6-15-2020

WHY ARE WE AVERSE TOWARDS ALGORITHMS? A COMPREHENSIVE LITERATURE REVIEW ON ALGORITHM AVERSION

Ekaterina Jussupow
University of Mannheim, jussupow@uni-mannheim.de

Izak Benbasat
University of British Columbia, izak.benbasatt@sauder.ubc.ca

Armin Heinzl
University of Mannheim, heinzl@uni-mannheim.de

WHY ARE WE AVERSE TOWARDS ALGORITHMS? A COMPREHENSIVE LITERATURE REVIEW ON ALGORITHM AVERSION

Research paper

Jussupow, Ekaterina, University of Mannheim, Germany, jussupow@uni-mannnheim.de
Benbasat, Izak, University of British Columbia, Vancouver, Canada, izak.benbasat@sauder.ubc.ca
Heinzl, Armin, University of Mannheim, Germany, heinzl@uni-mannheim.de

Abstract

With technological developments in artificial intelligence, algorithms are increasingly capable to perform tasks that were considered to be unique for humans. However, literature suggests that although algorithms are often superior in performance, users are reluctant to interact with algorithms instead of human agents – a phenomenon known as algorithm aversion. But, as algorithm aversion is attracting scientific attention, empirical findings are inconclusive and papers find the opposite effect of algorithm appreciation. With this literature review, we synthesize evidence from 29 publications with 84 distinct experimental studies to investigate how algorithm characteristics and human agents' characteristics influence algorithm aversion. We show how algorithm agency, performance, perceived capabilities and human involvement as well as human agents' expertise and social distance, influence whether users develop algorithm aversion, i.e., choose humans over algorithms, utilize humans' support more often and evaluate humans' actions more favourable. Furthermore, we provide a systematic conceptualization of aversion as a biased assessment and develop propositions for future research. With our work, we contribute to algorithm aversion literature and the contemporary discussion on the impact of algorithmic agents on the future of work. We indicate that the emerging literature stream on algorithm aversion is worth considering for human-computer interaction researchers.

Keywords: Algorithm, Aversion, Artificial Intelligence, Human-Computer Interaction.

1 Introduction

As artificial intelligence (AI) gains momentum, algorithms as computer agents are increasingly being applied to tasks that were previously reserved for and performed by humans. Algorithms, enhanced by advances in AI, are already faster and more accurate than humans in multiple domains. They outperform humans in CT images analyses (e.g., Cheng et al., 2016), offer better jail-or-release decisions (Kleinberg et al., 2018), or conduct superior employee performance forecasting (Highhouse, 2008). Algorithms as chatbots have started to replace humans in customer service (Luo, Tong, Fang and Qu, 2019). The future of algorithmic autonomy seems not too far away, as machine learning enables algorithms to learn new games themselves (Silver et al., 2017) or to drive cars autonomously (Bojarski et al., 2016).

All these exciting developments in algorithms and their power lead to an interesting challenge for us, their users. How do we decide whether we want to seek the assistance of a human expert or an algorithm for the same given task, if we had the choice? How does one make such a choice rationally? Interestingly, it is not even evident that the choice of a human versus an algorithm is solely based on objective, rational criteria. Recently, researchers began to show that users do not rationally react towards algorithms and prefer human agents to algorithmic support even if algorithms were proved to be superior.

This recently identified phenomenon of *algorithm aversion* (Dietvorst, Simmons and Massey, 2015, 2018; Castelo, Bos and Lehmann, 2019) propelled a new research stream in psychology and management, investigating how user interaction with human agents differ from their interaction with algorithms. Yet, as more studies dive into this phenomenon, the experimental results become inconclusive. For instance, Logg et al. (2019) revealed that decision makers have an appreciation for algorithmic support instead of algorithm aversion as they adjust more towards estimates of algorithms than towards estimates of human agents. Furthermore, it is unclear how research on algorithm aversion relates to established research streams such as advice taking (Bonaccio and Dalal, 2006), resistance towards the use of IT (Kim and Kankanhalli, 2009) and the impact of recommender systems on decision-making (Xiao and Benbasat, 2007). As there is currently no clear conceptualization of algorithm aversion, it is largely impossible to derive specific propositions as to whether decision makers develop aversion or appreciation of the algorithm. In particular, we know very little about which algorithm characteristics and which human characteristics lead to aversion. Against this backdrop, we have conducted a systematic literature review regarding the current experimental publications of algorithm aversion to answer the following research question: *Which algorithm and human characteristics influence whether users develop algorithm aversion or appreciation?*

Our study seeks to contribute (1.) to the current literature on algorithm aversion (Dietvorst et al., 2015, 2018) by providing an overview of empirical evidence on algorithm appreciation and aversion based on algorithm and human characteristics; (2.) to the current discussion on how intelligent algorithms shape the future of work (Demetis and Lee, 2018; Burton, Stein and Jensen, 2019; Rahwan et al., 2019) by showing why humans would accept algorithms as substitute for human experts; and (3.) to the practice of introducing algorithms to perform formerly exclusive human tasks by providing an overview of factors that lead users to develop algorithm aversion.

2 The literature review methodology

Following the guidelines by Webster and Watson (2002), we conducted a detailed literature review on algorithm aversion. We focused on articles in peer-reviewed journals and conference proceedings in management, information systems and psychology written in English in the past 17 years. As we were interested in understanding the mechanisms of how the interaction with an algorithm differs from an interaction with a human agent, we focused on publications which examined algorithm aversion as an *experimental* comparison between algorithms and human agents. We used the work of Dietvorst et al. (2015, 2018) as starting point for our literature search and in addition searched the databases Web of Knowledge, EBSCO and Google Scholar with the keywords algorithm, aversion, preference, decision making and decision aid. Based on this database search a total of over 2,300 articles were identified. Through screening of the abstract and title a final list of 74 papers was developed and assessed in detail. Of those 17 were included into the review as these papers compared human agents against algorithms using experimental methods. We then identified additional publications by conducting a backward literature search based on Dietvorst et al. (2015), and also on reviews by Castelo et al. (2019) and Burton et al. (2019). We *excluded* literature that aimed to develop human-like algorithms, as we focused on direct comparison between human agent and algorithm and algorithm aversion might be a possible reason for developing deploying this type of algorithm design. Similarly, papers that examined overreliance or trust in algorithms without a comparison with a human agent were *excluded* (for example Sutherland, Harteveld and Young, 2016; Banker and Khetani, 2019). The search resulted in a total 29 publications published between January 2002 and October 2019 (see Table 1 for an overview of journals) with 84 relevant experimental investigations with over 24.000 participants (without the multinational survey by Thurman, Moeller, Helberger and Trilling (2019)).

The typical experimental approach was a between-subject design in which one group interacted with an algorithm agent while the other group interacted with a human agent. Both groups received the same task and the same dependent variables. Most papers used either a vignette description or a controlled experimental set-up. Only a few papers allowed participants to interact with a real working algorithm (see e.g., Yeomans, Shah, Mullainathan and Kleinberg, 2017). The identified experiments examine algorithm aversion with a variety of subjects (consumers, managers, student samples), in various contexts (healthcare, accounting, finance, moral decisions) and compare algorithms against different human agents (experts, the crowd, human peers). We categorised the findings of each experiment based on (i.) algorithm characteristics, (ii.) human agent characteristics and (iii.) empirical results towards aversion or appreciation. Since multiple experiments had inconclusive findings, we added the category "inconclusive" which included studies without significant effects or with findings pointing towards both aversion and appreciation.

Domain	**Journal and number of papers in parenthesis**
Neuroscience	Cognition (1); Frontiers in Human Neuroscience (1); Social Neuroscience (1); Neuron (1)
Human factors	Human Factors (3); Computers in Human Behaviour (1)
Psychology	Journal of Experimental Psychology (1); Journal of Behavioural Decision Making (3); Organizational Behaviour and Human Decision Process (1)
Marketing	European Journal of Marketing (1); Journal of Marketing Research (1); Journal of Consumer Research (1)
Management	Academy of Management Discoveries (1); Management Science (1)
Information Systems and Computer Science	Journal of Management Information Systems (1); Advances in Intelligent Systems and Computing (1); Conference proceedings (ECIS, ACM) (3)
Domain specific journals	Journal of Forecasting (1); Medical Decision Making (3); International Journal of Selection and Assessment (1); Digital Journalism (1)

Table 1. Identified literature categorized by journals

3 Conceptual background

In the following, we first define the key concepts of our literature review, followed by the conceptual model and findings.

3.1 Algorithm

The origin of algorithm aversion research stems from comparing how people react towards mathematical or computational problem-solving approaches compared to intuitive, holistic approaches (Meehl, 1954). Yet, there was no distinction as to whether humans or computer agents applied the approaches. Only recently, the mathematical approach has been combined with how a computer would solve problems (Önkal et al., 2009; Dietvorst et al., 2015) and has been labelled as *algorithmic approach*. Yet, there is no unified label in algorithm aversion literature, as papers either stick to the label of a "mathematical approach" (Önkal et al., 2009; Diab, Pui, Yankelevich and Highhouse, 2011; Dietvorst et al., 2015; Gunaratne, Zalmanson and Nov, 2018) or use new labels such as "automated system" (Madhaven and Wiegmann, 2007), "algorithm" (Boorman, O'Doherty, Adolphs and Rangel, 2013; Castelo et al., 2019; Logg et al., 2019), "computer/ program / software" (Dzindolet, Pierce, Beck and Dawe, 2002; Pezzo and Pezzo, 2006; Promberger and Baron, 2006; Arkes, Shaffer and Medow, 2007; Beck, McKinney, Dzindolet and Pierce, 2009; Wolf, 2014; Palmeira and Spassova, 2015; Fuchs, Hess, Matt and Hoerndlein, 2016; Prahl and Van Swol, 2017), "recommender system" (Yeomans et al., 2017; Thurman et al., 2019), "machine" (Bigman and Gray, 2018) or "artificial intelligence/ supercomputer" (Bigman and Gray, 2018; Jago, 2019; Longoni, Bonezzi and Morewedge, 2019; Williams et al., 2019). But, due to the rise of artificial intelligence, research on algorithm aversion developed from considering computational approaches to investigating autonomous algorithms that are a black box for human decision makers. Across all identified papers, algorithms exert a certain degree of agency and have perceived capabilities which are similar or even superior to those of humans (Dietvorst et al., 2015; Yeomans et al., 2017; Castelo et al., 2019; Longoni et al., 2019). In this paper, we synthesize publications that use different definitions of algorithms, but we consider *algorithm* as a computer agent that applies rule-based or non-rule based (i.e., machine learning based) approaches to develop an output. Also, we refer to *human agent* as the human which is used as comparison for understanding whether we interact differently with algorithms compared to humans. Moreover, we refer to *user* as the participant or decision maker who decides, interacts or evaluates the two agents, algorithm and human.

3.2 Aversion

Currently, there is no unified or accepted definition of algorithm aversion. The most influential paper on this topic by Dietvorst et al. (2015) uses the following definition:

> *Research shows that evidence-based algorithms more accurately predict the future than do human forecasters. Yet when forecasters are deciding whether to use a human forecaster or a statistical algorithm, they often choose the human forecaster. This phenomenon, which we call algorithm aversion (…)* (p.114).

However, in their extension in 2018, the same authors provided a more precise definition as "people often fail to use (algorithms) after learning that they are imperfect, a phenomenon known as algorithm aversion." (p.1155). With this formulation, the authors limit their definition of algorithm aversion to the case where individuals learn that algorithms are imperfect. However, this definition is too narrow, since some papers demonstrate that users develop algorithm aversion well before they experience the failure of the algorithm (see e.g. Longoni et al., 2019). To account also for these studies, we define algorithm aversion in a more general way *as biased assessment of an algorithm which manifests in negative behaviours and attitudes towards the algorithm compared to a human agent*. In this sense, we refer to *algorithm appreciation* (Logg et al., 2019) as *positive* behaviour and attitudes towards the algorithm. We consider algorithm appreciation as the opposite side of algorithm aversion.

There are important elements of this definition that distinguish the concept of aversion from related constructs and research streams such as technology acceptance and resistance or rejecting a system. First, research on technology acceptance (Venkatesh, Thong and Xu, 2012) describes which factors determine usage of technology, but does not consider (1.) how technology persuades users to follow its assessment and (2.) how users react to the technology compared to a human agent. From a technology acceptance perspective, the technology is considered as a tool while in algorithm aversion and appreciation research the algorithm functions as social actor (Fogg, 2002). Thus, the technology interacts actively with the user and aims to convince users of its suggestions. This means that algorithms exert a persuasive influence which is not considered by technology acceptance research. Similarly, the research on resistance assumes the role of a tool and investigates how resistance develops from different threats from the technology (Lapointe and Rivard, 2005; Kim and Kankanhalli, 2009). Resistance can be understood as preference for the status quo. This can one reason why decision-makers might be hesitant towards replacing human work with algorithm work, but cannot explain whether there are systematic differences in how algorithms and human agents are evaluated. Lastly, aversion is a subjective evaluation of the algorithm, which is systematically distorted (Tversky and Kahneman, 1974). This differentiates aversion from rejection, which can often be based on objective characteristics, such as, the quality of the systems' suggestions (Bonaccio and Dalal, 2006; Xiao and Benbasat, 2007; Wang and Benbasat, 2013).

4 Measurement of aversion and conceptual model

If users have algorithm aversion they have a biased assessment of the algorithm which they do not display towards a human agent. Thus, how users interact with the algorithm differs compared to how they interact with a human agent. In our review, we identified three measurements of aversion. First, participants can *choose between an algorithm or the human* who provides an advice or performs a task. The choice can be either binary (see e.g., Bigman and Gray, 2018) or a preference rating scale with both agents as extreme points (see e.g., Longoni et al., 2019). In this design, the information about both agents is available for participants and they can choose one agent. Aversion is salient if participants choose humans over algorithms. Second, participants can *utilize the agent's assessment* to form their own decision. In experiments using this dependent variable, participants are often asked to make their initial estimate without the agent's influence and are then provided with the agent's assessment. Participants then decide about their final estimate. Often, a quantitative score is calculated that indicates how strongly participants accounted for the agent in their own decision (e.g., Logg et al., 2019). Aversion is salient if participants adjust less towards the algorithm than towards the human. Another approach for measuring *utilizing the agent's assessment* is to ask participants in a survey about how strongly they consider the agent's assessment (see e.g., Longoni et al., 2019). Higher preference scores for humans are then interpreted as aversion. Lastly, the *agent can be evaluated* by the participants. This includes both the outcome as well as the agent. For instance, participants evaluate their trust into the agent (e.g., Madhaven and Wiegmann, 2007; Önkal et al., 2009), but also the appropriateness of the agent's decision (Palmeira and Spassova, 2015; Bigman and Gray, 2018) or authenticity of the agent's action (Jago, 2019). If the algorithm is rated less favourable than the human, this is an indicator of aversion. Often identified paper used multiple dependent variables to provide evidence for aversion or appreciation.

We base our conceptual model on prior conceptual models of human-computer interaction (see e.g., Xiao and Benbasat, 2007). However, we only consider the influence of algorithm and human agents characteristics as factors and do not consider additional potentially influential factors that were studied in the context of advice-giving systems, such as user characteristics, task characteristics, and characteristics of the interaction, due to space limitations. Based on this restriction and the three dependent variables discussed above, we use the conceptual model depicted in Figure 1, to structure the findings of our literature review.

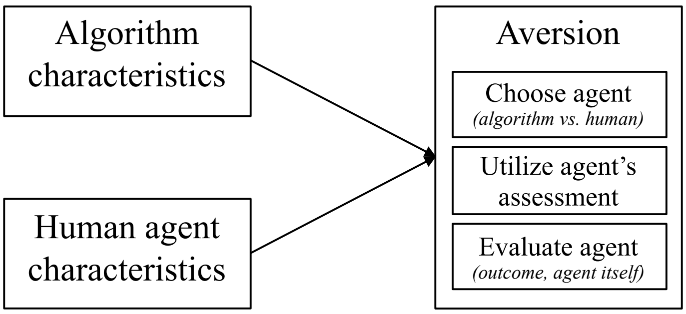

Figure 1. Conceptual model for structuring the findings of the literature review

5 Findings: algorithm characteristics

We have identified four characteristics of the algorithm that influence aversion: (1.) *algorithm agency* describes whether the algorithm advises users or performs tasks autonomously; (2.) *algorithm performance* includes algorithm failures and its reliability rate; (3.) perceived *algorithm capabilities* describes whether the algorithm is perceived to have the necessary capabilities to perform the task; (4.) *human involvement* in the algorithm describes how strongly humans in general (but not the user) are involved into training and using the algorithm. In the following, we describe each characteristic separately and relate it to the development of algorithm appreciation or aversion.

5.1 Algorithm agency

According to Nissen and Segupta (2006) and Komiak and Benbasat (2006), algorithms can be conceptualized based on their degree of autonomy as *performative algorithm* or *advisory algorithm.* A *performative algorithm* is able to accomplish independent actions by gathering information, decide and execute, leaving the human in the role to monitor outcome and algorithm performance. An *advisory algorithm,* on the contrary, only provides support to the user and does not act. The final decision remains with the user. Following Komiak and Benbasat (2006), algorithm agency is strongly influenced by how the user examines the algorithm's suggestions. In particular, when delegating the decision to the algorithm, users accept the decision without examining it in detail. However, when using the algorithm as an *advisor,* suggestions by the algorithm are considered more carefully. One reason why decision makers might interact differently with advisory compared to performative algorithms might be the loss of control (Burton et al., 2019) that decision makers experience when they interact with performative algorithms compared to advisory algorithms as the full decision authority is transferred to the algorithm. The table in the appendix provides an overview of the literature based on algorithm agency.

In the literature, performative algorithms often autonomously perform highly complex tasks such as managerial decisions (Leyer and Schneider, 2019), artistic work (Jago, 2019), selection of employees (Diab et al., 2011) or developed medical decisions which were then followed by doctors (Arkes et al., 2007; Shaffer et al., 2013). Thus, humans in general are replaced through the work of the algorithm. With performative algorithms users are left with two distinct decisions. First, they can decide whether to delegate the decision to the algorithm or not, and second, they can evaluate and monitor the outcome after the algorithm has been executed. Thus, choice of agent and use of agent (see Figure 1) are a holistic decision regarding the deployment of performative algorithms. For instance, Dietvorst et al. (2015) asked participants to bet their bonus either on the forecasting decision which they had made themselves or the decision of algorithm. The bonus payment the participants received depended upon their choice without the possibility to make joint decisions. Aversion was determined by how many participants did choose either their own estimates or the algorithm estimates. By contrast, Jago (2019) asked, among others, the participants to evaluate pieces of music or art already created either by a human or an algorithm. The study shows that decision makers preferred work performed by humans over work performed by algorithms although they evaluated exactly the same piece of work. Also, Bigman and Gray (2018) asked participants to evaluate consequences of decisions made by either algorithms or human agents. Thus, the algorithm already performed the work and then its outcome was evaluated by the user.

On the other hand, considering *advisory algorithms* takes a different perspective on algorithmic agency. In advice taking (see Bonaccio and Dalal, 2006), users can decide whether or not to integrate the advice into their reasoning and thus remain the agent of the decision. In our literature review, advisory algorithms often supported similar tasks as performative algorithms (see appendix), but left the final decision to the user. Also, literature which considered advisory algorithms often measured how strongly decision makers account for the advice as continuous variable (*advice utilization*). Decision makers could fully integrate the advice and completely follow the advice, or only partially follow the advice of the system. For instance, Logg et al. (2019) measured how strongly participants adjusted towards human or towards algorithm advice. Also, Önkal et al. (2009) calculated a relative measure which showed how strongly decisions makers estimate was influenced by the advice in a forecasting task. Also, participants could choose to interact with algorithms and evaluate the advisor or the outcome. For instance, Longoni et al. (2019) asked participants to evaluate a recommendation by a physician compared to a computer program.

Overall, the findings (see appendix) indicate that users seem to be averse towards both types of algorithms. Yet, more inconclusive results and even appreciation were found with advisory algorithms. Also, it is noteworthy, that Palmeira and Spassova (2015) experimentally tested in their second and third experiment how the perception of performative or advisory influenced algorithm aversion. They showed that that users perceived an advisory algorithm as more favourable than a performative algorithm. Thus, higher algorithm agency could be increasing algorithm aversion.

5.2 Algorithm performance

The performance of the algorithm plays a crucial role of how users interact and evaluate it. In particular, multiple studies provide evidence that decision makers struggle with algorithm errors. One explanation for this effect is based on expectation-disconfirmation theory (Bhattacherjee and Premkumar, 2004). Individuals believe that algorithms are perfect and do not make any mistakes (i.e., “perfect automation schema”) (Dzindolet et al., 2002; Madhaven and Wiegmann, 2007; Goodyear et al., 2016). When realizing, however, that the algorithm is not perfect, decision makers are more likely to blame and to punish the algorithm as an effect of this disconfirmation of their expectations. Thus, it is likely that individuals have higher expectations towards the accuracy and consistency of performance of algorithms compared to humans. However, although the algorithm might outperform the human objectively, the idea of aversion states that the performance of both, the algorithm and the human agent, seem to be evaluated in a different way (i.e., biased evaluation of the algorithm's performance). In the literature of algorithm aversion, there are three different approaches towards assessing the impact of algorithm performance: (1.) performance information; (2.) varying performance rate during interaction; and (3.) forcing algorithm failures.

The first approach provided information about the algorithm and human performance in vignettes often with a percentage score. This information was intended to show that the algorithm is well capable in performing the assigned task and is, thus, competent. Most papers did not provide any information about the performance of the algorithm or the human to reduce confounding effects of this information (Logg et al., 2019). However, if indicated, most paper used the same accuracy for the algorithm compared to the human agent. For instance, Longoni et al. (2019) described the success rate of a dermatologist and the AI system to be at 87%. However, some papers showed that the performance information does have crucial effects on the findings. For instance, Bigman and Gray (2018) demonstrated that humans do prefer another human to make a moral decision instead of an algorithm. However, in their last experiment (Study 8), the authors manipulated the accuracy rate of the algorithm and provided explicit information about it. First, the human and the algorithm had a similar accuracy rate of 75%. In the second condition, the algorithm had an accuracy rate of 95% while the humans' accuracy remained at 75%. When accuracy rates are similar, participants still preferred the human to make the moral decision. However, if the algorithm had a better accuracy rate, participants had chosen the algorithm instead. Also, Castelo et al. (2019) showed that the aversion for the algorithm relatively to a human was significantly lower when participants did not have any performance information compared to when they had performance information (Study 3). This effect was only statistically significant for objective tasks. Moreover, Yeomans et al. (2019) showed that explanations about how the algorithm performs reduces aversion and Dzindolet et al. (2002) showed that framing of the information influenced how decision makers rely on the provided algorithm. Thus, whether and how the performance information was provided did influence how strongly decision makers developed aversion.

The second approach actually varied the performance of the algorithm and let decision makers interact multiple times with this adjusted algorithm or human. This type of study was mostly applied in human-factors literature measuring a high number of interactions between humans and algorithms (Dzindolet et al., 2002; Madhaven and Wiegmann, 2007; Boorman et al., 2013; Goodyear et al., 2016, 2017; Williams et al., 2019). Often, the participants did not know about the actual reliability rate during the interaction. For instance, Goodyear et al. (2016, 2017) investigated whether participants react different to a luggage screening tool compared to a human peer. The reliability of the tool was manipulated and the tool provided good and bad advice. Overall, participants rated 64 images with the tool and their brain activity was measured in a fMRI scanner. The results of these two studies differ based on the type of performance error of the algorithm. For false positives, Goodyear et al. (2016) show that there was no difference in advice utilization for good advice, however, in the case of bad advice, the advice from the algorithm was used more frequently than the advice of the human. On the other hand, for false negatives Goodyear et al. (2017) indicate that the algorithm was more frequently discarded compared to the human. Moreover, Madhaven and Wiegmann (2007) asked participants to interact in 200 visual detection trials with a human or an algorithm. They varied both the framing of the algorithm as an expert and the actual reliability rate during the interaction. With 70% actual reliability, the authors demonstrated that participants started to use the algorithm which was previously framed as an expert less often compared to the human expert and the algorithm which was framed as novice. Compared to the first approach, decision makers often do not know about the actual reliability rate of the algorithm and develop their subjective assessment of it. Interestingly, studies which applied the second approach often did not find clear effects of algorithm aversion and even found appreciation as indicated by Goodyear et al. (2016) or Williams et al. (2019). One explanation might be that users did not know the actual reliability rate. Yet, this needs further testing.

Lastly, in the third approach, the experiment participants received explicit feedback about an algorithm failure. Papers following this approach also used faked performance feedbacks to experimentally enforce the algorithm failure situation. For instance, Dietvorst et al. (2015, 2018) showed that participants were less likely to delegate the decision to an algorithm and experienced loss beliefs in the forecasting algorithm after seeing it perform and err. Leyer and Schneider (2019) showed that decision makers develop more anger after learning about the algorithm failure compared to human failure. Interestingly, the results of publications describing algorithm failures point much more consistently towards algorithm aversion compared to the other two categories. Out of six papers which applied this approach (Pezzo and Pezzo, 2006; Dietvorst et al., 2015, 2018; Prahl and Van Swol, 2017; Bigman and Gray, 2018; Leyer and Schneider, 2019) only Bigman and Gray (2018) as well as Pezzo and Pezzo, (2006) found inconclusive results. Thus, aversion was more profound after seeing the algorithm fail compared to seeing the human agent make the same mistake. Table 2 summarizes how algorithm performance was assessed and how it influences algorithm aversion.

Algorithm performance	**Description**	**Key findings from literature**
Performance information	Provide information about the accuracy and reliability rates of the algorithm, often in percentage; sometimes richer explanation	Reduced aversion through performance information (Yeomans et al., 2017; Castelo et al., 2019); Algorithm appreciation if the algorithm was known to outperform the human and information about both agents was available (Bigman and Gray, 2018)
Performance rate	Experimentally adjust actual algorithm performance during the interaction with the participant	Findings inconsistent with evidence for both appreciation and aversion (Madhaven and Wiegmann, 2007; Boorman et al., 2013; Goodyear et al., 2016)
Algorithm failure	Forcing algorithm failure to assess whether participants evaluate the failure differently compared to a human failure	Relatively consistent findings towards aversion (Dietvorst et al., 2015, 2018)

Table 2. Overview of three approaches towards algorithm performance and the key findings from literature

5.3 Perceived algorithm capabilities

Perceived algorithm capabilities are one driver for algorithm aversion as algorithms are often perceived to lack capabilities which are necessary for the task. Especially for moral tasks, algorithms must exhibit a certain degree of expertise, autonomy and human capabilities (e.g., empathy) to make such decisions. For instance, Bigman and Gray (2018) investigated how much participants perceived that the algorithm possessed these perceived human capabilities, which were referred to as mind (i.e., skills such as thinking or communicating with others), and how the perception of mind mediated aversion. Moreover, Longoni et al. (2019) showed that users were only averse towards algorithms because they perceived that algorithms were unable to account for the users' unique characteristics. If algorithms were explicitly introduced to take into account the uniqueness of each patient case, aversion was reduced. Castelo et al. (2019) also argued that the interaction between task and perceived algorithm capabilities was an important driver for aversion. In particular, algorithms seemed not to be able to perform tasks that were highly subjective and which required perceived human capabilities while they were perceived to be better at performing more objective, quantifiable tasks. If this characteristic was taken into account, e.g. by adjusting the appearance of the algorithm such that was is more humanlike, this effect was reduced (Castelo et al. 2019). In sum, aversion is likely to decrease with higher perceived capabilities of the algorithm to perform the task.

5.4 Human involvement

Humans can be involved in two ways into the algorithm's performance and influence the perceived capabilities of the algorithm: (1.) humans ensure the quality of the algorithm as they are involved in the development and training; (2.) humans are involved to the use the algorithm as a "human-algorithm hybrid" (Palmeira and Spassova, 2015). As humans are more involved into usage of the algorithm, it appears to be more controllable (Dietvorst et al., 2018) but also as having the best of both worlds (Palmeira and Spassova, 2015).

First, humans are involved into the algorithms' development or training by providing expert knowledge. As most experiments are vignette-based, this characteristic was often described in the algorithms' introduction to create the impression that the algorithm is capable to address the task (see section 4.2. on algorithm performance). This is a way to experimentally ensure that participants acknowledge the algorithm as a valid source of information (see for instance the algorithm descripton by Dietvorst et al., 2015). However, Jago (2019) varied the type of training in his 4th experiment that an AI-based algorithm received and showed that specific types of human involvement could increase the perceived algorithm capabilities and decrease aversion. Also, Arkes et al. (2007) showed in their 3rd experiment that patients evaluated a doctor who was supported by a decision aid that was developed by a prestigious university more favourable compared to a doctor who used an unspecified decision aid. Thus, human expert training of the algorithm could reduce aversion.

Second, some studies asked participants to compare algorithm or human agents alone against human agents that work with algorithms as a hybrid. These findings are closely related towards the understanding of the algorithm agency (see section 5.1). For instance, Palmeira and Spassova (2015) identified that a hybrid provider is evaluated more favourable compared to an algorithm alone. In their experiments, Palmeira and Spassova (2015) included a hybrid in which the human makes the final decision and the algorithm had an advisory role (see section 5.1.). However, Arkes et al. (2007), Wolf (2014) and Shaffer et al. (2013) showed that physicians who used an algorithm as support were perceived to be less professional than physicians without any support. In contrast, they described the hybrid as a physician who delegated the decision towards the algorithm and, thus, increased the algorithm agency. Thus, the human was less involved compared to the study of Palmeira and Spassova (2015). Moreover, Pezzo and Pezzo (2006) combined human involvement with algorithm performance show that if physicians use an algorithm to support their decisions, failures of this hybrid are evaluated less negative compared to failures of the physician without algorithm support. Yet, positive outcomes of the hybrid are also evaluated less positive compared to positive decisions of physicians alone. Although the idea of a human-algorithm hybrid is closely related to algorithm agency, we decided to keep this type of human involvement as a distinct algorithm characteristic as it opens interesting further research questions. For instance, when comparing the objective performance, literature indicates that it's unclear whether a human-algorithm hybrid actually performs better compared to algorithms alone (Yeomans et al., 2017). Thus, while human-algorithm hybrids were perceived more favourable by users it is still unclear whether this is actually desirable. In sum, human involvement increases the perceived capabilities of the algorithm and also decreases the agency of the algorithm. Both result in a more favourable estimate of the algorithm.

6 Findings: human agent characteristics

As the literature on algorithm aversion aims to show how the interaction with algorithms differs from the interaction with human agents, all characteristics which we have described as algorithm characteristics are equally relevant for the human agent. For instance, if a study examined a performative algorithm, a performative human agent served as comparison. Also, if performance information was provided for the algorithm, the same information was provided for the human agent. However, we identified two characteristics of the human agent that were distinct from the algorithm characteristics: (1.) the expertise of the human agent; and (2.) the social distance of the human agent to the user.

6.1 Expertise

The human agent can be either framed as expert or non-expert. By framing the human agent as an expert the likelihood increases to follow its estimate as people are generally more willing to follow expert suggestions compared to non-expert suggestions (see for overview Bonaccio and Dalal, 2006). Thus, on the one hand, studies used physicians (Pezzo and Pezzo, 2006; Promberger and Baron, 2006; Arkes et al., 2007; Shaffer et al., 2013; Wolf, 2014; Palmeira and Spassova, 2015; Bigman and Gray, 2018; Longoni et al., 2019), qualified military personnel (Bigman and Gray, 2018), financial experts (Önkal et al., 2009), expert artists (Williams et al., 2019) or humans labelled as experts without clear profession (Madhaven and Wiegmann, 2007; Goodyear et al., 2016; Prahl and Van Swol, 2017; Castelo et al., 2019) as human expert agents. On the other hand, average humans or persons (Boorman et al., 2013; Bigman and Gray, 2018; Kramer, Schaich Borg, Conitzer and Sinnott-Armstrong, 2018; Jago, 2019; Logg et al., 2019), colleagues (Leyer and Schneider, 2019) or other participants from the study (Dzindolet et al., 2002; Madhaven and Wiegmann, 2007; Beck et al., 2009; Dietvorst et al., 2015) were used as non-expert agents. Only Madhaven and Wiegmann (2007) experimentally varied both, the algorithm's expertise (see algorithm performance) and the human agent's expertise. They identified that decision makers prefer the novice algorithm over the human novice while they prefer the human expert over the algorithm framed as an expert. Thus, if users could interact with human agents that had high expertise they were more likely to develop algorithm aversion.

6.2 Social distance

A different dimension of the human comparison agent is the social closeness towards the user. Instead of varying the agent expertise, some papers have chosen to use "oneself" as a comparison agent (Dietvorst et al., 2015) which is socially close to the user and increases the personal involvement. Social distance is known to have a crucial effect on how information is processed and also how different agents are perceived (Trope and Liberman, 2010). For instance, if something is perceived to be socially close, participants also relate more favourable attitudes towards it. Thus, a socially close comparison agent would influence whether users develop aversion. For instance, Beck et al. (2009) varied the personal investment of the participants into the task by experimentally manipulating whether participants choose between the algorithm or themselves or a prior participant. Participants who had the possibility to select their own estimates were more likely to disregard a superior algorithm compared to those who had the choice of prior participant's estimates. Similar findings were identified by Logg et al. (2019) in their third experiment. Participants could choose between either their own estimate or other participants' estimate and the algorithm for developing an estimated rank of US states by the number of airline passengers departing from it (task from Dietvorst et al. 2015). While participants preferred the algorithm, they preferred the algorithm less if they had the option to use their own estimate instead of using the estimate of another participant. Thus, if users could interact with a socially close agent they were more averse towards using algorithms.

7 Development of a refined model

Our review suggests that specific algorithm and human agent characteristics influence whether and how salient algorithm aversion is. Thus, we are able to refine the conceptual model from Figure 1 to a model that states specific propositions in Figure 2. While algorithm aversion emerged as a phenomenon from studying the impact computational problem-solving approaches (Meehl, 1954; see e.g., Eastwood, Snook and Luther, 2012), the importance of algorithm characteristics increased with the importance of AI-based systems. In particular, our review indicates that agency of the algorithm seems to be positively related to the development of aversion as participants develop stronger aversion with performative compared to advisory algorithms. On the contrary, performance is negatively related to aversion as lower degrees of performance through algorithm failures result in stronger aversion (see e.g., Dietvorst et al., 2015) while clearly superior performance compared to the human agent decreases aversion and can even reverse into appreciation (Bigman and Gray, 2018). Yet, it is important to note that in both cases, users evaluate the performance of the algorithm differently compared to the performance of the human. Thus, algorithmic errors are evaluated differently compared to the same error by a human agent. Moreover, if participants perceive the algorithm to be capable to perform the task (i.e., in the moral domain), they are less averse towards using the algorithm (Bigman and Gray, 2018; Castelo et al., 2019). Lastly, human involvement seems to reduce aversion through increasing the perceived capabilities of the algorithm and decreasing the agency as users prefer algorithms which were explicitly framed as trained by humans

(Jago, 2019) and perceive a human-algorithm hybrid more favourable (Palmeira and Spassova, 2015). Furthermore, on the other hand, distinct characteristics of the human comparison agent influence algorithm aversion. In particular, our review indicates that higher human agents' expertise increases aversion (e.g., Madhaven and Wiegmann, 2007). On the contrary, higher social distance of the comparison agent decreases aversion. This means that aversion seems to be higher when comparing between oneself and a different, unrelated other (see e.g., Logg et al., 2019).

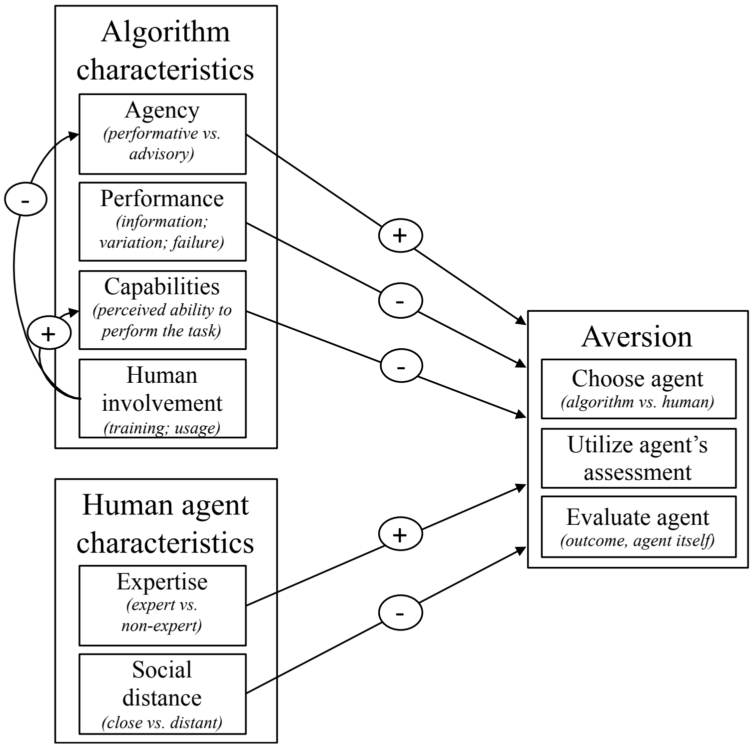

Figure 2. Adjusted model with propositions developed from literature review

8 Discussion

Overall, there are fundamental differences how we as users evaluate algorithms compared to human agents. For instance, Boorman et al. (2013) and Goodyear et al. (2016) even showed that there are differences in how our brain responds to the interaction with a human compared to an algorithm. With our review, we provided a systematic definition and developed specific propositions for future research on algorithm aversion. Our findings highlight that current literature is inconclusive because researchers on algorithm aversion often involuntary use different algorithm and human agent characteristics in their experimental investigations. Thus, we show that algorithm agency, performance, perceived capabilities and human involvement strongly influence aversion along with human agents' expertise and social distance.

With our work, we contribute to the emerging literature stream on algorithm aversion (Dietvorst et al., 2015, 2018) and algorithm appreciation (Logg et al., 2019) by showing how algorithm and human characteristics influence whether and how decision makers develop aversions against algorithms. Moreover, we provide a clearer conceptualization of what aversion is and offer specific propositions regarding its major determinants. This differentiates our work from two recent literature reviews. In comparison to Castelo et al. (2019), we indicate that the literature has not been consistent on whether participants develop aversion or appreciation and show how algorithm as well as human agent characteristics contribute to these conflicting findings. In comparison to Burton et al. (2019), we exclusively focus on experimental studies and provide a theoretical model which focuses on experimental settings. Thus, with our literature review, we provide specific propositions for future empirical research. We also contribute to the emerging research on the impact of intelligent algorithms on the future of work (Demetis and Lee, 2018; Faraj, Pachidi and Sayegh, 2018). Understanding how humans interact with algorithms compared to other humans is crucial to understand in which situations algorithm support might lead to more objective and better decisions or whether human support is needed instead. Moreover, it opens the avenue for future research on questions about how to combine different characteristics of algorithms with different human characteristics to increase acceptance of algorithms in knowledge work and society. Lastly, we contribute to practice by exhibiting which factors influence the adoption of algorithms and, thus, provide insights of how to implement these algorithms. By considering the proposed algorithm and human agent characteristics, our literature review suggests, for example, that it might be more difficult to substitute human expert work as participants prefer to consult the human expert in contrast to the algorithm. Furthermore, it is important to consider the best configurations of algorithm characteristics to mitigate aversion.

Our literature review has multiple limitations which offer directions for future research. First, we focused only on algorithm and human agent characteristics and left important further factors aside for future research. Future research should expand our proposed model by including task characteristics such as objective or subjective tasks (see e.g., Castelo et al., 2019), user characteristics such as user expertise (see e.g., Logg et al., 2019) and interaction characteristics such as the possibility for the user to influence how the algorithm works (Dietvorst et al., 2018). Furthermore, we see the need to include more contextualized factors regarding algorithm aversion (Hong et al., 2014) as, for instance, our findings indicate stronger aversion in the healthcare domain (Promberger and Baron, 2006; Shaffer et al., 2013; Longoni et al., 2019). Second, our model is based upon a synopsis of existent literature alone. Thus, future research should empirically test and validate the proposed model and evaluate whether the identified characteristics equally apply to all identified dependent variables. Moreover, future research should conduct a meta-analysis to compare different effect sizes of algorithm aversion and appreciation. Lastly, there is a lack of theorizing on the findings as most papers followed a phenomenological approach. For instance, advice taking literature (Logg et al. 2019), perfect automation schema (Dzindolet et al., 2002; Madhaven and Wiegmann, 2007) or negative attitudes towards the agent (Arkes et al., 2007; Dietvorst et al., 2015) have been proposed. The expectation-disconfirmation theory or trust formation theories might provide additional insight, but are not able to capture the full scope of phenomenon related to algorithm aversion research and thus require further research. Thus, more theorizing and more research on algorithm aversion is needed.

9 Appendix – Categorized literature

Empirical study	Finding	Tested effect of agent characteristics on finding	Underlying agent characteristics
Boorman et al. (2013)	Aversion	Performance rate (+)	Performative, non-expert
Diab et al. (2011)	Aversion	None	Performative
Jago (2019)	Aversion	Human involvement (-)	Performative
Arkes et al. (2007); Shaffer et al. (2013); Wolf (2014)	Aversion	Human involvement (-), but depends on type of involvement	Performative, expert
Promberger and Baron (2006)	Aversion	None	Expert
Fuchs et al. (2016)	Aversion	Performance information (-)	Advisory, expert
Longoni et al. (2019)	Aversion	Perceived capabilities (-) Social distance (+) Agency (-)	Advisory, expert
Önkal et al. (2009)	Aversion	None (multiple sources)	Advisory, expert
Dietvorst et al. (2015, 2018)	Aversion; Appr. in control	Algorithm failure (+) Human involvement (-)	Performative, close social distance
Bigman and Gray (2018)	Aversion; Appr. in study 9a	Performance information (-) Human involvement (-) Perceived capabilities (-) Algorithm failure (0)	Performative
Yeomans et al. (2019)	Aversion + Appr.	Performance information (-)	Advisory, non-expert
Castelo et al. (2019)	Aversion + Appr.	Perceived capabilities (+) Performance information (-)	Advisory, expert
Beck et al. (2009)	Inconclusive	Performance (+/-) Social distance (+/-)	Performative, non-expert
Leyer and Schneider (2019)	Inconclusive	Algorithm failure (+/-)	Performative, social distance, non-expert
Kramer et al. (2018)	Inconclusive	Performance information (-)	Performative, non-expert
Palmeira and Spassova (2015)	Inconclusive	Human involvement (-) Agency (-)	Expert
Dzindolet et al. (2002)	Inconclusive	Performance information (+/-)	Advisory, non-expert
Goodyear et al. (2016, 2017)	Inconclusive	Performance rate (+/-)	Advisory, expert
Madhaven and Wiegmann (2007)	Inconclusive	Performance rate (+) Perceived capabilities (+/-) Expertise (+)	Advisory
Pezzo and Pezzo (2006)	Inconclusive	Algorithm failure (+/-) Social distance (+/-)	Advisory, expert
Prahl and Van Swol (2017)	Inconclusive	Algorithm failure (+)	Advisory, expert
Gunaratne et al. (2018)	Appreciation	None	Advisory, non-expert
Logg et al. (2019)	Appreciation	Social distance (-)	Advisory, non-expert
Thurman et al. (2019)	Appreciation	None	Performative
Williams et al. (2019)	Appreciation	Performance rate (0)	Advisory, expert

Notes. – effect on finding is diminished; + effect on finding is increased, +/- complex effect, 0 no effect

References

Arkes, H. R., V. A. Shaffer and M. A. Medow. (2007). “Patients derogate physicians who use a computer-assisted diagnostic aid.” *Medical Decision Making*, *27*(2), 189–202.

Banker, S. and S. Khetani. (2019). “Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-Being.” *Journal of Public Policy and Marketing*, *38*(4), 500–515.

Beck, H. P., J. B. McKinney, M. T. Dzindolet and L. G. Pierce. (2009). “Effects of human-machine competition on intent errors in a target detection task.” *Human Factors*, *51*(4), 477–486.

Bhattacherjee, A. and G. Premkumar. (2004). “Understanding Changes in Belief and Attitude toward Information Technology Usage: A Theoretical Model and Longitudinal Test.” *MIS Quarterly*, *28*(2), 229–254.

Bigman, Y. E. and K. Gray. (2018). “People are averse to machines making moral decisions.” *Cognition*, *181*, 21–34.

Bojarski, M., D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, … K. Zieba. (2016). *End to End Learning for Self-Driving Cars*.

Bonaccio, S. and R. S. Dalal. (2006). “Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences.” *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151.

Boorman, E. D., J. P. O’Doherty, R. Adolphs and A. Rangel. (2013). “The behavioral and neural mechanisms underlying the tracking of expertise.” *Neuron*, *80*(6), 1558–1571.

Burton, J. W., M. Stein and T. B. Jensen. (2019). “A systematic review of algorithm aversion in augmented decision making.” *Journal of Behavioral Decision Making*, 1–20.

Castelo, N., M. W. Bos and D. R. Lehmann. (2019). “Task-Dependent Algorithm Aversion.” *Journal of Marketing Research*, *56*(5), 809–825.

Cheng, J.-Z., D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, … C.-M. Chen. (2016). “Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans.” *Nature Scientific Report*, *6*(1), 24454.

Demetis, D. S. and A. S. Lee. (2018). “When humans using the IT artifact becomes IT using the human artifact.” *Journal of the Association of Information Systems*, *19*(10), 929–952.

Diab, D. L., S. Y. Pui, M. Yankelevich and S. Highhouse. (2011). “Lay perceptions of selection decision aids in US and non-US samples.” *International Journal of Selection and Assessment*, *19*(2), 209–216.

Dietvorst, B., J. P. Simmons and C. Massey. (2015). “Algorithm aversion: People erroneously avoid algorithms after seeing them err.” *Journal of Experimental Psychology: General*, *144*(1), 114–126.

Dietvorst, B., J. P. Simmons and C. Massey. (2018). “Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them.” *Management Science*, *64*(3), 1155–1170.

Dzindolet, M. T., L. G. Pierce, H. P. Beck and L. A. Dawe. (2002). “The Perceived Utility of Human and Automated Aids in a Visual Detection Task.” *Human Factors*, *44*(1), 79–94.

Eastwood, J., B. Snook and K. Luther. (2012). “What People Want From Their Professionals: Attitudes Toward Decision-making Strategies.” *Journal of Behavioral Decision Making*, *25*(5), 458–468.

Faraj, S., S. Pachidi and K. Sayegh. (2018). “Working and organizing in the age of the learning algorithm.” *Information and Organization*, *28*(1), 62–70.

Fogg, B. J. (2002). *Persuasive Technology: Using Computers to Change What We Think and Do*. *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco: Morgan Kaufmann.

Fuchs, C., T. Hess, C. Matt and C. Hoerndlein. (2016). “Human vs. Algorithmic recommendations in big data and the role of ambiguity." In: *AMCIS 2016: Surfing the IT Innovation Wave - 22nd Americas Conference on Information Systems*.

Goodyear, K., R. Parasuraman, S. Chernyak, E. de Visser, P. Madhavan, G. Deshpande and F. Krueger. (2017). "An fMRI and effective connectivity study investigating miss errors during advice utilization from human and machine agents." *Social Neuroscience*, *12*(5), 570–581.

Goodyear, K., R. Parasuraman, S. Chernyak, P. Madhavan, G. Deshpande and F. Krueger. (2016). "Advice Taking from Humans and Machines: An fMRI and Effective Connectivity Study." *Frontiers in Human Neuroscience*, *10*(542).

Gunaratne, J., L. Zalmanson and O. Nov. (2018). "The Persuasive Power of Algorithmic and Crowdsourced Advice." *Journal of Management Information Systems*, *35*(4), 1092–1120.

Highhouse, S. (2008). "Stubborn Reliance on Intuition and Subjectivity in Employee Selection." *Industrial and Organizational Psychology*, *1*(3), 333–342.

Hong, W., F. K. Y. Chan, J. Y. L. Thong, L. C. Chasalow and G. Dhillon. (2014). "A framework and guidelines for context-specific theorizing in information systems research." *Information Systems Research*, *25*(1), 111–136.

Jago, A. S. (2019). "Algorithms and Authenticity." *Academy of Management Discoveries*, *5*(1), 38–56.

Kim, H.-W. and A. Kankanhalli. (2009). "Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective." *MIS Quarterly*, *33*(3), 567–582.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig and S. Mullainathan. (2018). "Human decisions and machine predictions." *Quarterly Journal of Economics*, *133*(1), 237–293.

Komiak, S. Y. X. and I. Benbasat. (2006). "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents." *MIS Quarterly*, *30*(4), 941–960.

Kramer, M. F., J. Schaich Borg, V. Conitzer and W. Sinnott-Armstrong. (2018). "When Do People Want AI to Make Decisions?" In: *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 204–209).

Lapointe, L. and S. Rivard. (2005). "A Multilevel Model of Resistance to Information Technology Implementation." *MIS Quarterly*, *29*(3), 461–491.

Leyer, M. and S. Schneider. (2019). "Me, You or Ai? How Do We Feel About Delegation." *Proceedings of the 27th European Conference on Information Systems (ECIS)*.

Logg, J. M., J. A. Minson and D. A. Moore. (2019). "Algorithm appreciation: People prefer algorithmic to human judgment." *Organizational Behavior and Human Decision Processes*, *151*, 90–103.

Longoni, C., A. Bonezzi and C. K. Morewedge. (2019). "Resistance to Medical Artificial Intelligence." *Journal of Consumer Research*.

Luo, X., S. Tong, Z. Fang and Z. Qu. (2019). "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases." *Marketing Science*.

Madhaven, P. and D. A. Wiegmann. (2007). "Effects of Information Source, Pedigree, and Reliability on Operator Interaction With Decision Support Systems." *Human Factors*, *49*(5), 773–785.

Meehl, P. E. (1954). *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. *The American Journal of Psychology*. Minneapolis: University of Minnesota Press.

Nissen, M. E. and K. Segupta. (2006). "Incorporating Software Agents Into Supply Chains: Experimental Investigation with a Procurement Task." *MIS Quarterly*, *30*(1), 145–166.

Önkal, D., P. Goodwin, M. Thomson, S. Gönül and A. Pollock. (2009). "The relative influence of advice from human experts and statistical methods on forecast adjustments." *Journal of Behavioral Decision Making*, *22*(4), 390–409.

Palmeira, M. and G. Spassova. (2015). "Consumer reactions to professionals who use decision aids." *European Journal of Marketing*, *49*(3/4), 302–326.

Pezzo, M. V. and S. P. Pezzo. (2006). "Physician evaluation after medical errors: Does having a computer decision aid help or hurt in hindsight?" *Medical Decision Making*, *26*(1), 48–56.

Prahl, A. and L. Van Swol. (2017). "Understanding algorithm aversion: When is advice from automation discounted?" *Journal of Forecasting*, *36*(6), 691–702.

Promberger, M. and J. Baron. (2006). "Do patients trust computers?" *Journal of Behavioral Decision Making*, *19*(5), 455–468.

Rahwan, I., M. Cebrian, N. Obradovich, J. Bongard, J. F. Bonnefon, C. Breazeal, … M. Wellman. (2019). "Machine behaviour." *Nature*.

Shaffer, V. A., C. A. Probst, E. C. Merkle, H. R. Arkes and M. A. Medow. (2013). "Why do patients derogate physicians who use a computer-based diagnostic support system?" *Medical Decision Making*, *33*(1), 108–118.

Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, … D. Hassabis. (2017). "Mastering the game of Go without human knowledge." *Nature*, *550*, 354–359.

Sutherland, S. C., C. Harteveld and M. E. Young. (2016). "Effects of the Advisor and Environment on Requesting and Complying With Automated Advice." *ACM Transactions on Interactive Intelligent Systems*, *6*(4), 1–36.

Thurman, N., J. Moeller, N. Helberger and D. Trilling. (2019). "My Friends, Editors, Algorithms, and I: Examining audience attitudes to news selection." *Digital Journalism*, *7*(4), 447–469.

Trope, Y. and N. Liberman. (2010). "Construal-level theory of psychological distance." *Psychological Review*, *117*(2), 440–463.

Tversky, A. and D. Kahneman. (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science*, *185*(4175), 1124–1131.

Venkatesh, V., J. Y. L. Thong and X. Xu. (2012). "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology." *MIS Quarterly: Management Information Systems*, *36*(1), 157–178.

Wang, W. and I. Benbasat. (2013). "A Contingency Approach to Investigating the Effects of User - System Interaction Modes of Online Decision Aids." *Information Systems Research*, *24*(3), 861–876.

Webster, J. and R. Watson. (2002). "Analyzing the past to prepare for the Future: Writing a Literature Review." *MIS Quarterly*, *26*(2), xiii–xxiii.

Williams, A., I. Sherman, S. Smarr, B. Posadas and J. E. Gilbert. (2019). "Human trust factors in image analysis." *Advances in Intelligent Systems and Computing*, *778*, 3–12.

Wolf, J. R. (2014). "Do IT students prefer doctors who use IT?" *Computers in Human Behavior*, *35*, 287–294.

Xiao, B. and I. Benbasat. (2007). "E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact." *MIS Quarterly*, *31*(1), 137–209.

Yeomans, M., A. Shah, S. Mullainathan and J. Kleinberg. (2017). "Making sense of recommendations." *Journal of Behavioral Decision Making*, *32*(4), 403–414.