

I do not need to know how a bus works, for taking the bus!?

Why and how to boost people's risk literacy for public discourse on scoring

Felix G. Rebitschek

Imagine that a bus driver does what the bus says, the bus takes routes that you do not know, and you are the only passenger, and you know that the bus sometimes breaks down unexpectedly – you better should know how a bus works and the whole traffic system – or avoid taking the bus at all. It is not about choosing the right bus to take. It is not about complying with the expectations towards the passenger. It is not even about understanding how to mend the bus. It is about your participation in decisions about where and how public transportation should be available.

Aiming at empowered citizens in terms of superscoring means to assure their influence on the digital environment, equal to or larger than in the analogous environment. This ideal of an educative approach to laypeople who face scoring systems equipped with methods of machine learning, however, seems counted out before commenced. Accordingly, my presentation will be guided along the main objections against efforts to build up algorithm sovereignty of the general population.

- 1) Most citizens would not be aware of scoring issues.
- 2) Most citizens would not be interested in scoring algorithms.
- 3) Most citizens could not be taught about scoring algorithms.
- 4) It would not be relevant to teach citizens about scoring algorithms.
- 5) Aiming at scoring-empowered citizens would be like cosmetics undermining regulatory efforts.

What is algorithm sovereignty? It does not mean to strengthen sovereignty of users, for instance by increasing self-control in using digital services. It does not mean to strengthen data sovereignty, for instance by giving an actually informed consent in data processing. It is even not about information sovereignty, how people find appropriate and quality-assured information. Algorithm sovereignty actually is about being the object of others' analyses, or better, about putting control on that.

Most citizens would not be aware of scoring issues.

One major issue, for sure, is that scoring algorithms like tests and any instrument to draw inferences from observations are not perfect. So, they produce errors, in classification, for instance misses and false alarms. The general population in Germany has expectations about how many errors different scoring algorithms make. Gerd Gigerenzer, Gert G. Wagner, and I surveyed the innovation sample of the socioeconomic panel about credit scoring, recidivism prediction, people analytics, and health

behaviour assessment (Rebitschek, Gigerenzer, & Wagner, in preparation). The expectations about error rates vary across domains and they are not so different from expectations about experts in place of decision-support algorithms. But this representative household survey tells us: the accepted error rates are much lower than the expected rates, particularly for algorithms. A quasi-representative online study from 2017 from the RisikoAtlas project already showed us that the majority of people in Germany does not stick to the illusion of certainty any longer, when asked about the certainty of different test results (Rebitschek, Jenny, & McDowell, in preparation). There has been a development of the public about the reliability of evidence and about how certain conclusions are. Most citizens expect low but substantial errors of algorithms, of both types in classification, and they seem to consider how potential costs vary per error (Rebitschek, Gigerenzer, & Wagner, in preparation).

Most citizens would not be interested in scoring algorithms.

This can be falsified rather directly, as we know from a representative survey commissioned by the Advisory Council for Consumer Affairs that most citizens of the German population show preferences about the features that enter scoring systems: specifically, telematics driver scoring systems, health behavioural proto-scoring, and credit scoring as well (SVRV, 2018). Given that, surveying the audience of a scoring cinema event showed remarkable reflections on scoring-related surveillance, personal data and privacy concerns with potential for abuse, and on behavioural control with loss of autonomy, but also on norms and behaviour standardisation, on validity of scoring outcomes, economization with misplaced incentives and loss of solidarity, and on the potential for social interaction disorders (Rebitschek, Groß, Keitel, Brümmer, Gigerenzer, & Wagner, 2018). Now imagine that hot brands such as Apple or LV install consumer programs that underline their exclusivity. Climate protection, for instance, is an excellent marketing vehicle, that allows to justify a green consumer scoring, according to which only those who prove sufficiently surveilled climate-friendly activities across many life domains have access to high-class products. There would be a lot of consumer interest in understanding how to comply with this algorithm.

Most citizens could not be taught about scoring algorithms.

One may feel to agree with this statement when it comes to teaching programming languages. However, this is not meant by algorithm sovereignty. Algorithm literacy must focus on the functional concepts on which the algorithms are based. The goal is that every citizen knows what she has to check before taking part in a scoring. Algorithm sovereignty should thus be stimulated in schools and adult education by learning what to ask: What is the purpose of the scoring algorithm in question? How has this purpose been achieved so far? How valid is the predicted target and what are the consequences of having chosen this target? Which possible benefits and damages for the use of algorithms have been determined on an individual, social and societal level? Which characteristics of

the individual are included and to what extent? How good and representative were the data with which the algorithm was 'built'? What is the quality of the data that the algorithm uses to score? What is the quality and reliability of the algorithm, on average and for individuals with certain feature combinations? How are the different types of algorithm errors weighted and to what extent do those affected agree with this balance? Which qualities of fairness does the algorithm match?

To empower consumers accordingly, in the RisikoAtlas (2019) project – which is funded by the Ministry of Justice and for Consumer Protection – part of our research aimed at different educative consumer tools which address three levels of questions: how to make an informed scoring system choice (e.g. telematics in car insurances), how to avoid feeding in potential scoring systems (e.g. people analytics), and how to recognize that informed participation is impossible (e.g. health bonus programmes). We show that consumers equipped with simple decision trees can detect when they cannot make informed choice and understand what they can do in order to increase personal control with regards to scoring.

It would not be relevant to teach citizens about scoring algorithms.

One may argue that improving algorithm literacy of the general population is wasted effort, because experts need to deal with complex scoring systems primarily: developers, auditors, and regulators. However, improving algorithm literacy of those who are subject to scoring programs is an essential precondition for positive benefit-harm ratios of algorithm implementations (Gigerenzer, Rebitschek, & Wagner, 2018). Like medical interventions in health care, any intended influence of scorings, shown at the laboratory level, reaches its limits with real-world implementation. Instead of planned effect strengths, side effects will occur that affect many citizens strongly and are hardly controllably. The chance for a positive benefit-harm ratio is tied to the ability of the scored people to detect, assess, and correct undesirable effects from implemented algorithms. Here not only algorithm literacy but risk literacy comes into play: a set of features and skills that allows the individual to act on problems of uncertainty (also and particularly in case of decision support) in a way that implies a positive benefit-harm ratio for oneself.

Aiming at scoring-empowered citizens would be like cosmetics undermining regulatory efforts.

It is not a question of transparency – even if you mean ‘comprehensibility’ for lay people about why a certain output is produced. I am just not able to consent encroaching my personal rights in an informed way, particularly if the potential by my data is unknown. So, it is not in the hands of the individual to defend herself against encroachments on her fundamental rights. Yet as a scoring model has the potential to determine the chances of scored individuals to which extent they can participate in economy and society, the regulation has to defend fundamental rights primarily. However,

regulatory initiatives are fed by citizens and voters who need more than just cosmetics for public discourse.

With regards to the perspective of Germany being a highly innovative nation, it could become a matter of research funding. When you think about the first years of Internet in Germany, much research funding went into data security and usability. Since the beginning of the 21st century the preferences of customers about algorithms went into research focus; opinions, and evaluations. And in the last three years knowledge transfer became the subject, as awareness and knowledge about algorithms did. Research projects about interpretable algorithms preceded research on how to design the digital environment for increasing its benefit-harm ratio for the inhabitants. But this is still remarkably limited because it deals with the deceptive idea of analogous preconditions in the digital and analogous environment and it ignores the critical difference that you can reverse, reject, or replace an algorithmic environment. Citizens do not have to accept it. Research is needed about how to empower citizens successfully so that they question the design of algorithmic environments and produce political pressure that restricts or corrects undesirable environments.

Acknowledgments

The RisikoAtlas project is supported (was supported) by funds of the Federal Ministry of Justice and Consumer Protection (BMJV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme.

References

Gigerenzer, G., Rebitschek, F. G., & Wagner, G. G. (2018). Eine vermessene Gesellschaft braucht Transparenz. *Wirtschaftsdienst*, 98(12), 860-868.

Rebitschek, F. G., Gigerenzer, G., & Wagner, G. G. (in preparation). Western people accept and expect different types and levels of decision errors from experts and algorithms: a large-scale representative survey.

Rebitschek, F. G., Groß, C., Keitel, A., Brümmer, M., Gigerenzer, G., & Wagner, G. G. (2018). Dokumentation einer empirischen Pilot-Studie zum Wissen über und zur Bewertung von Verbraucher-Scoring (Working Paper / Sachverständigenrat für Verbraucherfragen). Berlin: Sachverständigenrat für Verbraucherfragen.

Rebitschek, F. G., Jenny, M. A., & McDowell, M. (in preparation). Risk literacy across six European countries.

RisikoAtlas (2019). A research project to promote risk literacy. Press statement.
<https://www.risikoatlas.de>

SVRV (2018). Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen. Berlin: Sachverständigenrat für Verbraucherfragen.