The New Imbroglio:[1] Living with Machine Algorithms

Mireille Hildebrandt

> Every day a piece of computer code is sent to me by e-mail from a website to which I subscribe called IFTTT. Those letters stand for the phrase "if this then that," and the code is in the form of a "recipe" that has the power to animate it. Recently, for instance, I chose to enable an IFTTT recipe that read, "if the temperature in my house falls below 45 degrees Fahrenheit, then send me a text message." It's a simple command that heralds a significant change in how we will be living our lives when much of the material world is connected—like my thermostat—to the Internet.
>
> *Sue Halpern, 2014*[2]

> Since the *present futures* co-determine *the future present*, predictions basically enlarge the probability space we face; they do not reduce but expand both uncertainty and possibility. The question is about the distribution of the uncertainty and the possibility: who gets how much of what?
>
> *Mireille Hildebrandt, 2016*

Two Types of Algorithmic Governance

IFTTT stands for 'if this than that' (Halpern 2014). IFTTT is how computers 'think'. It suggests that computers can only run like closed systems that are deterministic by definition. Nevertheless, due to their processing power, connectivity, and our inventiveness, computing systems have now reached unprecedented levels of complexity, generating previously unforeseen levels of indeterminacy. It turns out that a recursive series of deterministic instructions (IFTTT) is capable of producing emergent behaviours that surprise even those who wrote the initial recipes (an algorithm, ultimately, is nothing more or less than a rather precise recipe or set of instructions). This is the result of advances in a sub discipline of artificial intelligence, called 'machine learning' (ML). We should, however, not mistake deterministic computing systems that follow clear and simple rules to provide *automation* of well-defined tasks (IFTTT), for systems that reconfigure their own behaviours to *autonomically* improve their performance (ML). There is flexibility, a recursiveness and an unpredictability in ML that is absent in 'dumb' IFTTTs. The term 'dumb' here is not meant in a pejorative sense; it merely refers to non-learning algorithms that do not adapt their own IFTTTs on the basis of their 'experience', though they may adapt their behaviour, based on their IFTTTs. A simple thermostat runs on a 'dumb' algorithm: its IFTTT determines that whenever the temperature drops below (or above) a certain degree the heating or air conditioner will be turned on – until that same temperature is reached. A smart energy grid will require a continuous learning process, to calibrate energy demand and (local and centralised) energy supply in a way that enables load balancing as well as energy efficiency. As I hope to clarify, 'dumb' can be smart as far as dependence on computing systems is concerned.

In this essay I will suggest that 'dumb' IFTTTs and ML can each have added value as well as drawbacks, depending on how they are used for what purpose. On the one hand, *automated decision systems* in public administration may, for instance, have the added value of being predictable while sustaining accountability, precisely because they are 'dumb' (they do not learn, they just do as instructed). The drawback will probably be that automated decisions are rigid and may be flawed because only a limited set of data points is taken into account. On the other hand, *autonomic decision support systems* for medical diagnosis may, for instance, have the added value of coming up with

[1] An imbroglio has been defined as 'a confused mass; an intricate or complicated situation (as in a drama or novel); an acutely painful or embarrassing misunderstanding', cf. 'Imbroglio' Merriam-Webster.com, *Merriam-Webster*, accessed 2 August 2016. In this essay the concept is used as a reference to the complex entanglement of deterministic and learning algorithms that (in)form our increasingly data-driven environment.

[2] See also https://ifttt.com.

unforeseen correlations between previously unrelated data points, precisely because such systems improve their performance due to recurrent feedback cycles. Here the drawback may be that the system is not transparent (its inner operations are black boxed) and its output cannot be explained, other than referring to often-irretrievable statistics. I believe that a discussion that turns on whether one is for or against algorithmic governance in general would ignore the difference between two types of algorithms and thereby obfuscate what is at stake. Instead, the discussion about algorithms should focus on the type of problems that benefit from either strict application of 'dumb' algorithms or from adaptive algorithms that are not entirely predictable.

This being said, the question of when to use what algorithms is neither a purely technical question (if there is such a thing) nor a purely political one (if there were such a thing).[3] The decision to engage either 'dumb' IFTTTs or ML may have far reaching consequences for those subjected to the outcome of algorithmic machines, which turns it into a political question. Even if the outcome is checked or applied by a human person, the impact of algorithmic governance is momentous, because that person may not be capable of explaining the outcome and she may have no competence to amend it. If democracy is about self-government we need to find ways and means to involve those affected by algorithmic governance in the choices that must be made.[4] Democratically informed decisions on what algorithms to use will therefore require public understanding of how their employment is constraint by what is possible and feasible in computational terms.[5] This may sound like an insurmountable challenge, but it is not unlike the challenge of alphabetising an entire population, which ultimately enabled self-government since the era of the printing press. On top of that, the political question depends on coming to terms with the *distribution* of beneficial and adverse effects amongst citizens, commercial and governmental players. It may be that a small set of players benefits, whereas others pay the costs. To chart such effects, we need to understand the difference between automation ('dumb' IFTTTs) and autonomics (ML). As to the latter, I will suggest that the distribution of benefits and costs is contingent on the *distribution of the uncertainty* that is created by predictive and pre-emptive analytics. Though one may intuitively assume that predictions reduce uncertainty, this is agent-dependent. Those with access to the predictions have additional reasons to change course, whereas those who remain in the dark are confronted with decisions they could not have foreseen.

Finally, this brief essay will draw the line between, on the one hand, legal certainty and, on the other hand, both arbitrary decision-making and rigid application of inflexible rules. The Rule of Law aims to create an institutional environment that enables us to foresee the legal effect of what we do, while further instituting our agency by stipulating that such effect is contestable in a court of law – also against big players.[6] Such a – procedural – conception of the Rule of Law implies that both automation and autonomics should be constraint in ways that open them up to scrutiny and render their computational judgements liable to being nullified as a result of legal proceedings. This will not solve all of the problems created by algorithmic governance. It should, nevertheless, create the level playing field needed to partake in the construction of the choice architectures that determine both individual freedom and 'the making of' the public good.[7]

Automation and Autonomics in an Onli*f*e World

[3] On the relationship between technology, morality, political issues and law: Chapter 7 and 8 in Hildebrandt 2015a.

[4] Cf. Hildebrandt and Gutwirth 2007, where we discriminate between aggregate, deliberative and participatory democratic practices. Instead of discussing these practices in terms of either/or we propose that each has an important role to play. On participatory democratic theory see notably Dewey (1927) and the excellent analysis of Marres (2005). Democratic participation does not assume that consensus can be reached, but vouches that those who suffer the consequences of a policy must be involved, cf. Mouffe (2000), who speaks of agonistic debate as a precondition for sound democratic decision making.

[5] Cf. Wynne 1995.

[6] I agree with Waldron (2008) that the core of the Rule of Law depends on an effective right to see to it that justice is done, by appealing to an independent court that has authority to decide the applicable interpretation of the law.

[7] The concept of a 'choice architecture' refers to the type of choices one can and cannot make and the default settings that favour specific options, taking note that most choices are made implicitly, based on heuristics rather than rational deliberation. See Thaler and Sunstein 2008, and – more interesting - Gigerenzer 2000.

The idea of an 'onli*f*e' world was initiated by Floridi and taken up by a group of philosophers, social scientists, neuroscientists, and computer scientists who wrote the Onli*f*e Manifesto, on 'being human in a hyperconnected era'.[8] The original idea was to signal the conflation of online and offline, as this distinction is becoming increasingly artificial. In my book on *Smart Technologies and the End(s) of Law,*[9] I have further developed the notion of an onli*f*e world, suggesting that autonomic computing systems develop a specific kind of mindless agency that animates our new 'social' environment. The onli*f*e world is not merely a matter of turning everything online but also a matter of *things* seemingly coming *alive*. This relates to the difference between automation and autonomics.[10]

Since the advent of the steam engine and electricity we are familiar with the automation of menial tasks, delegating physical labour to machines, either because of their enhanced 'horse power' or because of their ability to endlessly and rapidly repeat specific tasks.[11] Currently, however, a new type of automation has developed, automating cognitive tasks that require autonomic behaviour.[12] Mere repetition will not do here, as autonomic systems must be capable of reconfiguring their own rules when responding to changes in the environment.[13] Though both automation and autonomics operate on the basis of algorithms, the first are static whereas the second are adaptive, dynamic, and more or less transformative. Though both can be black boxed by whoever employs them,[14] algorithms that generate machine learning are inherently opaque – even for those who develop and employ them. This is evident when the system runs on deep learning multi-level artificial neural networks that conduct unsupervised learning, but even in other cases there is no easy way to explain why the algorithms decided as they did. Machine learning is an inductive process, its output is liable to inductive bias (bias in the data as well as bias in the algorithms) and its usage is liable to various types of inductive fallacies (e.g. mistaking the outcome for the truth, or deriving an 'ought' from an 'is'). This entails that, as with inductive science, we need certain scepticism as to the extrapolation of patterns detected in the observed data (called the training set in ML) to similar patterns in new data (called the test set in ML). Not only because these new data may contain a black swan, but also because we could probably have used other patterns to describe the initial data points and these other patterns may turn out to better fit with the mechanisms that determine both the training and the test set.[15]

In other words, the temporality of our being rules out that either human or machine learning is infallible. ML does not reduce uncertainty but extends it by adding new predictions that will trigger new responses that in turn call for updated predictions.[16] Moreover, a prediction is a 'present future' that influences the 'future present' because actors will change their behaviour based on such a prediction.[17] ML, based on predictive analytics, will therefore create new uncertainties, since we do not know how actors will respond to the new range of 'present futures'.[18] To the extent that the capability to anticipate uncertainty is a crucial characteristic of living organisms, ML can indeed be said to turn our machine environment onli*f*e. We may even come to a point where 'dumb' algorithms will come to the rescue, consolidating and stabilising cascading uncertainties by simply acting as stipulated, behaving as coded, contributing to a reasonable level of predictability.

It is not that simple, of course. If the onli*f*e world is an imbroglio of 'dumb' as well as smart algorithmic governance, the question will be when to endorse either one and how to foresee their interoperability (or, the lack thereof). As to the frustrations generated by automation let me provide a topical example. Imagine entering the Leiden Railway Station during maintenance work on the tourniquets, so it is not possible to check-in with your public transport chip card. Those at work

---

[8] Floridi 2014. I was part of the Onli*f*e Initiative, see <https://ec.europa.eu/digital-single-market/en/onlife-original-outcome>.
[9] Hildebrandt 2015a.
[10] Hildebrandt 2011.
[11] See Latour (2000) on delegation to technologies.
[12] Chess, Palmer, and White 2003, Hildebrandt and Rouvroy 2011.
[13] Steels 1995.
[14] Pasquale 2015.
[15] Mitchell 2006, Wolpert 2016.
[16] Gabor 1963.
[17] Esposito 2011.
[18] Hildebrandt 2016.

suggest you can just go through without checking in. At the next station, when checking out, you are automatically fined for traveling without having checked-in. When calling the help-desk the lady ensures you that this is no problem because you will get your money back. However, she ends the conversation with a warning: you can only get your money back three times per year – after that you will have to pay. Trying to explain to her that this was not your mistake does not ring any bells with her; she is just repeating the rules.[19] Note that 'dumb' IFTTTs cannot adjust their own rules based on feedback, which also means that any wrong input will cascade through the system (errors are also automated). Smart systems may be more flexible and improve their performance based on recurrent feedback loops. The question remains, however, who determines the performance metric in the light of what goals. Moreover, as indicated above, the opacity of ML systems may reduce both the accountability of their 'owners' and the contestability of their decisions.[20]

The Political, the Technical, and the Legal

The question of when to employ what type of algorithms is both a political and a technical question. At some point it should also be a legal question, because under the Rule of Law individuals should have effective means to challenge decisions that violate their rights and freedoms. Lawyers call this the right to effective remedies to uphold one's fundamental rights (e.g. codified in art. 13 European Convention of Human Rights). Algorithmic governance easily implies that one is not aware of how decisions have been prepared, moulded or even executed in the intestines of various computational systems. Autonomic computing systems, however, enable the profiling, categorising and targeting as citizens or consumers in terms of high or low risk for health, credit, profitable employment, failure to pass a grade in one's educational institution, for tax and social security fraud, for criminal or terrorist inclinations, and in terms of access to buildings, social security, countries or medical assistance. Such personalised targeting will determine what cognitive psychologists and behavioural economists call 'the choice architecture' that decides which options individuals have, and whether and how these options are brought to the attention of the 'user'.[21] It enables subliminal influencing of individual people, based on techniques like AB research-designs that trace and track how we respond to different interfaces, approaches, and options.[22] To the extent that 'dumb' algorithms rely on the input generated by ML the problems they generate are expounded. This results in an imbroglio of invisibly biased decision systems that mediate our access to the world (search engines, online social networks, smart energy grids and the more), potentially creating unprecedented uncertainty about how our machine-led environment will interpret and sanction our behaviours.

Such gloomy prophecies need not, however, come true. We have struggled against the arbitrary rule of dictators as well as the power of private actors capable of twisting our hand. We have developed ways and means to protect human dignity and individual liberty, achieving the kind of legal certainty that safeguards both the predictability and trustworthiness of our social and institutional environment and its open texture in the face of legitimate argumentation.[23] The point is that we cannot take for granted that remedies that worked in the era of printing press, steam engine, and electricity will necessarily protect us in an onli*f*e world. This will require rethinking as well as reinventing the Rule of Law, for instance by making the intestines of the emerging imbroglio transparent and by making its decisions contestable. In recent articles the so-called right to profile transparency of the EU General Data Protection Regulation has been heralded for its spot-on approach to ML algorithms.[24] This right means that automated decisions (whether based on dumb or smart algorithms) that significantly affect people, trigger the fundamental right to data protection. More specifically, such decisions 'automatically'

[19] The example modulates a similar experience of colleagues Aernout Schmidt and Gerrit-Jan Zwenne, as recounted during the Annual Meeting of the Netherlands Lawyers Association, 10th June 2016.
[20] For attempts to chart the legal issues of automated and autonomic bureaucratic decision making, see Citron 2007 and Citron and Pasquale 2014.
[21] Thaler, Sunstein and Balz 2010.
[22] This also regards attempts to influence voting, e.g. Christian and Griffiths 2016, explaining AB research-design and its usage in the US presidential elections. On the capacity to influence voting by merely tweaking a search algorithm see Epstein and Robertson 2015.
[23] Waldron 2011. See also, arguing for legality and against legalism Hildebrandt 2015b.
[24] E.g. Goodman and Flaxman 2016, cf. Hildebrandt 2012.

generate two obligations and one right: first, people must be told about the existence of automated decisions, second, they must be given meaningful access to the logic of such decision, and, third, those concerned have a right to object against being subject to such decisions. It is interesting to note that the right to profile transparency is framed – by some - as a clash between US based AI companies and the EU, or even between innovation and Luddite hesitation.[25] I believe that such labels are out-dated and stand in the way of global progress. Progress involves hesitation as well as innovation, high risk and high gain, but not at the cost of those already disadvantaged, or vulnerable to data-driven exclusion. At some point any individual person faced with the onli*f*e imbroglio - that we are already a part of – may be disadvantaged by and vulnerable to unfair and degrading treatment by interacting automated and autonomic computing systems.

Profile transparency implies that the uncertainty generated by ML should be contained, to prevent mishaps. It also implies that decisions that seriously affect individuals' capabilities must be constructed in ways that are comprehensible as well as contestable.[26] If that is not possible, or, *as long as* this is not possible, such decisions are unlawful. In that case we may have to employ dumb algorithms, though even the outcome of dumb algorithms must be comprehensible and contestable. As to ML, we need to invest in the engineering of choice architectures that re-instates our agency instead of manipulating it. This is not about 'the more choice the better'.[27] It can only be about involving those whose onli*f*e is at stake in the construction of the choice architectures that will define their capabilities – and thus, their effective rights.

[25] E.g. Metz 2016.
[26] My use of the term capability in inspired by Sen (2004), where the capability is the substance that is to be protected, while the right itself co-constitutes the capability by safeguarding its sustainability.
[27] Van den Berg 2016.

References

Van den Berg, B. 2016. Coping with Information Underload: Hemming in Freedom of Information through Decision Support. In *Information, Freedom and Property: The Philosophy of Law Meets the Philosophy of Technology*, edited by Mireille Hildebrandt and Bibi van den Berg. Abingdon, Oxon UK: Routledge.

Chess, D. M., Palmer, C. C., and White, S. R. 2003. Security in an Autonomic Computing Environment. *IBM Systems Journal* 42:1, 107–18.

Christian, B, and Griffiths, T. 2016. *Algorithms to Live By: The Computer Science of Human Decisions*. New York: Henry Holt and Co.

Citron, D. K. 2007. Technological Due Process. Washington University Law Review 85, 1249–1313.

Citron, D. K. and Pasquale, F. 2014. The Scored Society: Due Process for Automated Predictions. Washington Law Review 89:1, 1–33.

Dewey, J. 1927. *The Public & Its Problems*. Chicago: The Swallow Press.

Epstein, R, and Robertson, R. E. 2015. The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections. *Proceedings of the National Academy of Sciences* 112:33, E4512–21.

Esposito, E. 2011. *The Future of Futures: The Time of Money in Financing and Society*. Cheltenham: Edward Elgar Publishing.

Floridi, L. 2014. *The Onlife Manifesto - Being Human in a Hyperconnected Era*. Dordrecht: Springer.

Gabor, D. 1963. *Inventing the Future*. Secker & Warburg.

Gigerenzer, G. 2000. *Adaptive Thinking: Rationality in the Real World*. Oxford; New York: Oxford University Press.

Goodman, B, and Flaxman, S. 2016. European Union Regulations on Algorithmic Decision-Making and A "right to Explanation". *arXiv:1606.08813 [Cs, Stat]*, June. http://arxiv.org/abs/1606.08813.

Halpern, S. 2014. The Creepy New Wave of the Internet. *The New York Review of Books*. http://www.nybooks.com/articles/2014/11/20/creepy-new-wave-internet/.

Hildebrandt, M. 2011. Autonomic and Autonomous "Thinking": Preconditions for Criminal Accountability. In *Law, Human Agency and Autonomic Computing*. Abingdon: Routledge.

———. 2012. The Dawn of a Critical Transparency Right for the Profiling Era. In *Digital Enlightenment Yearbook 2012*. Amsterdam: IOS Press. 41–56.

———. 2015a. *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology*. Cheltenham: Edward Elgar.

———. 2015b. Radbruchs Rechtsstaat and Schmitt's Legal Order: Legalism, Legality, and the Institution of Law. *Critical Analysis of Law* 2:1. http://cal.library.utoronto.ca/index.php/cal/article/view/22514.

———. 2016. New Animism in Policing: Re-Animating the Rule of Law? In *The SAGE Handbook of Global Policing*, edited by B. Bradford, B. Jauregui, I. Loader, and J. Steinberg. London: SAGE.

Hildebrandt, M., and Gutwirth, S. 2007. (Re)presentation, pTA Citizens' Juries and the Jury Trial. *Utrecht Law Review.* Accessed at http://www.utrechtlawreview.org/. 3, 1.

Hildebrandt, M, and Rouvroy, A. 2011. *Law, Human Agency and Autonomic Computing. The Philosophy of Law Meets the Philosophy of Technology*. Abingdon: Routledge.

Latour, B. 2000. The Berlin Key or How to Do Words with Things. In *Matter, Materiality and Modern Culture*, edited by P.M. Graves-Brown. London: Routledge. 10–21.

Marres, N. 2005. *No Issue, No Public. Democratic Deficits after the Displacement of Politics*. Amsterdam, available via: http://dare.uva.nl.

Metz, C. 2016. Artificial Intelligence Is Setting Up the Internet for a Huge Clash With Europe. *WIRED*, July 11.

Mitchell, T. M. 2006. *Machine Learning*. Mcgraw-Hill Computer Science Series

Mouffe, C. 2000. Deliberative Democracy or Agonistic Pluralism. Department of Political Science, Institute for Advanced Studies, Vienna.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.

Sen, Amartya. 2004. Elements of a Theory of Human Rights. *Philosophy & Public Affairs* 32: 4, 315–56.

Steels, Luc. 1995. When Are Robots Intelligent Autonomous Agents? *Robotics and Autonomous Systems.* 15: 3–9.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Thaler, Richard H., Cass R. Sunstein, and John P. Balz. 2010. Choice Architecture. SSRN Scholarly Paper ID 1583509. Rochester, NY: Social Science Research Network. <http://papers.ssrn.com/abstract=1583509>.

Waldron, Jeremy. 2008. The Concept and the Rule of Law. *Georgia Law Review* 43:1, 1.

———. 2011. Are Sovereigns Entitled to the Benefit of the International Rule of Law? *European Journal of International Law* 22:2, 315–43.

David H. 2013. Ubiquity Symposium: Evolutionary Computation and the Processes of Life: What the No Free Lunch Theorems Really Mean: How to Improve Search Algorithms. *Ubiquity* 2013 (December): 2:1, 2-15.

Wynne, Brian. 1995. Public Understanding of Science. In *Handbook of Science and Technology Studies*, edited by Sheila Jasanoff, Gerald E. Markle, James C. Petersen, and Trevor Pinch. Thousand Oaks, CA: Sage. 361–89.