In: Emre Bayamlıoğlu, Irina Baraluic, Liisa Janssens and Mireille Hildebrandt (eds), BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen, 2018: Amsterdam University Press, 16-19. DOI 10.5117/9789463722124/CH1.

From agency-enhancement intentions to profile-based optimisation tools: what is lost in translation

*Sylvie Delacroix**

Abstract

Whether it be by increasing the accuracy of web searches, educational interventions or policing, the level of personalisation that is made possible by increasingly sophisticated profiles promises to make our lives better. Why 'wander in the dark', making choices as important as that of our lifetime partner, based on the limited amount of information we humans may plausibly gather? The data collection technologies empowered by wearables and apps mean that machines can now 'read' many aspects of our quotidian lives. Combined with data mining techniques, these expanding datasets facilitate the discovery of statistically robust correlations between particular human traits and behaviors, which in turn allow for increasingly accurate profile-based optimisation tools. Most of these tools proceed from a silent assumption: our imperfect grasp of limited data is at the root of most of what goes wrong in the decisions we make. Today, this grasp of data can be perfected in ways unimaginable even twenty years ago. The profile-based optimisation tools built thanks to this data 'boon' thus promise to lift us out of our murky meanderings: if precise algorithmic recommendations can replace the flawed heuristics that preside over most of our decisions, why think twice? The above line of argument often informs the widely shared assumption that today's profile-based technologies are agency-enhancing, supposedly facilitating a fuller, richer realisation of the selves we aspire to be.

This 'provocation' not only questions this assumption; it also highlights the extent to which the agency-compromising, unintended effects of such technologies threaten the very resources that could be relied on to mitigate these agency-compromising aspects. Could it be the case that a hereto poorly acknowledged side effect of our profile-based systems (and the algorithmic forms of government they empower) is that it leaves us sheep-like, unable to mobilise a normative muscle that has gone limp? The more we are content to offload normative decisions to profile-based, 'optimised' algorithms, the more atrophied our 'normative muscles' would become. Considered at scale, the (endless) normative holidays that would result from such 'offloading' would spell the end of agency (in spite of the noble, agency-enhancing intentions that drove their development).

Keywords: ethical agency, fallibility, profile-based optimisation, reflectivity, enhancement

Introduction

Whether it be by increasing the accuracy of web searches, educational interventions or policing, the level of personalisation that is made possible by increasingly sophisticated profiles promises to make our lives better. Why 'wander in the dark', making choices as important as that of our lifetime partner, based on the limited amount of information we humans may plausibly gather? The data collection technologies empowered by wearables and apps mean that machines can now 'read' many aspects of our quotidian lives. Combined with fast evolving data mining techniques, these expanding datasets facilitate the discovery of statistically robust correlations between particular human traits and behaviours, which in turn allow for increasingly accurate profile-based optimisation tools. Most of these tools proceed from a silent assumption: our imperfect grasp of data is at the root of most of what goes wrong in the decisions we make. Today, this grasp of data can be perfected in ways not necessarily foreseeable even 10 years ago, when *Profiling the European Citizen* defined most of the issues discussed in this volume. If data-perfected, precise algorithmic recommendations can replace the flawed heuristics that preside over most of our decisions, why think twice? This line of argument informs the widely-shared assumption that today's profile-based technologies are agency-enhancing, supposedly facilitating a fuller, richer realisation of the selves we aspire to be. This 'provocation' questions this assumption.

Fallibility's inherent value

Neither humans nor machines are infallible. Yet our unprecedented ability to collect and process vast amounts of data is transforming our relationship to both fallibility and certainty. This manifests itself not just in terms of the epistemic confidence sometimes wrongly generated by such methods. This changed relationship also translates in an important shift in attitude, both in the extent to which we strive for control and 'objective' certainty and in the extent to which we retain a critical, questioning stance.

The data boon described above has reinforced an appetite for 'objective' certainty that is far from new. Indeed, one may read a large chunk of the history of philosophy as expressing our longing to overcome the limitations inherent in the fact that our perception of reality is necessarily imperfect, constrained by the imprecision of our senses (de Montaigne 1993). The rationalist tradition which the above longing has given rise to is balanced by an equally significant branch of philosophy, which emphasizes the futility of our trying to jump over our own shoulders, striving to build knowledge and certainty on the basis of an overly restrictive understanding of objectivity, according to which a claim is objectively true only if it accurately 'tracks' some object (Putnam 2004) that is maximally detached from our own perspective. Such aspiring for a Cartesian form of objectivity (Fink 2006) is futile, on this account, because by necessity the only reality we have access to is always already inhabited by us, suffused with our aspirations.

To denigrate this biased, 'subjective' perspective as 'irrational' risks depriving us of an array of insights. Some of these simply stem from an ability for wonder, capturing the rich diversity of human experience, in all its frailty and imperfection. Others are best described as 'skilled intuitions' (Kahneman and Klein 2009) gained through extensive experience in an environment that provides opportunity for constructive feedback, the insights provided by such skilled intuitions are likely to be dismissed when building systems bent on optimizing evidence-based outcomes. Instead of considering the role played by an array of non-cognitive factors in decisions 'gone wrong', the focus will be on identifying what machine-readable data has been misinterpreted or ignored. If factors such as habits and intuitions are known to play a role, they are merely seen as malleable targets that can be manipulated through adequate environment architecture, rather than as valuable sources of insights that may call into question an 'irrationality verdict'.

Similarly, the possibility of measuring the likely impact of different types of social intervention by reference to sophisticated group profiles is all too often seen as dispensing policy-makers from the need to take into account considerations that are not machine-readable (such as the importance of a landscape). Indeed the latter considerations may not have anything to do with 'data' per se, stemming instead from age-old ethical questions related to the kind of persons we aspire to be. Some believe those ethical questions lend themselves to 'objectively certain' answers just as well as the practical problems tackled through predictive profiling. On this view, perduring ethical disagreements only reflect our cognitive limitations, which could in principle be overcome, were we to design an all-knowing, benevolent superintelligence. From that perspective, the prospect of being able to rely on a system's superior cognitive prowess to answer the 'how should we [I] live' question with certainty, once and for all, is a boon that ought to be met with enthusiasm. From an 'ethics as a work in progress' by contrast, such a prospect can only be met with scepticism at best or alarm at worst (Delacroix 2019b): on this view, the advent of AI-enabled moral perfectionism would not only threaten our democratic practices, but also the very possibility of civic responsibility.

Civic responsibility and our readiness to question existing practices

Ethical agency has always been tied to the fact that human judgment is imperfect: we keep getting things wrong, both when it comes to the way the world is and when it comes to the way it ought to be. The extent to which we are prepared to acknowledge the latter, moral fallibility—and our proposed strategies to address it—have significant, concrete consequences. The latter can be felt at a personal and at an institutional, political level. A commitment to acknowledging our moral fallibility is indeed widely deemed to be a key organising principle underlying the discursive practices at the heart of our liberal democracies (Habermas 1999). This section considers the extent to which the data-fed striving for 'objective certainty' is all too likely to compromise the above commitment.

Now you may ask: why is such a questioning stance important? Why muddle the waters if significant, 'data-enabled' advances in the way we understand ourselves (and our relationship to our environment) mean that some fragile state of socio-political equilibrium has been reached? First, one has to emphasise that it is unlikely that any of the answers given below will move those whose metaphysical or ideological beliefs already lead them to deem the worldview informing such equilibrium to be 'true', rather than 'reasonable' (Habermas 1995). The below is of value only to those who are impressed enough by newly generated, data-backed knowledge to be tempted to upgrade their beliefs from 'reasonable' to 'true'. A poor understanding of the limitations inherent in both the delineation of the data that feeds predictive models and the models themselves is indeed contributing to a shift in what Jasanoff aptly described as the culturally informed 'practices of objectivity'. In her astute analysis of the extent to which the ideal of policy objectivity is differently articulated in disparate political cultures, Jasanoff highlights the United States' marked preference for quantitative analysis (Jasanoff 2011). Today the recognition of the potential inherent in a variety of data mining techniques within the public sector (Veale, Van Kleek, and Binns 2018) is spreading this appetite for quantification well beyond the United States.

So why does the above matter at all? While a commitment to acknowledging the fallibility of our practices is widely deemed a cornerstone of liberal democracies, the psychological obstacles to such acknowledgment -including the role of habit- are too rarely considered. All of the most influential theorists of democratic legitimacy take the continued possibility of critical reflective agency as a presupposition that is key to their articulation of the relationship between autonomy and authority. To take but one example: in Raz's account, political authority is legitimate to the extent that it successfully enables us to comply with the demands of 'right reason'(Raz 1986). This legitimacy cannot be established once and for all: respect for autonomy entails that we keep checking that a given authority still has a positive 'normal justification score' (Raz 1990). If the 'reflective individual' finds that abiding by that authority's precepts takes her away from the path of 'right reason', she has a duty to challenge those precepts, thereby renewing the fabric from which those normative precepts arise. In the case of a legal system, that fabric will be pervaded by both instrumental concerns and moral aspirations. These other, pre-existing norms provide the material from which the 'reflective individual' is meant to draw the resources necessary to assessing an authority's legitimacy. Much work has gone into analysing the interdependence between those different forms of normativity; not nearly enough consideration has been given to the factors that may warrant tempering political and legal theory's naive optimism—including that of Delacroix (2006)—when it comes to our enduring capacity for reflective agency.

Conclusion

To live up to the ideal of reflectivity that is presupposed by most theories of liberal democracy entails an ability to step back from the habitual and question widely accepted practices (Delacroix 2019a). Challenging as it is to maintain such critical distance in an 'offline world', it becomes particularly arduous when surrounded by some habit-reinforcing, optimised environment at the service of 'algorithmic government'. The statistical knowledge relied on by such form of government does not lend itself to contestation through argumentative practices, hence the temptation to conclude that such algorithmic government can only be assessed by reference to its 'operational contribution to our socio-economic life' (Rouvroy 2016). That contribution will, in many cases, consist in streamlining even the most personal choices and decisions thanks to a 'networked environment that monitors its users and adapts its services in real time' (Hildebrandt 2008). Could it be the case that a hereto poorly acknowledged side effect of our profile-based systems (and the algorithmic forms of government they empower) consists in its leaving us sheep-like, unable to mobilise a normative muscle that has gone limp through lack of exercise? The more efficient those systems are, the more we are content to offload normative decisions to their 'optimised' algorithms, the more atrophied our 'normative muscles' would become. Considered at scale, the (endless) normative holidays that would result from such 'offloading' would spell the end of agency, and hence the end of legal normativity. All that in spite of the noble, agency-enhancing intentions that prompted the creation of such systems in the first place.

* Sylvie Delacroix is a Professor in Law and Ethics at the University of Birmingham and a fellow of the Alan Turing Institute.

References

Bostrom, Nick. 2014. Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

de Montaigne, Michel. 1993. The complete essays. Translated and edited with an Introduction and Notes by M. A. Screech. London: Penguin Books.

Delacroix, Sylvie. 2006. Legal norms and normativity: an essay in genealogy. Oxford: Hart Publishing.

Delacroix, Sylvie. 2019a. Habitual Ethics? Oxford: Hart Publishing.

Delacroix, Sylvie. 2019b. "Taking Turing by surprise? Designing autonomous systems for morally-loaded contexts." arXiv:1803.04548.

Fink, Hans. 2006. "Three Sorts of Naturalism." European Journal of Philosophy 14(2): 202-21.

Habermas, Jurgen. 1995. "Reconciliation Through the Public use of Reason: Remarks on John Rawls's Political Liberalism." The Journal of Philosophy 92(3): 109-31.

Hildebrandt, Mireille. 2008. "Defining Profiling: A New Type of Knowledge?" In Profiling the European Citizen: Cross-Disciplinary Perspectives, edited by Mireille Hildebrandt and Serge Gutwirth, 17-45. Dordrecht: Springer.

Jasanoff, Sheila. 2011. “The Practices of Objectivity in Regulatory Science” In Social Knowledge in the Making, edited by Charle Camic, Neil Gross, and Michèle Lamont, 307-37. Chicago: University of Chicago Press.

Kahneman, Daniel, and Gary Klein. 2009. "Conditions for intuitive expertise: a failure to disagree." Am Psychol 64(6): 515-26. doi: 10.1037/a0016755.

Putnam, Hilary. 2004. Ethics without ontology. Cambridge, MA: Harvard University Press.

Raz, Joseph. 1986. The morality of freedom. Oxford: Clarendon Press.

Raz, Joseph. 1990. Practical Reason and Norms. Princeton, NJ: Princeton University Press.

Rouvroy, Antoinette. 2016. "La gouvernementalité algorithmique: radicalisation et stratégie immunitaire du capitalisme et du néolibéralisme?” La Deleuziana, (3): 30-36.

Veale, Michael, Max Van Kleek, and Reuben Binns. 2018. "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. doi: 0.1145/3173574.317401.