First, can typical cases of real-time profiling be deemed as misleading in legal terms? Article 6(1) UCPD defines a misleading commercial practice as one that "contains false information and is therefore untruthful or . . . deceives or is likely to deceive the average consumer". We have already clarified that real-time profiling does not necessarily involve false, mis- or disinformation as such, and it can be simply presenting truthful information at an opportunistic time. As regards deception, we pointed out that core cases of real-time profiling are more likely to fall within the scope of non-deceptive manipulation. Whether a legal – as opposed to philosophical – concept of deception can capture this phenomenon would be a separate question, but in the current absence of legislative guidance or clear case law on this matter, it would probably at best be a stretch to consider real-time profiling as deceptive without a strong conceptual support.

Second, in terms of aggressive practices, the UCPD has a particular emphasis on "harassment, coercion, including the use of physical force, or undue influence" (European Commission 2005, art 8). Our earlier example of MoodX involves no harassment or coercion (although neither of those two terms are defined in the UCPD) but can nevertheless be seen as a form of manipulation. When it comes to undue influence, while conceptually it is debatable whether MoodX's practices are undue, the law has a relatively narrow definition of undue influence, namely "exploiting a position of power in relation to the consumer so as to apply pressure . . . in a way which significantly limits the consumer's ability to make an informed decision" (European Commission 2005, art 2(j)). Though we see that real-time profilers "exploit a position of power", the definition does not fit many cases of real-time profiling because real-time profilers, as a rule, do not pressurise users.

Third, and due to the unsatisfactory coverage of the legal definitions of "misleading" and "aggressive" practices, the next question would be whether real-time profiling falls within the more generic concept of unfair practices. Under the UCPD, an unfair practice is one that meets two criteria: (a) it breaches professional diligence; and (b) it distorts the economic behaviour of the average consumer. Real-time profiling presents a particularly interesting case to condition (b), because on the one hand, it clearly shows potentially distortive effect on the economic behaviour of the consumers, which rests at the heart of the very idea of psychological hijacking in a commercial context. On the other hand, however, condition (b) has a particular emphasis on the average consumer – whether with regard to the entire market or a targeted group – not an individual consumer, which can be problematic in the case of real-time profiling. With its hyper-personalisation nature, it is unclear how the average consumer standard may apply to individualised manipulation. Indeed, current online marketing practices have evolved from targeting a group audience to "an audience of one" (Summers, Smith, and Reczek 2016). Laux et al. have highlighted some of the similar challenges in the context of online behavioural advertising and call for a stricter average consumer test (Laux, Wachter, and Mittelstadt 2021). For consumer protection law to fully capture real-time profiling and similarly manipulative practices, it would either entail further legislative, judicial, or regulatory guidance to expand the legal concept of "misleading", "aggressive" or "unfair", or a new provision specifically covering manipulative practices.

5.2 Data Protection Law

To the extent that typical real-time profiling techniques involve the collection of personal data, data protection law may stand out as a promising regulatory forum in restricting the use of personal data and hence real-time profiling practices. The earlier discussions on the technical and moral nature of these practices, however, reveal some conceptual challenges in applying data protection law to real-time profiling.

First, on a technical level, it has been pointed out how real-time profiling can be particularly intrusive by identifying the exact moment where the targeted individual would be susceptible to the influence. Yet, this does not necessary involve "sensitive data" as defined by Article 9 of the General Data Protection Regulation (GDPR).[19] Under Article 9, sensitive data is defined as

> data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. (European Commission 2016, art 9)

Real-time profiling does not necessarily involve any of such categories of data but can be equally revealing and exploitative. The challenge, as such, lies at the current data protection regime's inability to clearly capture manipulative practices that do not rely on data that is classically considered prone to discrimination or manipulation. It should be noted that other parts of the GDPR still apply to non-sensitive uses of personal data, but the level of protection would be significantly lower.

Second, on a moral level, and perhaps on a more optimistic note, our conceptualisation of real-time profiling elaborates why such practices are legally challengeable in the absence of clearly applicable rules or guidance. The disrupted power dynamics exhibited in the case of real-time profiling provides an articulation of how such practices can possibly be held unlawful. For example, as a matter of data protection principle, Article 5(1)(a) requires that personal data must be "processed lawfully, fairly and in a transparent manner", rendering any unfair uses of personal data illegal. While the law does not further clarify what amounts to "unfairness", our discussion on psychological hijacking and gateway wrongs presents a conceptual case against the acceptance of real-time profiling as a "fair use" of personal data. Another example is how the profiler's exploitation of its position of power, as fleshed out with the notion of *subjection*, would create a power imbalance. The legal consequence of the establishment of such an imbalance is that any consent given by the data subject would no longer be considered "freely given" (European Data Protection Board 2020, 7–9), rendering the uses of personal data reliant on such consent no longer lawful. Of course, specific provisions directly addressing real-time profiling would be the most effective way to regulate it, but before new rules are put in place, courts and regulators would have to rely on a theoretical explanation of the moral wrongness of real-time profiling.

5.3 The Digital Services Act and the Artificial Intelligence Act: An Opportunity?

Given the limitations of consumer protection and data protection law, arguably a more targeted regulatory approach is needed to effectively address the unique challenges of real-time profiling. There have been ongoing regulatory efforts initiated in the EU on manipulation. In December 2020, for example, the European Commission published the long-expected proposal for a Digital Services Act (DSA) (European Commission 2021). Two draft provisions might be of particular interest. The proposed Article 24 requires online platforms to disclose the factors that determine how adverts are targeted to internet users. The transparency requirement here may partly cover commercial real-time profiling through third-party platforms, but as discussed earlier, the fact that the manipulee is aware of the manipulator's intention does not fundamentally change the moral status of the action. The proposed Article 26(1), on the other hand, would impose a duty on key online platforms to monitor the spread of information with regard to public interest, which could cover manipulative political – but not necessarily commercial – real-time profiling.

More recently, in April 2021, the Commission tabled a proposal for the Artificial Intelligence Act (AIA). While not all real-time profiling techniques will involve what the AIA defines as AI, the relatively broad definition[20] would likely capture a large part of real-time profiling systems, especially the more sophisticated ones.

Article 5 of the draft AIA prohibits, among other things, two types of manipulative AI systems, one that "deploys subliminal techniques beyond a person's consciousness", the other that "exploits any of the vulnerabilities of a specific group of persons" (European Commission 2021, art 5). Both banned practices must however "materially distort a person's behaviour" and cause "physical or psychological harm" (European Commission 2021, art 5). As a preliminary assessment, it seems typical real-time profiling practices may count as a "subliminal technique" but would probably not involve vulnerabilities as currently limited to only "age, physical or mental disability". More importantly, while our analysis shows successful attempts of real-time profiling could create behavioural distortion, the "physical or psychological harm" bar is perhaps too high a legal test to cover the more subtle, yet wrongful, forms of real-time profiling.

The Commission is clearly mindful of the interplays between the AIA and other areas of law by stating "[o]ther manipulative or exploitative practices affecting adults that might be facilitated by AI systems could be covered by the existing data protection, consumer protection and digital service legislation" (European Commission 2021, 13), but our analysis has exposed some of the regulatory challenges in those areas. Building on our theoretical enquiry into the nature of real-time profiling, further legal research could – and should – be carried out to uncover how the regulatory regime could be better equipped to address novel forms of online manipulation.

6 Conclusion

Real-time profiling is already a part of our online environment. All suggests that it is here to stay. We have shown that some cases of real-time profiling closely resemble wrongful manipulative practices and, thus, raise similar ethical concerns. To highlight such concerns, we have provided an account of what makes real-time profiling wrong. Real-time profiling is wrong both because the profiler engages in what we have called "psychological hijacking" and because it is a gateway wrong. This diagnosis has led us to identify shortcomings that might help the potential regulation of real-time profiling and related online manipulative practices. Whether real-time profiling needs to be regulated and how to precisely go about it are questions that we cannot tackle in this chapter. But if, as some have envisaged, real-time profiling is the future of online manipulation, these questions cannot be ignored in further discussions.

Notes

1. *We thank Fleur Jongepier, Himani Bhakuni, Kalle Grill, Michael Klenk, Moti Gorin, and Pei-Hua Huang for their helpful written comments and suggestions on a previous draft. For discussion and feedback, we also thank the audiences of the Manipulation Online Workshop Series and the Maastricht Law and Tech Lab.
2. Among them, the increase in the use of smart devices, the increase in computing capacity which allowed for more sophisticated forms of data collection/analysis, the growth of digital services that monetise users' data, among others. For more examples, see Zuboff (2019).
3. We assume here that the category "manipulative but justified" is not an empty one. It is worth noting, however, that this is not uncontroversial, as some philosophers may adopt a thick conception of manipulation according to which manipulation would be wrongful by definition. Given that we focus on the wrongful instances of manipulation (and on real-time profiling), nothing in our argument would change if the thick conception of manipulation turns out to be correct. For an overview of thick and thin conceptions of manipulation, see Jongepier and Klenk, in this volume.
4. For example, Sunstein (2016, 82) defines manipulation as "an effort to influence people's choices . . . to the extent that it does not sufficiently engage or appeal to their capacity for reflection and deliberation." Along the same lines, others have highlighted the fact that manipulators influence behaviour by "adjusting [the manipulee's] psychological levers" (Noggle 1996, 44).
5. The scenario is fictional, but not fictitious. See Sam Levin (2017).
6. There are different ways in which a plan can be said to be unilateral: when (i) the *design* of one's plan is underpinned by an agenda set out without the manipulee's input, consent, or awareness; when (ii) the *implementation* of one's plan is not actually accepted or would not be accepted by the target in idealised conditions; or when (iii) the *primary* motivating reason for implementing the plan is its benefit to the planner. Though many instances of manipulative influences (and real-time profiling) are unilateral in all three senses, we circumscribe our use of "unilateral" to the third sense just specified. It is this sense that helps in explaining why manipulative influences in general, and real-time profiling in particular, seem morally suspicious even when it favours the manipulee's interests or well-being. We return to this in Section 3. Thanks to Kalle Grill for pressing us on this point.
7. Note here that we are not suggesting that MoodX's influence *bypasses* users' deliberative capacity. Some accounts of manipulation do require the manipulator to either bypass or disengage the manipulee's deliberative capacity. We explain why we think this is inadequate in Section 3.

8. The kind of cases that interest us are cases like (MoodX), (Uber), and other examples we cited in Section 1.
9. From this point onwards whenever we use "wrong" and related terms we mean "*pro tanto* wrong". For short, we also suppress the qualifier "when wrong".
10. Conditions (a) and (b) are stated separately to highlight the fact that one can *have* a plan (in the sense of having a layout of the steps that will lead to a certain end) but can still decide to commit to the plan or not.
11. We use "psychological states" broadly and include phenomena that might not be strictly or purely part of someone's psychology. For example, we would include moods, feelings, preferences, motives, reasons, dispositions, beliefs, and other propositional attitudes.
12. Our point here follows Marcia Baron's general claims about what she calls the "*Mens Rea* of manipulation" (Baron 2003, 2014).
13. It could, for example, be seen as wrongfully paternalistic. See Grill (2012).
14. On why purely instrumental (and hierarchical) relations such as the ones we have been considering are (at least) *pro tanto* wrongful, see Jongepier and Wieland, in this volume.
15. Also, we do not deny the possibility that large-scale and continued imposition of psychological hijacking may lead to the materialisation of domination in the long term. For accounts that associate forms of online manipulation with relations of domination, see Gorin and Capasso, both in this volume.
16. For a helpful discussion on whether manipulation necessarily involves bypassing, see Gorin (2014).
17. A detailed and recent argument along these lines can be found in Dorst (2020).
18. European Commission 2005.
19. European Commission 2016.
20. The draft AIA defines AI as, in short, software developed with machine learning, logic- and knowledge-based, and statistical approaches, Bayesian estimation, search and optimization methods. See art 3(1), Annex 1, ibid.

7 References

Alvarez, Sandra. 2017. "Mood Tracking & Emotional Advertising: What Does the Future Hold – NMPi." Accessed August 23, 2021. https://nmpidigital.com/us/mood-tracking-emotional-advertising-future-hold/.

Baron, Marcia. 2003. "Manipulativeness." *Proceedings and Addresses of the American Philosophical Association* 77 (2): 37. doi:10.2307/3219740.

Baron, Marcia. 2014. "The Mens Rea and Moral Status of Manipulation." In Coons and Weber 2014, 98–109.

Bradshaw, Samantha, and Philip N. Howard. 2018. "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation." https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/.

Brown, Bruce. 2018. "Fitbit Mental Health Physiology Monitoring App." Accessed August 23, 2021. https://healthtechinsider.com/2018/07/13/fitbit-mental-health-physiology-monitoring-app/.

Capasso, Marianna. 2022. "Manipulation as Digital Invasion: A Neo-republican Approach." In *The Philosophy of Online Manipulation*, edited by Jongepier, F. and Klenk, M., 180–198. New York, NY: Routledge.

Charara, Sophie. 2020. "A New Fitbit Claims to Track Your Stress Levels. Can it Really do it?" *WIRED UK*, August 28. Accessed August 23, 2021. www.wired.co.uk/article/fitbit-stress-tracking-eda.

Chen, Jiahong. 2021. *Regulating Online Behavioural Advertising through Data Protection Law*. Cheltenham: Edward Elgar Publishing.

Coons, Christian, and Michael Weber, eds. 2014. *Manipulation: Theory and Practice*. Oxford: Oxford University Press.

Dorst, Kevin. 2020. "Reasonably Polarized: Why Politics is More Rational Than You Think." Accessed August 23, 2021. www.kevindorst.com/stranger_apologies/rp.

European Commission. 2005. Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive') (Text with EEA relevance). UCPD. May 11. https://eur-lex.europa.eu/eli/dir/2005/29/oj.

European Commission. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 ('GDPR'). GDPR. April 27.

European Commission. 2021. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (ARTIFICIAL INTELLIGENCE ACT) and amending certain Union legislative acts. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.

European Data Protection Board. 2020. Guidelines 05/2020 on consent under Regulation 2016/679.

Gädeke, D. 2020. "Does a Mugger Dominate? Episodic Power and the Structural Dimension of Domination." *Journal of Political Philosophy* 28: 199–221.

Gorin, Moti. 2022. "Gamification, Manipulation, and Domination." In *The Philosophy of Online Manipulation*, edited by Jongepier, F. and Klenk, M., 199–215. New York, NY: Routledge.

Gorin, Moti. 2014. "Towards a Theory of Interpersonal Manipulation." In Coons and Weber 2014, 73–97.

Grill, Kalle. 2012. "Paternalism." In *Encyclopedia of Applied Ethics*, 2nd ed, edited by R. Chadwick, 359–69. London: Elsevier.

Jongepier, Fleur, and Michael Klenk. 2022a. "Online Manipulation: Charting the Field." In *The Philosophy of Online Manipulation*, edited by Jongepier, F. and Klenk, M., 15–48. New York, NY: Routledge.

Jongepier, Fleur, and Michael Klenk, eds. 2022b. *The Philosophy of Online Manipulation*. New York, NY: Routledge.

Jongepier, Fleur, and J. W. Wieland. 2022. "Microtargeting People as a Mere Means." In *The Philosophy of Online Manipulation*, edited by Jongepier, F. and Klenk, M., 156–179. New York, NY: Routledge.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York, NY: Farrar Straus Giroux.

Laux, Johann, Sandra Wachter, and Brend Mittelstadt. 2021. "Neutralizing Online Behavioural Advertising: Algorithmic Targeting with Market Power as an Unfair Commercial Practice." *Common Market Law Reviews* 58 (3): 719.

Levin, Sam. 2017. "Facebook Told Advertisers it Can Identify Teens Feeling 'Insecure' and 'Worthless'." *The Guardian*, January 5. Accessed August 23, 2021. www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens.

Lindsay, Jessica. 2019. "Does Uber Charge More if Your Battery is Lower?" *Metro.co.uk*, September 27. Accessed August 23, 2021. https://metro.co.uk/2019/09/27/uber-charge-battery-lower-10778303/.

Mahwadi, Arwa. 2018. "Uber Developing Technology that Would Tell if You're Drunk." *The Guardian*, November 6. Accessed August 23, 2021. www.theguardian.com/technology/2018/jun/11/uber-drunk-technology-new-ai-feature-patent.

Metz, Rachel. 2016. "Control Your Smartphone with Your Eyes." *MIT Technology Review*, January 7. Accessed August 23, 2021. www.technologyreview.com/2016/07/01/159012/control-your-smartphone-with-your-eyes.

Noggle, Robert. 1996. "Manipulative Actions: A Conceptual and Moral Analysis." *American Philosophical Quarterly* 33 (1): 43–55.

Rudgard, Olivia. 2018. "Gambling Firms Could Use GPS to Tempt 'Vulnerable' Customers." *The Telegraph*. Accessed August 23, 2021. www.telegraph.co.uk/news/2018/06/25/gambling-firms-could-use-gps-tempt-vulnerable-customers/.

Sax, Marijn. 2021. "Optimization of What? For-profit Health Apps as Manipulative Digital Environments." *Ethics and Information Technology* 23 (3): 345–61.

Shapiro, Tom. 2016. "How Emotion-Detection Technology Will Change Marketing." October 17. Accessed August 23, 2021. https://blog.hubspot.com/marketing/emotion-detection-technology-marketing.

Spencer, Shaun B. 2020. "The Problem of Online Manipulation." *University of Illinois Law Review* 2020 (3): 959–1006. doi:10.2139/ssrn.3341653.

Summers, Christopher A., Robert W. Smith, and Rebecca W. Reczek. 2016. "An Audience of One: Behaviorally Targeted Ads as Implied Social Labels." *Journal of Consumer Research* 43 (1): 156.

Sunstein, Cass R. 2016. *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge: Cambridge University Press.

Susser, Daniel, Beate Roessler, and Helen Nissenbaum. 2019. "Online Manipulation: Hidden Influences in A Digital World." *Georgetown Law Technology Review* 4 (1): 1–45. Accessed February 27, 2020.

Valliappan, Nachiappan, Na Dai, Ethan Steinberg, Junfeng He, Kantwon Rogers, Venky Ramachandran, Pingmei Xu et al. 2020. "Accelerating Eye Movement Research via Accurate and Affordable Smartphone Eye Tracking." *Nature Communications* 11 (1): 4553. doi:10.1038/s41467-020-18360-5.

Vrousalis, Nicholas. 2019. "How Exploiters Dominate." *Review of Social Economy* 1.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, NY: PublicAffairs.