

TRUST, AUTOMATION BIAS AND AVERSION: ALGORITHMIC DECISION-MAKING IN THE CONTEXT OF CREDIT SCORING

Rita Gsenger^{1,*} and Toma Strle²

¹University of Vienna, Vienna University of Economics
Vienna, Austria

²University of Ljubljana
Ljubljana, Slovenia

DOI: 10.7906/indecs.19.4.7
Regular article

Received: 7 August 2021.
Accepted: 27 September 2021.

ABSTRACT

Algorithmic decision-making (ADM) systems increasingly take on crucial roles in our technology-driven society, making decisions, for instance, concerning employment, education, finances, and public services. This article aims to identify peoples' attitudes towards ADM systems and ensuing behaviours when dealing with ADM systems as identified in the literature and in relation to credit scoring. After briefly discussing main characteristics and types of ADM systems, we first consider trust, automation bias, automation complacency and algorithmic aversion as attitudes towards ADM systems. These factors result in various behaviours by users, operators, and managers. Second, we consider how these factors could affect attitudes towards and use of ADM systems within the context of credit scoring. Third, we describe some possible strategies to reduce aversion, bias, and complacency, and consider several ways in which trust could be increased in the context of credit scoring. Importantly, although many advantages in applying ADM systems to complex choice problems can be identified, using ADM systems should be approached with care – e.g., the models ADM systems are based on are sometimes flawed, the data they gather to support or make decisions are easily biased, and the motives for their use unreflected upon or unethical.

KEY WORDS

algorithmic decision-making, credit scoring, trust, automation bias, algorithmic aversion

CLASSIFICATION

APA: 2910

JEL: O3

*Corresponding author, *η*: rita.gsenger@wu.ac.at; +436 991 072 6227;
Heinrich-Collin-Str. 8-14/6/24, 1140 Vienna, Austria

INTRODUCTION

The process of decision-making is prone to a diverse range of biases that can lead, at least in certain contexts, to erroneous judgments or disadvantageous choices (e.g., [1-4]). From the perspective of classical (especially economic) models of decision-making – where the decision-maker is seen as a kind of a globally rational agent (e.g., [5]), approximately capable of and motivated to maximise her utility – the range of contexts where people systematically fall prey to erroneous judgment or make disadvantageous choice is remarkable. For instance: people are quite prone to irrelevant anchors when making judgments or choices [4, 6]; much more responsive to losses than gains [7]; people's choices are strongly affected by how choice problems are formulated [8, 9]; people are likely to choose smaller, immediate rather than larger, more distant rewards [10]; prone to remain with default choices even if disadvantageous [11] and quite indecisive [12]; people even choose disadvantageously from the perspective of their own happiness [13]; and are even blind to reasons of their own, seemingly deliberate and simple, choices [14]. Decision-makers are, in a rather important sense, quite bounded in their “rationality”. As Herbert Simon lucidly states: “[...] the concept of “economic man” (and, I might add, of his brother “administrative man”) is in need of fairly drastic revision ... Broadly stated, the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist.” [5; p.99].

Several ways of amending the imperfection of human decision-making are available. One strategy is to try to educate decision-makers through various debiasing strategies (e.g. [15]). Another is to modify choice environments and thus help decision-makers make better decisions – e.g., the strategy of the nudge programme [11]. Another solution, increasingly used in administrative and economic sectors, is to use algorithms to support human decision-making or to entrust decision-making to algorithmic systems altogether.

Some would say that using algorithms for decision-making purposes promises a reduction of biases in judgment and decision-making as they are, for instance, able to consider more information [16]. Furthermore, some have argued that ADM systems enable fairer and more objective decisions, since algorithms are not, for instance, affected by emotions [17], or because their decision-making process is, at least in principle, more transparent and accountable than humans' [18]. Moreover, ADM can provide relief from cognitive workload of users and decision-makers having to make choices in a rather complex world [19-21]. ADM systems have been, for these and other reasons, employed in many different contexts, such as to determine loans [16, 22] and insurance premiums [23], to investigate tax evasion [24], to calculate credit scores [25-27], to predict the likelihood of criminal activity [16, 28-30], in policing [31-33], healthcare [34, 35] and within social media platforms [36, 37].

All in all, the ubiquitous use of ADM systems has widespread economic and social consequences, as it is transforming business sectors and creating new ways of social organisation [16]. It must be noted, however, that the consequences of using such systems can be quite dire: from biased models leading to disadvantaging the already marginalized groups to enabling new and effective ways of manipulating people's behaviour [16]. The consequences, risks, and ethical questions within ADM should thus be taken seriously and critically reflected upon [16, 18, 38, 39]. Considering the whole range of consequences and risks involved in using ADM systems surpasses the scope of this article.

Instead, we aim to understand how people's attitudes towards and beliefs about ADM systems affect their use, influence, and success of application. In the first part of the article, we briefly discuss some main characteristics and types of ADM systems. In the second, we introduce the functioning and employment of ADM systems and argue that their results are

often perceived differently from human recommendations. In the third part, we focus on trust, automation bias, complacency, aversion, and the resulting behaviours, drawing on research from various areas such as psychology, human-computer-interaction, and cognitive science. In the fourth part, we apply findings from previous research to the context of credit scoring where such research is scarce. Credit scoring is an interesting use case for ADM systems primarily due to its application in various domains having a pervasive influence on many areas of people's lives. Moreover, in credit scoring customers can hardly object to a credit score calculated by an ADM system. To affect their credit scores certain groups of people, for instance, engage in "strategic data-generating performances" [25, p.349], deliberately creating data to produce a more favourable credit score [25]. We end the article with a brief discussion of ADM systems for credit scoring from the perspective of human-centric approach to AI and touch upon certain ethical issues and challenges.

ALGORITHMIC DECISION-MAKING SYSTEMS

Various kinds of automated or partially automated systems that support decision-making can be distinguished. Castelluccia and Le Métayer [18] suggest that three classes of systems can be distinguished. First, systems that aim to improve knowledge and technology by analysing big datasets to support, for example, climate forecasts or research in healthcare (e.g., to assist the process of discovering a new virus). Second, systems that support decisions by making recommendations and predictions, utilised, for instance, "to improve logistics (optimal product placement in stores, optimal road constructions or the frequency of refuse collection), finance (real-time auctions) or security (automated detection of vulnerabilities in computer systems)" [18; p.5]. Systems of that category are used additionally to optimize and improve services that have been performed by humans so far. Third, systems that enable inanimate objects to act and decide to some extent on their own. In this context, the algorithms, for example, autonomous cars or robots, are making decisions on behalf of the users. (Anthropomorphic systems such as robots were excluded from our analysis as these might elicit different reactions due to their form [40].)

Some ADM systems allow operators or users control over recommendations, suggestions, and the degree of the systems' use. For instance, in social media, users can disable suggestions about personalised advertisements or content [36, 37]. In the context of managerial or governmental decisions, however, people often have little choice in following recommendations and using these systems [20]. Generally, different levels of automation can be distinguished: from no assistance by the computer to full automation whereby the ADM system decides and/or acts autonomously. The computer could, for instance, provide multiple or only one option, wait for the approval or allow a veto by a human operator, provide information only if asked or solely if the computer decides to do so. Higher automation levels might be beneficial for tasks that do not require flexibility and systems with a low chance of failure [40].

Here, we will be concerned with ADM systems giving recommendations to support the decision-making process of a user or operator, irrespective of their influence on the decision outcome. We will consider systems that perform services or parts of services that humans used to be in charge of, such as credit scoring. As we will focus on various attitudes towards ADM systems and the resulting behaviours of users or operators, we will consider different kinds of individuals influenced by or related to ADM systems: operators and users (like customers of a bank), developers and designers of the systems, etc.

ATTITUDES TOWARDS ALGORITHMIC DECISION-MAKING SYSTEMS

Perceptions of and attitudes towards ADM systems can be investigated on several different levels. First, the perception varies between stakeholders and people, such as designers and

developers, operators, users, the media and the public. Second, the attitude toward the systems might affect the perceptions of their decisions [20]. Therefore, the latter needs to be considered as it influences the successful employment of such systems. Third, the perception can vary according to the influence of individual aspects, for instance, the cultural background of the users [41] or their expertise and knowledge [19, 42]. Overconfidence in the algorithmic systems might lead, for instance, to their unnecessary use [43]. Furthermore, peers and/or the media might alter certain expectations people have about ADM systems. Moreover, they cause a different perception of the advice given by ADM systems compared to advice from humans, even if the content of the advice itself is the same [42]. Algorithmic systems might be preferable to human decision-makers in some contexts as they outperform human experts in prediction across different domains such as climate forecasts [44], the discovery of new viruses [18], and clinical diagnosis [45].

According to Lee and See [46], the perception of and the beliefs about ADM systems might be positive due to various reasons: First, users and operators might judge them more apt and objective [47, 48], as more information is readily available to them [28, 46]. Second, the systems are less influenced by emotions. Therefore, their decision-making might be more competent [20]. Third, in some instances, users and operators perceive systems as value-neutral in their decision-making [28]. However, for some choices, intuition is understood as useful or even required – accordingly, users might perceive systems as less competent in decision-making [49].

TRUST IN ALGORITHMIC DECISION-MAKING SYSTEMS

Algorithms can only be useful to support human decision-making if users, operators, and stakeholders trust them [50]. Gaining trust is influenced by expectations [51], familiarity [52, 53] and non-verbal cues during an interaction [54]. In psychological, behavioural, and neuroscientific research, trust has been described as an attitude [46], a behaviour [55], a relationship [56], and a brain activation pattern [57]. Trust in any automated system includes specific influences such as reliability, utility, robustness, and a false-alarm rate [58]. Moreover, research has shown that some people tend to trust automated systems more and perceive them as more reliable than human individuals, a phenomenon called the automation bias [59] (for details on the automation bias, see section 2.3). Overall, trust in automated systems depends largely on performance, such as the response time of the system [60]. The process of establishing trust depends on the operator's knowledge about the system, its design features, and other situational influences such as the expertise of the truster [41].

The participants of a study about the trustworthiness of ADM systems [20] regarded both humans' and algorithms' decisions as equally trustworthy if they concerned scenarios of mechanical tasks. Algorithmic decisions were perceived as less trustworthy in more human tasks, such as scheduling in the workplace. Most participants were aware that an algorithmic system could exhibit glitches. Therefore, no participant trusted the algorithm without reservations [20].

Adopting trust towards automated systems facilitates the navigation of complexity, replaces supervision, and enables reliance when the system is too complex to be understood completely. Reliance, however, cannot always be accurate in terms of the capabilities of the automated system. Blind reliance on automated systems can be just as detrimental to its application as not trusting the system at all. If operators trust the system blindly, mistakes might not be detected. If they do not trust it at all, cooperative decision-making is not possible. Trusting the system too much or not enough might be described in terms of misuse and disuse: "Misuse refers to the failures that occur when people inadvertently violate critical assumptions and rely on automation inappropriately, whereas disuse signifies failures that

occur when people reject the capabilities of automation” [46; p.50]. Inappropriate reliance resulting in disuse and misuse of automation is frequently caused by a mismatch between the system’s capabilities and the trust invested. This discrepancy is described in terms of (i) calibration, (ii) overtrust, and (iii) resolution [46]. The first aspect, calibration, refers to a mismatch between the trust invested and the system’s capabilities. Overtrust concerns the phenomenon of trusting the system too much due to poor calibration. Resolution describes “how precisely a judgment of trust differentiates levels of automation capability” [46; p.55]. If the resolution is low, large changes in the system are met with small variations of trust. Misuse and disuse of automated systems can be decreased by greater specificity – meaning the flexible adaptation of trust over time, high resolution, and good calibration of trust in the system’s capabilities [46].

Estimating a system’s capabilities correctly and placing enough trust in it depends on the knowledge about its capabilities and functioning, as a study by Alexander et al. [19] on over – or underreliance on ADM systems has shown. In the study, participants were given recommendations by ADM systems in a problem-solving game. To make the choice nontrivial, participants had to pay the algorithm to support them in making money. Participants had to solve two-dimensional mazes, getting a reward of 5 \$ if they solved one in 60 seconds or less. The support of the algorithms would cost 2 \$ and they could either adopt the suggestions of the algorithm or ignore them. Participants were in one of four conditions with varying information about the system: the first group was not given any information about the suggested algorithm; the second one was told the algorithm had a 75 % accuracy rate; the third group was told that 54 % of people used this algorithm; the fourth group was told that 70 % of people used it. The study measured the neurophysiological response, cardiac rate, and behaviour of participants to determine if they relied too much or not enough on the algorithm. By measuring heart rate variability, researchers determined the cognitive load and arousal of participants. According to the study, the social proof was the most effective tool in convincing people of the algorithmic system. Moreover, the study found that the adoption of the algorithm reduced cognitive load in all conditions. That might suggest that participants did not monitor the algorithm after its adoption. Therefore, the performance was lowest when participants included the algorithm. Generally, the attention of participants adopting the algorithm was lower than non-adopters’, but it was still higher than baseline, meaning they did pay attention to what the algorithm was doing.

AUTOMATION COMPLACENCY

In supporting human decision-makers, ADM systems are often designed to reduce erroneous judgments. However, they can cause other types of errors, such as automation complacency, leading to disadvantageous decisions. Automation complacency is defined as a human operator monitoring an automated system and missing a system failure or malfunction due to substandard monitoring [21]. Complacency as well as automation bias (for details see the next section 2.3.) were first researched in the aviation sector [61]. There, pilots, air traffic controllers, and other responsible personnel, can underestimate threats and work under the assumption that everything is fine, even though there is evidence to the contrary. Their negligence ultimately results in an accident. The term automation complacency was coined regarding automated aviation systems [21]. Operators of automated systems mostly passively observe and control the functioning of the system. Even as that has increased speed and efficiency, automation also gave rise to the misuse of automation [62]. Due to the assumption that everything is working correctly, operators insufficiently inspect automated systems compared to systems under manual control. Consequently, system malfunction or failure might be missed, or reactions might be delayed [62].

Parasuraman and Manzey [21] have shown that complacency occurs especially for highly reliable systems. The detection rate of failures increases if the system is not entirely reliable. However, these results vary depending on the expertise of the participants. Accordingly, Parasuraman and Manzey [21] distinguish between complacency potential and behaviour, whereby the latter occurs only if the potential is given with other circumstances, such as a high workload [21]. Moreover, research indicates that complacency might be a compensatory mechanism to deal with a high workload [62].

Complacency in ADM systems has been researched in the context of the control problem. The problem arises when operators supervise a task execution, which is increasingly the case, for instance, in aviation, where the plane flies automatically while the pilot monitors the situation. When using reliable automated systems, pilots might become “complacent, overreliant or unduly diffident when faced with the outputs” [63; p.556]. The complacency effect affects novices as well as experts and might have damaging consequences such as accidents. Generally, the less human intervention is necessary for a system to function, the greater the likelihood of the control problem occurring [63].

More recently, automation complacency has been observed in decision-making systems based on machine learning. For instance, in predictive policing, officers go on the recommended route without challenging it. Otherwise, they would have to justify the divergence from a fixed procedure dictated by the algorithm [64]. In another example, as shown by Eubanks [38], caseworkers who are responsible for child welfare in Pennsylvania and work in a governmental agency using an ADM system were more inclined to adapt their own risk estimates to the model’s estimates instead of taking advantage of the scope of action they had [38].

AUTOMATION BIAS

Automation bias refers to the human tendency of ignoring or not inquiring about contradictory information about a computer-generated solution, which is accepted as correct [61]. Moreover, automation bias is enforced when a system gives the wrong advice, whereas complacency occurs if the system does not give advice, even though it should [65]. Automation bias is defined, similarly to other biases, as the use of a “heuristic replacement for vigilant information seeking and processing” [21; p.391], but contrary to other decision biases, it results specifically from the interaction with an automated system [21].

Parasuraman and Manzey [21] identify three causes of automation bias: 1) The cognitive-miser hypothesis, stating that humans prefer to reduce their cognitive load and thus decide according to simple decision rules and comprehensive heuristics, which might result in automation bias, as operators do not undertake any thorough analysis; 2) Automated systems are perceived as powerful agents, believed to have more analytic capabilities than humans, and thus they are trusted more; 3) Responsibility might be handed over to the automated system, as people try to reduce their own contributions when work is shared. If the automated system is part of a team, other team members might reduce their efforts and refrain from analysing additional aspects or inspecting the automated system’s decisions. Studies suggest, however, that systems that support analysis and information integration are less prone to lead to automation biases than systems, which recommend specific actions due to their analysis [21]. Therefore, automation bias might occur due to cognitive overload and could be reduced by decreasing cognitive load [35]. According to Parasuraman and Manzey [21], the effects of automation bias are twofold: First, operators following incorrect recommendations are committing a commission error. Second, an error of omission happens when operators neglect a critical situation because they were not informed by the system (see, for instance, the Enbridge Pipeline Disaster [66]).

Reducing cognitive load, trusting algorithmic systems more than humans, and handing over responsibility increase the occurrence automation bias. Moreover, the degree to which operators perceive themselves socially accountable – for instance, when in direct interaction with a customer to whom they must justify a decision – plays a role regarding the frequency of omission and commission errors. People who feel accountable were more thoroughly examining the decisions taken by an algorithmic system and verifying them more often [21, 67]. Accountability creates pressure for people to include more information and process it more thoroughly. Moreover, accountability enables decision-makers to “employ more multi-dimensional, self-critical and complex information processing strategies and to put more effort into identifying appropriate responses” [67; p.703].

ALGORITHMIC AVERSION

The previous sections highlighted the misuse of automated systems due to overestimation, complacency, and bias. Here we describe the phenomenon of algorithmic aversion where a negative attitude towards ADM systems might lead to a disregard of their help compared to a person’s advice [68]. As Dietvorst et al. spell out various characteristics of algorithmic aversion: people often “prefer humans’ forecasts to algorithms’ forecasts, [...] more strongly weigh human input than algorithmic input, [...] and more harshly judge professionals who seek out advice from an algorithm rather than from a human” [69; p.114].

Additionally, the type of the decision plays a role regarding the degree of aversion towards an ADM system. Lee [20] has shown that participants have similar emotional responses to decisions made by algorithms and humans if the decision requires solely mechanical skills and does not require subjective judgment or emotions. Conversely, regarding human skills, emotions towards the systems’ decisions were more negative compared to humans’ decisions. Generally, participants felt less positive about managerial decisions made by algorithmic systems [20].

After giving bad advice, advice utilisation decreases more for an automated system than for a human advice giver. This probably happens because people expect automated systems to be more “perfect” compared to “flawed” human beings. Moreover, participants might have confidence in human advisors to perceive and correct their own errors [70]. People were prone to prefer the advice of ADM systems compared to humans before mistakes in their decisions were known to them [69, 70]. This seems to be consistent with multiple studies done by Logg et al. [71] where people clearly preferred the recommendations by ADM systems (regarding forecasts of popular songs, romantic attraction, and numeric estimates).

Expectations about the functioning and the capabilities of such systems shape the users’ perceptions [42]. Multiple studies done by Yeomans et al. [72] reveal a considerable aversion towards ADM systems if the functioning of the systems is not known. Participants preferred human recommendations as they reportedly understood them better. Therefore, the researchers conclude that not only increased accuracy, as often suggested, but also an understanding of ADM systems would decrease aversion. Overall, the aversion against these systems depends to some degree on the understanding of and knowledge about those systems [72]. These studies contradict the findings of Logg et al. [71] where experts on ADM systems and forecasts tended to rely less on the recommendations by ADM systems than lay people.

The degree of aversion depends on the expectations and beliefs about the algorithm influence, the need for control by users, and the capability of alignment with the outcome of an algorithmic decision. Often, ADM systems dominate the decision-making instead of enabling a transparent process that includes the algorithm and the human in an aligned decision-making process. Algorithmic aversion develops if the consequences following a decision are not the same as expected, and the human user loses confidence in the system [42].

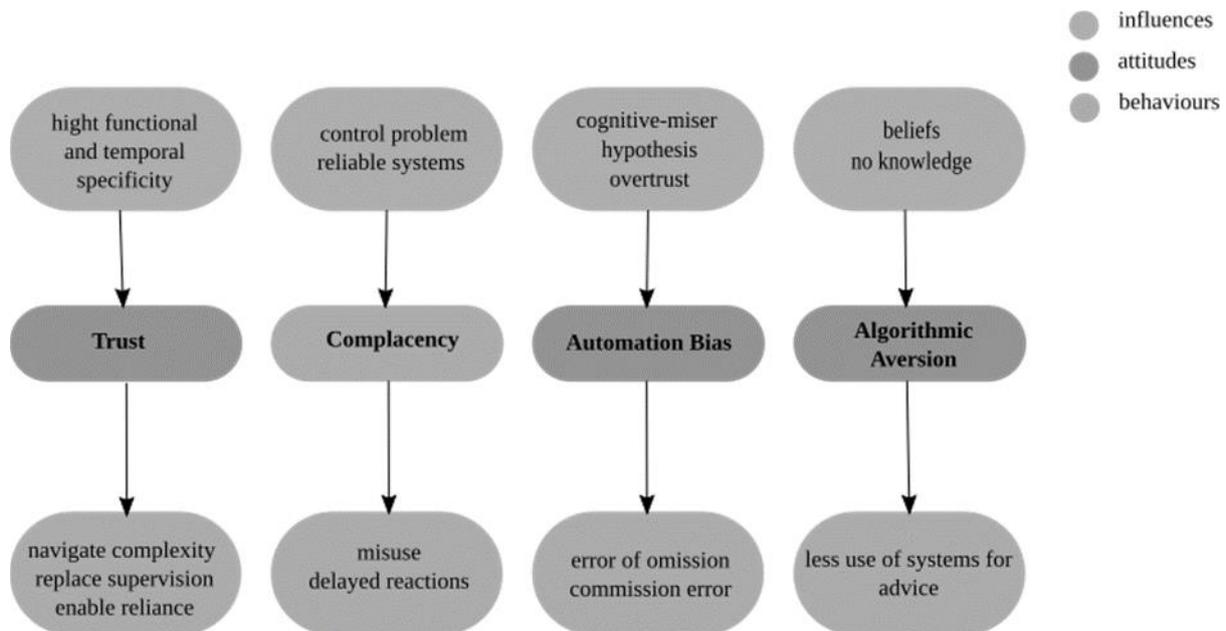


Figure 1. Summary of influences on and effects of trust, complacency, automation bias and algorithmic aversion.

CREDIT SCORING

In this part, the functioning and the use of credit scoring is first briefly explained. Second, since insight into the attitudes towards ADM systems within credit scoring is limited, we consult research and theory on attitudes towards ADM systems from other domains (presented in the previous sections) and apply it to credit scoring.

WHAT IS CREDIT SCORING?

Traditionally, if the customer of a bank wanted to get a loan, she would need to go to the bank and be interviewed. Subsequently, a credit manager would evaluate her trustworthiness, reliability, and the risk of defaulting [16]. Furthermore, the evaluation would rely on her financial status and factors such as marital status, gender, address, employment, housing and criminal history.

However, the evaluation system has been increasingly computerized since the 1950ies, leading to centralization and standardization of the evaluation process. Due to computerization, requests can be processed faster, and unskilled workers are hired instead of skilled bankers, reducing personnel costs [26]. Moreover, the evaluation systems seem to have removed the human element from credit scoring by excluding personal contact between borrower and lender. By doing so, the systems promised to remove the prejudice credit managers might have towards customers. Additionally, they “reduced personal creditworthiness to the sum of statistical probabilities” [26; p.232]. An applicant is perceived as a part of a risk population, determined by demographic and economic qualities. Therein, intervention of credit managers is viewed as a distortion. Moreover, the process of determining creditworthiness is seemingly separated from characteristics such as honesty, responsibility, and morality of the customer [26]. Recently and increasingly, companies are trying to improve the credit scoring system by evaluating creditworthiness based on personality traits such as patience, impulsiveness, risk preference, and trustworthiness, among others [73], including social network data [27].

Credit scoring is used worldwide in different domains such as individual loans, mortgages [74] or contracts such as telephone contracts [75]. The mechanism applies a statistical model “that tries to predict the future behaviour of accounts and customers based on data from the same or a similar group of accounts and customers from the relatively recent past” [74; p.59]. Every business uses its own model based on different methodologies. In the US, the Fair Isaac Corporation (FICO) developed an algorithm used by three major credit reporting agencies. The algorithm itself is unknown but it is most likely based on the ratio of debt to available credit [73]. Dozens of different commercially available scores for different kinds of debts can be distinguished, such as credit card debt, personal loans, or automobile loans [26]. A model produces a certain score, wherein a low score would mean low quality and high risk. The acceptance of a credit application and the credit conditions are decided according to the determined score. Edelman, however, points out that credit scoring is additionally a business process, which includes the “data quality, credit policy, profitability, model stability and what to do with decisions that the bank or the branches think are not correct” [74; p.60]. Until the late 1980s, only data from the credit application informed the decision of the credit scoring system, meaning there was one evaluation to determine creditworthiness. Nowadays, however, the creditworthiness is evaluated continuously after the acceptance of a credit application. These evaluations are done by scoring algorithms that examine if customers pay in time and if they are still profitable for the credit bureau, developing risk models. Risk is low enough if they are “carrying an interest-generating balance without maxing out” [26; p.252], meaning defaulting on the credit. Simultaneously, the risk and the performance of an individual can be tracked across all her accounts, even if she borrowed money from multiple credit bureaus [26]. Scorecards play an important role in determining creditworthiness. These are tools embedded in software packages to select customers and calculate credit scores. The software includes “back-stage statisticians, electronic data warehouses, risk managers, and front-stage marketing campaigns” [76; p.284]. Scorecards differ according to the way they translate the conditions in which risk is analysed.

Calculating credit scores by using statistical models can operate under more certainty if data about the consumers are available, which do not stem from the consumers themselves. Digital data analysis allows for replacing dependency of information given by the consumers directly to the credit lending institution [76]. The data used for statistical credit scoring comprises up to 400 variables provided by credit reference agencies [74]. In the era of big data, often no differentiation between credit data and other data is made and many other variables which are not connected to the credit history of a customer are included [27, 77-80]. For instance, an applicant’s college, her use of capital letters in applications (whereby the use of all caps writing is interestingly a warning sign) and social media data [26], including online tracking and behavioural profiling [79, 81]. Moreover, data harvested by specific apps from smartphones might be included [82]. Furthermore, other network-based data is included, developing a social credit score based on the individuals’ position in a social structure [79]. Additionally, the inclusion of network data allows targeted advertising of credit products to new customers [83] and the inclusion of individuals who previously did not have access to credit [77].

As for the previously used consumer credit assessment, which only included data from the application made by the customers concerning their credit history, three factors are generally assessed through credit scoring: Stability, honesty, and the ability to repay a credit. This assessment is usually repeated on a regular basis [74]. Paying back debt depends not only on the ability to do so but also on the willingness to pay. Behavioural tendencies, for instance, trustworthiness, reliability, impulsivity, and risk attitude are used as defining characteristics to determine individuals’ willingness to pay back the debt [73]. For a detailed history of credit assessment and the quantification of creditworthiness, see [26]. For a historical perspective on the FICO score, see [76].

Credit scores determine evaluations of other areas of people's lives as well. For instance, some employers use the credit score to determine if the customer is a responsible employee and trustworthy [73]. Individuals who pay their bills on time are presumed to be responsible in the workplace as well, not accounting for many other factors that could cause a bad credit score. That might lead to a negative feedback loop as people with bad credit subsequently have more difficulties finding a job, making their credit even worse [16]. Moreover, sometimes tax inspectors use credit scores to decide whom to investigate [74].

ATTITUDES TOWARDS CREDIT SCORING ALGORITHMS

Different attitudes and behaviours are formed while using ADM systems that we have described in chapter two: trust, complacency, automation bias, and algorithmic aversion. Each of these attitudes has different causes, dependencies, and results (see Figure 1). Some of them can be observed or applied to the use of ADM systems in credit scoring.

In the context of credit scoring, systems barely permit human intervention. Often, employees are required to use the system to calculate the credit score and the contract's conditions [26]. Moreover, not being able to influence the credit score even if it is perceived as unjust might increase algorithmic aversion.

Furthermore, credit decisions are not mechanical, making them possibly inept to be taken over by such systems. Algorithms used for credit scoring are criticized for introducing standardization in a highly complex area and entailing negative consequences for individuals [16].

Burton et al. [42] suggest that ADM systems as support along every step of a decision-making process would enable their adoption by more users and include more application domains. Such differentiated systems might be beneficial for credit scoring as well, as increasing flexibility and adaptation to the needs of customers. Burton et al. [42] define the successful use of ADM systems as the shared decision-making capability of the human operator and the system. Complete trust in or disregard of the systems point to the failure of the interaction between the operator and the system, even as in some application domains, full automation is beneficial [42]. That, however, does not seem to be the case for ADM systems used in credit scoring, as a shared agency, and avoidance of complacency and aversion could benefit the person affected by a credit score. Combining the capacities of the systems, considering large quantities of data on the one hand and the knowledge of human circumstances and individual situations, on the other hand, could make such systems more successful. Overall, a more human-centred approach to ADM systems would be beneficial for their successful use, as such an approach could solve problems of accuracy, bias, and transparency [84].

A study done by Schäufele [75] shows that operators of credit score systems often do not question the decisions of the system even if they could, indicating automation bias. Operators in that study did not see any reason to question the system even though they could object to decisions by filing a complaint. They seem to give away responsibility for the decision to the ADM system. Moreover, operators reduce their cognitive load by relying on the system too much for the complex credit score calculation [75]. These factors indicate automation bias in credit scoring systems.

As previously mentioned, automation bias might lead to commission and omission errors [21]. In the case of credit scoring, commission errors seem to occur, as the type of contract a person gets depends on the inference of data from a profile made about her, using a model that might be biased. The commission error has dire consequences for some, who receive, for instance, bad conditions for a credit, which keeps them in debt. These consequences are especially difficult as a bad credit score influences other domains such as employment. In

consequence, people have more difficulty paying back their debt as they are not hired. Automation bias thus results in a negative feedback loop [16].

Furthermore, uncritically accepting credit scores could let customers believe the systems' recommendations without comparison to other credit providers [62], thus causing complacency [16].

All in all, credit scoring would need to account for the unique and complex life situations of very different individuals and situations. And although ADM systems promise to reduce biases of decisions about credit scores, and the systems have more data available and a bigger processing capacity, the models employed are still human-made, mostly not user-centred, and thus criticized as biased [16], perceived to be unfair to certain individuals and can lead to negative feedback loops. For instance, some groups of people, to circumvent a negative credit score, even "play [...] the credit score game" [25; p.346], finding strategies to improve their credit score (also to enable upward social mobility [78]) by producing positive data; for instance, by joining lending circles where people lend money to each other without interest to build credit [85] (see also [86-88] for similarly created loops within systems, rich with social interaction).

HUMAN-CENTRIC ADM SYSTEMS FOR CREDIT SCORING

The attitude towards ADM systems is crucial for their successful and beneficial use. A survey conducted among U.S. adults in 2018 by the Pew Research Center shows that 31 % of participants deem using automated decision-making systems for credit scoring acceptable for companies. Respondents who would not consider such a system acceptable voiced concerns such as the violation of privacy, the accuracy of online data representing a person, and the irrelevance of online habits and behaviours for an individual's creditworthiness [89]. The exploration and increasing use of alternative data sources for credit scoring, including social media data or information from smartphones [79, 82, 83, 90], is perceived rather critically in research [27, 83] and by most participants as inquired by the Pew Research Center [89]. Users and customers (people who these systems decide about) seem to have a different attitude towards ADM systems than managers and executives of companies or institutions who decide about these systems. The latter seem to emphasize the timeliness and efficacy of their companies due to using these systems [28]. To alleviate users' concerns and to make ADM systems successful and possibly fairer, a human-centric framework is necessary.

A human-centred framework provides strategies to use AI systems to improve capabilities instead of replacing the workforce, including "*human factors design* to ensure AI solutions are explainable, comprehensible, useful, and usable" [91; p.44] (emphasis in original). Ethical design principles to ensure fairness and justice [91] are crucial for an AI systems' human-centric framework. The social accountability of operators and managers is important to establish and maintain the fairness of systems. Studies show that participants who know to be accountable when using ADM systems committed significantly fewer omission and commission errors than control groups [21]. Another study by Lee and Baykal [92] found that interpersonal power as well as knowledge of programming influence the attitude towards decisions by mathematically fair algorithms compared to group decisions. (Fair division algorithms use a mathematical definition of fairness, which in most cases uses equity as a central concept. Equity, in contrast to equality, does not advocate for treating every person the same, but accounts for individual differences to guarantee a fair distribution [92].) Their results show that participants perceived decisions made through group discussions as fairer. Furthermore, algorithmic decisions were thought to be unfair if the algorithm did not "account for multiple concepts of fairness and cognitive and social behaviours in groups, such as the presence of altruism and group dynamics" [92; p.1035]. The authors attribute the

increased perception of fairness in group discussions to the decision's transparency and the possibility of individual group members' intervention. As individuals were held accountable, the perception of fairness increased [92]. Overall, adapting the decision-making algorithm to specific tasks by increasing functional and temporal specificity might ensure fairness and reduce algorithmic aversion. Moreover, the social accountability of the operators could reduce omission and commission errors, especially for group decisions.

Aside from social accountability, a legal framework is necessary to regulate the use of ADM systems and ensure algorithmic accountability. The General Data Protection Regulation 2016/679 (GDPR) grants several new rights to citizens of member states of the European Union, including the right to be forgotten and to have their data deleted (Art. 17) or rectified (Art. 16), the right to be informed about ADM systems and their use, including the consequences of such systems and profiling (Art. 13) and the right not to be subject of an automated decision, which includes profiling (Art. 22). Exceptions to Article 22 of the GDPR can be granted if (1) the data subject's informed consent is provided, (2) if the legislation of a member state allows such decision-making, or (3) if the decision is necessary to fulfil a contract (Art. 22(2)). Furthermore, the GDPR grants the right to a human reassessment of the system's decision if perceived as unfair or incorrect [93].

Critics claim that the legal framework provides too much freedom to data controllers and insufficiently protects individuals [94]. Furthermore, as ADM systems are very complex, the information should be presented in a comprehensible manner for each individual, and the system's "intentions" made clear [94]. What is problematic is that, not all parties involved have a right to an explanation, for instance, the general public. Providing the public with information concerning the ADM systems' functioning, however, would be beneficial to reduce public concerns and to improve individuals' understanding of the use of their data [94]. By increasing knowledge about these systems, social accountability could be created, and influence could be exercised to make these systems more human-centric. Especially regarding credit-scoring systems, which possess sensitive data about individuals, creating accountable and transparent systems is crucial to ensure a fair distribution of credit.

CONCLUSION

This article aimed to identify peoples' attitudes towards ADM systems and ensuing behaviours when dealing with ADM systems with a particular consideration for credit scoring.

Trust and algorithmic aversion were identified as common attitudes adopted towards ADM systems, automation bias and complacency as key behaviours. Trust, complacency, automation bias, and algorithmic aversion were consulted to shed light on the attitudes towards ADM systems for credit scoring. In credit scoring, all these aspects could be identified, complacency resulting in overreliance and automation bias engendering commission errors by the operators, causing the misuse of ADM systems. Moreover, complacent users might not question the credit score assigned to them. ADM systems' decisions could be most beneficial for the service providers because they might be designed to find the most cost-efficient solutions leading to complacency by the operators and managers. These solutions might not be the most beneficial for the users or customers who must live with the consequences.

Furthermore, aversion could be influential for operators and users. First, credit scoring systems do not allow for human interference, and second, they might not be perceived as fair. Multiple strategies are suggested in research to reduce errors and biases, such as highly functional and temporal specificity and a human in the loop, to reduce complacency effects [46]. Moreover, social accountability and transparency of the decision-making process by the algorithmic

systems might be useful strategies to establish trust on the one hand and reduce bias, aversion, and complacency on the other hand. The design of human-centred ADM systems would benefit customers and operators alike. That, however, would require designing systems based on explainability and transparency instead of data that are often opaque and biased but are used due to easy access and availability [89].

All in all, ADM systems are increasingly used, influencing decisions made by companies, policymakers, and individuals [18]. Even as these systems are frequently advertised to be more objective and reliable than human decision-makers [28, 46], and many advantages in applying ADM systems to complex choice problems can be identified, using ADM systems should be approached with care since they are sometimes based on biased models, and the motives for their use unreflected upon or unethical. Often unreflected use of ADM systems might thus too easily result in dire consequences for individuals and, more often than not, for the already disadvantaged groups [16, 38, 39].

REFERENCES

- [1] Kahneman, D.: *A perspective on judgment and choice: Mapping bounded rationality*. American Psychologist **58**(9), 697-720, 2003, <http://dx.doi.org/10.1037/0003-066X.58.9.697>,
- [2] Kahneman, D. and Klein, G.: *Conditions for intuitive expertise: A failure to disagree*. American Psychologist **64**(6), 515-526, 2009, <http://dx.doi.org/10.1037/a0016755>,
- [3] Thaler, R.H. and Sunstein, C.R.: *Nudge: Improving Decisions about Health, Wealth and Happiness*. Yale University Press, London, 2008,
- [4] Tversky, A. and Kahneman, D.: *Judgment under Uncertainty: Heuristics and Biases*. Science **185**(4157), 1124-1131, 1974, <http://dx.doi.org/10.1126/science.185.4157.1124>,
- [5] Simon, H.A.: *A Behavioral Model of Rational Choice*. The Quarterly Journal of Economics **69**(1), 99-118, 1955, <http://dx.doi.org/10.2307/1884852>,
- [6] Englich, B.; Mussweiler, T. and Strack, F.: *Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making*. Personality and Social Psychology Bulletin **32**(2), 188-200, 2006, <http://dx.doi.org/10.1177/0146167205282152>,
- [7] Kahneman, D. and Tversky, A.: *Prospect Theory: An Analysis of Decision under Risk*. Econometrica **47**(2), 263-291, 1979, <http://dx.doi.org/10.2307/1914185>,
- [8] Ruggeri, K., et al.: *Replicating patterns of prospect theory for decision under risk*. Nature Human Behaviour **4**, 622-633, 2020, <http://dx.doi.org/10.1038/s41562-020-0886-x>,
- [9] Tversky, A. and Kahneman, D.: *The Framing of Decisions and the Psychology of Choice*. Science **211**(4481), 453-458, 1981, <http://dx.doi.org/10.1126/science.7455683>,
- [10] Green, L.; Fry, A.F. and Myerson, J.: *Discounting of Delayed Rewards: A Life-Span Comparison*. Psychological Science **5**(1), 33-36, 1994, <http://dx.doi.org/10.1111/j.1467-9280.1994.tb00610.x>,
- [11] Sunstein, C.R.: *Default Rules Are Better Than Active Choosing (Often)*. Trends in Cognitive Sciences **21**(8), 600-606, 2017, <http://dx.doi.org/10.1016/j.tics.2017.05.003>,

- [12] Anderson, C.J.: *The psychology of doing nothing: Forms of decision avoidance result from reason and emotion*.
Psychological Bulletin **129**(1), 139-167, 2003,
<http://dx.doi.org/10.1037/0033-2909.129.1.139>,
- [13] Hsee, C.K. and Hastie, R.: *Decision and experience: why don't we choose what makes us happy?*
Trends in Cognitive Sciences **10**(1), 31-37, 2006,
<http://dx.doi.org/10.1016/j.tics.2005.11.007>,
- [14] Johansson, P.; Hall, L.; Sikström, S. and Olsson, A.: *Failure to detect mismatches between intention and outcome in a simple decision task*.
Science **310**(5745), 116-119, 2005,
<http://dx.doi.org/10.1126/science.1111709>,
- [15] Soll, J.B.; Milkman, K.L. and Payne, J.W.: *A User's Guide to Debiasing*.
In: Keren, G.; Wu, G., eds.: *Wiley-Blackwell Handbook of Judgment and Decision Making*.
Wiley Blackwell, Chichester, pp.924-952, 2015,
- [16] O'Neil, C.: *Weapons of Math Destruction. How big data increases inequality and threatens democracy*.
Crown, New York, 2016,
- [17] Tolan, S.: *Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges*.
European Commission, Seville, 2018,
- [18] Castelluccia, C. and Le Métayer, D.: *Understanding algorithmic decision-making: Opportunities and Challenges*.
European Parliamentary Research Service, Scientific Foresight Unit (STOA) PE 624.261, 2020,
- [19] Alexander, V.; Blinder, C. and Zak, P.J.: *Why trust an algorithm? Performance, cognition, and neurophysiology*.
Computers in Human Behaviour **89**, 279-288, 2018,
<http://dx.doi.org/10.1016/j.chb.2018.07.026>,
- [20] Lee, M.K.: *Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management*.
Big Data & Society **5**(1), 1-16, 2018,
<http://dx.doi.org/10.1177/2053951718756684>,
- [21] Parasuraman, R. and Manzey, D.H.: *Complacency and Bias in Human Use of Automation: An Attentional Integration*.
Human Factors **52**(3), 381-410, 2010,
<http://dx.doi.org/10.1177/0018720810376055>,
- [22] Lohr, S.: *Big Data Underwriting for Payday Loans*.
<https://bits.blogs.nytimes.com/2015/01/19/big-data-underwriting-for-payday-loans>,
- [23] De Mayer, J.: *The use of big data and artificial intelligence in insurance*.
BEUC. The European Consumer Organisation, 2020,
- [24] De Laat, P.B.: *Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?*
Philosophy and Technology **31**(4), 525-541, 2018,
<http://dx.doi.org/10.1080/03085147.2017.1412642>,
- [25] Kear, M.: *Playing the credit score game: algorithms, 'positive' data and the personification of financial objects*.
Economy and Society **46**(3-4), 346-368, 2017,
<http://dx.doi.org/10.1080/03085147.2017.1412642>,
- [26] Lauer, J.: *Creditworthy: a history of consumer surveillance and financial identity in America*.
Columbia University Press, New York, 2017,
- [27] Rosenblatt, E.: *Credit Data and Scoring. The First Triumph of Big Data and Big Algorithms*.
Academic Press, London, 2020,

- [28] Christin, A.: *Algorithms in practice: Comparing web journalism and criminal justice*. Big Data & Society **4**(2), 1-14, 2017, <http://dx.doi.org/10.1177/2053951717718855>,
- [29] Chiao, V.: *Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice*. International Journal of Law in Context **15**(2), 126-139, 2019, <http://dx.doi.org/10.1017/S1744552319000077>,
- [30] Zweigl, K.A.; Wenzelburger, G. and Krafft, T.D.: *On Chances and Risks of Security. Related Algorithmic Decision Making Systems*. European Journal of Security Research **3**, 181-203, 2018, <http://dx.doi.org/10.1007/s41125-018-0031-2>,
- [31] Harcourt, B.E.: *Against prediction: profiling, policing, and punishing in an actuarial age*. University of Chicago Press, Chicago, 2007,
- [32] Kubler, K.: *State of urgency: Surveillance, power, and algorithms in France's state of emergency*. Big Data & Society **4**(2), 1-10, 2017, <http://dx.doi.org/10.1177/2053951717736338>,
- [33] Bennett Moses, L. and Chan, J.: *Algorithmic prediction in policing: assumptions, evaluation, and accountability*. Policing and Society **28**(7), 806-822, 2018, <http://dx.doi.org/10.1080/10439463.2016.1253695>,
- [34] Reich, A.: *Disciplined doctors: The electronic medical record and physicians' changing relationship to medical knowledge*. Social Science & Medicine **74**(7), 1021-1028, 2012, <http://dx.doi.org/10.1016/j.socscimed.2011.12.032>,
- [35] Lyell, D. and Coiera, E.: *Automation bias and verification complexity: a systematic review*. Journal of the American Medical Informatics Association **19**(1), 121-127, 2016, <http://dx.doi.org/10.1136/amiajnl-2011-000089>,
- [36] Bucher, T.: *The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms*. Information, Communication & Society **20**(1), 30-44, 2017, <http://dx.doi.org/10.1080/1369118X.2016.1154086>,
- [37] Eslami, M., et al.: *'I always assumed that I wasn't really that close to [her]': Reasoning about Invisible Algorithms in News Feeds*. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, Seoul, pp.153-162, 2015, <http://dx.doi.org/10.1145/2702123.2702556>,
- [38] Eubanks, V.: *Automating Inequality. How high-tech tools profile, police, and punish the poor*. St. Martin's Press, New York, 2018,
- [39] Barocas, S. and Selbst, A.D.: *Big Data's Disparate Impact*. California Law Review **104**(3), 671-732, 2016,
- [40] De Visser, E.J., et al.: *A Little Anthropomorphism Goes a Long Way: Effects of Oxytocin on Trust, Compliance, and Team Performance With Automated Agents*. Human Factors **59**(1), 116-133, 2017, <http://dx.doi.org/10.1177/0018720816687205>,
- [41] Hoff, K.A. and Bashir, M.: *Trust in automation integrating empirical evidence on factors that influence trust*. Human Factors **57**(3), 407-434, 2015, <http://dx.doi.org/10.1177/0018720814547570>,
- [42] Burton, J.W.; Stein, M.K. and Jensen, T.B.: *A systematic review of algorithm aversion in augmented decision making*. Journal of Behavioural Decision Making **33**(2), 220-239, 2020, <http://dx.doi.org/10.1002/bdm.2155>,

- [43] Ashton, A.H.; Ashton, R.H. and Davis, M.N.: *White-Collar Robotics: Levering managerial decision making*.
Management Review **37**(1), 83-109, 1994,
<http://dx.doi.org/10.2307/41165779>,
- [44] Jones, N.: *How machine learning could help to improve climate forecasts*.
Nature **548**(7668), 379-380 2017,
<http://dx.doi.org/10.1038/548379a>,
- [45] Grove, W.M., et al.: *Clinical versus mechanical prediction: A meta-analysis*.
Psychological Assessment **12**(1), 19-30, 2000,
<http://dx.doi.org/10.1037/1040-3590.12.1.19>,
- [46] Lee, J.D. and See, K.A.: *Trust in Automation: Designing for Appropriate Reliance*.
Human Factors **46**(1), 50-80, 2004,
http://dx.doi.org/10.1518/hfes.46.1.50_30392,
- [47] Gillespie, T.: *The relevance of algorithms*.
In: Boczkowski, P.J.; Foot, K.A. and Gillespie, T., eds.: *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press, Cambridge, pp.167-194, 2014,
- [48] Christin, A.: *From Daguerreotypes to Algorithms: Machines, Expertise, and Three Forms of Objectivity*.
ACM SIGCAS Computers and Society **46**(1), 27-32, 2016,
<http://dx.doi.org/10.1145/2908216.2908220>,
- [49] Akter, S., et al.: *Analytics-based decision-making for service systems: A qualitative study and agenda for future research*.
International Journal of Information Management **48**, 85-95, 2019,
<http://dx.doi.org/10.1016/j.ijinfomgt.2019.01.020>,
- [50] Rader, E. and Wash, R.: *Trustworthy Algorithmic Decision-Making*.
<http://bitlab.cas.msu.edu/trustworthy-algorithms>,
- [51] Castelfranchi, C. and Falcone, R.: *Trust theory: a socio-cognitive and computational model*.
John Wiley & Sons, Chichester, 2010,
- [52] Tjøstheim, T.A.; Johansson, B. and Balkenius, C.: *A Computational Model of Trust-, Pupil-, and Motivation Dynamics*.
Proceedings of the 7th International Conference on Human-Agent Interaction. Association for Computing Machinery, Kyoto, pp.179-185, 2019,
- [53] Alarcon, G.M.; Lyons, J.B. and Christensen, J.C.: *The effect of propensity to trust and familiarity on perceptions of trustworthiness over time*.
Personality and Individual Differences **94**, 309-315, 2016,
<http://dx.doi.org/10.1016/j.paid.2016.01.031>,
- [54] DeSteno, D., et al.: *Detecting the Trustworthiness of Novel Partners in Economic Exchange*.
Psychological Science **23**(12), 1549-1556, 2012,
<http://dx.doi.org/10.1177/0956797612448793>,
- [55] Fehr, E.; Fischbacher, U. and Kosfeld, M.: *Neuroeconomic Foundations of Trust and Social Preferences: Initial Evidence*.
American Economic Review **95**(2), 346-351, 2005,
<http://dx.doi.org/10.1257/000282805774669736>,
- [56] Resnik, D.B.: *Scientific Research and the Public Trust*.
Science and Engineering Ethics **17**(3), 399-409, 2011,
<http://dx.doi.org/10.1007/s11948-010-9210-x>,
- [57] Krueger, F. and Meyer-Lindenberg, A.: *Toward a Model of Interpersonal Trust Drawn from Neuroscience, Psychology, and Economics*.
Trends in Neurosciences **42**(2), 92-101, 2019,
<http://dx.doi.org/10.1016/j.tins.2018.10.004>,
- [58] Hoffman, R.R.; Johnson, M.; Bradshaw, J.M. and Underbrink, A.: *Trust in Automation*.
IEEE Intelligent Systems **28**(1), 84-88, 2013,
<http://dx.doi.org/10.1109/MIS.2013.24>,

- [59] De Visser, E.J., et al.: *Almost human: Anthropomorphism increases trust resilience in cognitive agents*.
Journal of Experimental Psychology: Applied **22**(3), 331-349, 2016,
<http://dx.doi.org/10.1037/xap0000092>,
- [60] Efendić, E.; Van de Calseyde, P.P.F.M. and Evans, A.M.: *Slow response times undermine trust in algorithmic (but not human) predictions*.
Organizational Behavior and Human Decision Processes **157**, 103-114, 2020,
<http://dx.doi.org/10.1016/j.obhdp.2020.01.008>,
- [61] Cummings, M.L.: *Automation bias in intelligent time critical decision support systems*.
AIAA 1st Intelligent Systems Technical Conference, American Institute of Aeronautics and Astronautics, Chicago, 2004,
<http://dx.doi.org/10.2514/6.2004-6313>,
- [62] Bahner, J.E.; Hüper, A.-D. and Manzey, D.: *Misuse of automated decision aids: Complacency, automation bias and the impact of training experience*.
International Journal of Human-Computer Studies **66**(9), 688-699, 2008,
<http://dx.doi.org/10.1016/j.ijhcs.2008.06.001>,
- [63] Zerilli, J.; Knott, A.; Maclaurin, J. and Gavaghan, C.: *Algorithmic Decision-Making and the Control Problem*.
Minds & Machines **29**(4), 555-578, 2019,
<http://dx.doi.org/10.1007/s11023-019-09513-7>,
- [64] Villani, C.: *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*.
Mission assigned by the Prime Minister Édouard Philippe, 2018,
- [65] Wickens, C.D.; Clegg, B.A.; Vieane, A.Z. and Sebok, A.L.: *Complacency and Automation Bias in the Use of Imperfect Automation*.
Human Factors **57**(5), 728-739, 2015,
<http://dx.doi.org/10.1177/0018720815581940>,
- [66] Wesley, D. and Dau, L.: *Complacency and Automation Bias in the Enbridge Pipeline Disaster*.
Ergonomics in Design **25**(1), 17-22, 2017,
<http://dx.doi.org/10.1177/1064804616652269>,
- [67] Skitka, L.J.; Mosier, K. and Burdick, M.D.: *Accountability and automation bias*.
International Journal of Human-Computer Studies **52**(4), 701-717, 2000,
<http://dx.doi.org/10.1006/ijhc.1999.0349>,
- [68] Önkal, D., et al.: *The Relative Influence of Advice From Human Experts and Statistical Methods on Forecast Adjustments*.
Journal of Behavioural Decision Making **22**(4), 390-409, 2009,
<http://dx.doi.org/10.1002/bdm.637>,
- [69] Dietvorst, B.J.; Simmons, J.P. and Massey, C.: *Algorithm aversion: People erroneously avoid algorithms after seeing them err*.
Journal of Experimental Psychology: General **144**(1), 114-126, 2015,
<http://dx.doi.org/10.1037/xge0000033>,
- [70] Prahla, A. and Van Swol, L.: *Understanding algorithm aversion: When is advice from automation discounted?*
Journal of Forecasting **36**(6), 691-702, 2017,
<http://dx.doi.org/10.1002/for.2464>,
- [71] Logg, J.M.; Minson, J.A. and Moore, D.A.: *Algorithm Appreciation: People Prefer Algorithmic to Human Judgment*.
Organizational Behavior and Human Decision Processes **151**, 90-103, 2019,
<http://dx.doi.org/10.1016/j.obhdp.2018.12.005>,
- [72] Yeomans, M.; Shah, A.; Mullainathan, S. and Kleinberg, J.: *Making sense of recommendations*.
Journal of Behavioral Decision Making **32**(4), 403-414, 2019,
<http://dx.doi.org/10.1002/bdm.2118>,

- [73] Arya, S.; Eckel, C. and Wichman, C.: *Anatomy of the credit score*. Journal of Economic Behaviour & Organization **95**, 175-185, 2013, <http://dx.doi.org/10.1016/j.jebo.2011.05.005>,
- [74] Edelman, D.: *Credit this: how the banks decide your credit score*. Significance **5**(2), 59-61, 2008, <http://dx.doi.org/10.1111/j.1740-9713.2008.00287.x>,
- [75] Schäufele, F.: *Profiling zwischen sozialer Praxis und technischer Prägung*. Springer, Wiesbaden, 2017,
- [76] Poon, M.: *Scorecards as Devices for Consumer Credit: The Case of Fair, Isaac & Company Incorporated*. The Sociological Review **55**(2), 284-306, 2007, <http://dx.doi.org/10.1111/j.1467954X.2007.00740.x>,
- [77] Aitken, R.: 'All data is credit data': *Constituting the unbanked*. Competition & Change **21**(4), 274-300, 2017, <http://dx.doi.org/10.1177/1024529417712830>,
- [78] Hurley, M. and Adebayo, J.: *Credit Scoring in the Era of Big Data*. Yale Journal of Law and Technology **18**(1), 148-216, 2017,
- [79] Wei, Y.; Yildirim, P.; Van den Bulte, C. and Dellarocas, C.: *Credit Scoring with Social Network Data*. Marketing Science **35**(2), 234-258, 2016, <http://dx.doi.org/10.1287/mksc.2015.0949>,
- [80] Saif, M.A.; Prisyazhny, A.V. and Medvedeva, M.A.: *On the Model of Credit Score Calculation Using Social Networks Data*. Marketing Science **35**(2), 234-258, 2018, <http://dx.doi.org/10.1063/1.5044043>,
- [81] Deville, J. and van der Velden, L.: *Seeing the invisible algorithm: the practical politics of tracking the credit trackers*. In: Amooore, L. and Piotukh, V., eds.: *Algorithmic life. Calculative devices in the age of big data*. Routledge, New York, pp.87-106, 2016, <http://dx.doi.org/10.4324/9781315723242>,
- [82] Lohokare, J.; Dani, R. and Sontakke, S.: *Automated data collection for credit score calculation based on financial transactions and social media*. International Conference on Emerging Trends & Innovation in ICT. IEEE, Pune, pp.134-138, 2017,
- [83] Siddiqi, N.: *Intelligent Credit Scoring. Building and Implementing Better Credit Risk Scorecards*. Wiley, Hoboken, 2017,
- [84] Springer, A.: *Accurate, Fair and Explainable: Building Human-Centred AI*. Ph.D. Thesis. UC Santa Cruz, Santa Cruz, 2019,
- [85] Cohen, P.: *In Lending Circles, a Roundabout Way to a Higher Credit Score*. New York Times, 2014, <http://www.nytimes.com/2014/10/11/your-money/raising-a-credit-score-from-zero-to-789-in-26-months.html>,
- [86] Strle, T.: *Looping minds: How cognitive science exerts influence on its findings*. Interdisciplinary Description of Complex Systems **16**(4), 533-544, 2018, <http://dx.doi.org/10.7906/indecs.16.4.2>,
- [87] Strle, T. and Markič, O.: *Looping effects of neurolaw and the precarious marriage between neuroscience and the law*. Balkan Journal of Philosophy **10**(1), 17-26, 2018, <http://dx.doi.org/10.5840/bjp20181013>,
- [88] Strle, T.: *The Image of Bounded Rationality and Feedback Effects of Modifying Choice Environments*. Cognitive Science: Proceedings of the 22nd International Multiconference Information Society – IS 2019. Institut Jožef Stefan, Ljubljana, pp.56-60, 2019,

- [89] Smith, A.: *Public Attitudes Toward Computer Algorithms*.
<http://www.pewresearch.org/internet/2018/11/16/attitudes-toward-algorithmic-decision-making>,
- [90] Taylor, A. and Sadowski, S.: *How Companies Turn Your Facebook Activity Into a Credit Score*.
<http://www.thenation.com/article/archive/how-companies-turn-your-facebook-activity-credit-score>,
- [91] Xu, W.: *Toward Human-Centred AI: A Perspective from Human-Computer Interaction*.
Interactions **26**(4), 42-46, 2019,
<http://dx.doi.org/10.1145/3328485>,
- [92] Lee, M.K. and Baykal, S.: *Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division*.
Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW. Association for Computing Machinery, Portland, 2017,
- [93] Dreyer, S. and Schulz, W.: *The General Data Protection Regulation and Automated Decision-making: Will it deliver? Potentials and limitations in ensuring the rights and freedoms of individuals, groups and society as a whole*.
Bertelsmann Stiftung, 2019,
<http://dx.doi.org/10.11586/2018018>,
- [94] Castets-Renard, C.: *Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making*.
Fordham Intellectual Property, Media and Entertainment Law Journal **30**(1), 91-137, 2019.