Received: 26 October 2021 | Revised: 5 November 2022 | Accepted: 12 November 2022

DOI: 10.1111/padm.12901

ORIGINAL ARTICLE

public administration WILEY

Algorithms in the public sector. Why context matters

Georg Wenzelburger[1] | Pascal D. König[2] | Julia Felfeli[3] | Anja Achtziger[3]

[1]Department of European Studies, Political Science, Saarland University, Saarbrücken, Germany

[2]Department of Social Sciences, Minda de Gunzburg Center for European Studies, TU Kaiserslautern, Harvard University, Cambridge, Massachusetts

[3]Department of Political and Social Sciences, Zeppelin University Friedrichshafen, Friedrichshafen, Germany

Correspondence
Pascal D. König, Department of Social Sciences, Harvard University, Minda de Gunzburg Center for European Studies, TU Kaiserslautern, Erwin-Schrödinger-Str. 57, 67663 Kaiserslautern, Germany.
Email: pascalkoenig@fas.harvard.edu

Funding information
Volkswagen Foundation, Grant/Award Number: 19-0087

Abstract

Algorithms increasingly govern people's lives, including through rapidly spreading applications in the public sector. This paper sheds light on acceptance of algorithms used by the public sector emphasizing that algorithms, as parts of socio-technical systems, are always embedded in a specific social context. We show that citizens' acceptance of an algorithm is strongly shaped by how they evaluate aspects of this context, namely the personal importance of the specific problems an algorithm is supposed to help address and their trust in the organizations deploying the algorithm. The objective performance of presented algorithms affects acceptance much less in comparison. These findings are based on an original dataset from a survey covering two real-world applications, predictive policing and skin cancer prediction, with a sample of 2661 respondents from a representative German online panel. The results have important implications for the conditions under which citizens will accept algorithms in the public sector.

Zusammenfassung

Algorithmen bestimmen zunehmend das Leben der Menschen, auch weil sie vermehrt im öffentlichen Sektor Verbreitung finden. Dieser Artikel untersucht die Akzeptanz von Algorithmen im öffentlichen Sektor. Er trägt dabei besonders dem Umstand Rechnung, dass Algorithmen als

Teil sozio-technischer Systeme immer in einen spezifischen sozialen Kontext eingebettet sind. Die Ergebnisse zeigen, dass die Akzeptanz eines Algorithmus stark davon abhängt, wie Personen bestimmte Aspekte dieses Kontexts bewerten. So gehen eine höhere subjektive Wichtigkeit des Problems, welches ein Algorithmus adressiert, sowie ein höheres Vertrauen in die Organisation, die den Algorithmus einsetzt, mit höherer Akzeptanz einher. Dagegen beeinflusst die objektive Leistung eines präsentierten Algorithmus die Akzeptanz viel weniger. Diese Ergebnisse beruhen auf neuen Daten aus einer Umfrage zu zwei realen Anwendungen, der vorhersagenden Polizeiarbeit und der Hautkrebsvorhersage. Die Stichprobe bildeten 2661 Befragte aus einem repräsentativen deutschen Online-Panel. Die Ergebnisse haben wichtige Konsequenzen für die Bedingungen, unter denen Bürger Algorithmen im öffentlichen Sektor akzeptieren.

1 | INTRODUCTION

Today's information societies increasingly delegate decisions in various life areas to algorithmic decision-making systems. These systems—abbreviated as algorithms in the following—can build decision-models from data that, in turn, serve to deal with a task that would otherwise require human cognitive capabilities. Citizens already interact in various ways with algorithms, often without being aware of this. As consumers, citizens are for instance taking advantage of algorithm-based services when relying on software that recommends movies or TV series based on customers' previous choices. More importantly, however, people are also affected by algorithms in their role as citizens because governments and public services increasingly rely on such systems when they make decisions (Busuioc, 2021). Indeed, countries such as the United States (Coglianese & Ben Dor, 2020), New Zealand (New Zealand Government, 2018), and European Union member states (Chiusi et al., 2020) are rapidly adopting algorithms in the public sector. These assist in policing activities, in matching qualification measures to the unemployed, or in identifying the most efficient school bus routes, among other things.

The use of algorithms not only entails a partial reorganization of bureaucracies and individual agencies (Meijer et al., 2021) as well as their relation to citizens (Miller et al., 2022; Prokop & Tepe, 2022), but also changes the bases of decision-making in a context that demands democratic accountability and legitimacy (Busuioc, 2021; Levy et al., 2021; Veale & Brass, 2019). Clearly, the use of algorithms by state authorities is quite different from commercial applications such as movie recommendation systems at least for two reasons. First, citizens often cannot "opt out" from being treated by an algorithm and, second, state decisions often intervene substantially in society and people's lives (Krafft et al., 2022). The potential far-reaching impact of algorithms is one of the reasons why research on how to ensure fair, accountable, and transparent (FAccT) algorithms has proliferated in recent years (e.g., Kroll et al., 2017; Lepri et al., 2018; Mittelstadt et al., 2016; Wachter et al., 2017). The importance of these features is also mirrored in research dealing with how laypersons perceive and evaluate algorithmic systems. This work is commonly interested in the role of an algorithm's performance, including with regard to its fairness, as well as its transparency and accountability (e.g., Binns et al., 2018; Grgic-Hlaca et al., 2018; Juravle et al., 2020; Langer et al., 2018;

Shin, 2020, 2021). This focus on algorithm characteristics is also visible in studies that look more specifically at algorithmic systems used in the public sector, such as for detecting welfare fraud or predicting criminal behavior (Grimmelikhuijsen, 2022; Kennedy et al., 2022; Schiff et al., 2022).

Much less attention has been directed at the question of how the context (i.e., a certain life area) in which an algorithm is used affects people's acceptance of algorithms. The research on algorithm perceptions thus mirrors what Brown et al. (2021, p. 2) have diagnosed in the literature on the evaluation of ethical aspects of algorithms, namely that "[t]he context of the algorithm is one of the most overlooked elements." This is an important gap, because while exhibiting certain technical features, an algorithm is also always bound up with its social context: it is necessarily deployed by certain actors, applied in a concrete domain, and realizes a specific purpose (e.g., optimizing routes for emptying trash cans or predicting crime). And citizen's perceptions of algorithms might already depend notably on their attitudes toward these contextual characteristics, the domain of application, the realized objective, and who uses the algorithms.

Importantly, this social embeddedness of the algorithm may be much more important for the acceptance of algorithms than the technical features of the systems themselves. It is in the concrete area of application where the algorithm has direct repercussions on life chances. At the same time, citizens will hardly be familiar or even become aware of an algorithm's technical characteristics. How citizens view familiar aspects of the larger social context in which an algorithm is deployed may thus drive the evaluation of algorithms much more strongly than technical features. This argument does not imply that studying people's evaluation of certain technical aspects of an algorithm (e.g., their effectiveness or fairness in prediction errors) is not important. However, merely concentrating on technical features of algorithms may miss an important facet that is much closer to how people actually develop perceptions about algorithms in their daily lives.

To shed light on this under-investigated question of how the social context in which an algorithm is applied affects its public acceptance, we explore to what extent support for these systems depends on how people evaluate two context-related factors. We look at (1) how important the purposes realized by an algorithm (e.g., security) are for an individual's acceptance of it and (2) how a person's attitude toward the organizations which deploy the system in a given context affects the evaluation of an algorithm. To test these attitudes, we choose two prominent use cases from contexts in which fundamental values are at stake for most people: policing (with algorithms used for in predictive policing), and health care (with algorithms for predicting skin cancer). We also examine the performance of the algorithm as a central technical feature and analyze how much it matters as compared to context-related characteristics and personal dispositions. This analysis is based on a representative survey with data from more than 2600 respondents from an online panel that reflects the composition of the German population.

The paper is structured as follows. The following, Section 2 presents the theoretical assumptions and hypotheses. Section 3 covers the methods of data collection and analyses, followed by the presentation of the findings in Section 4 and a discussion and summary of the main findings in Section 5.

2 | THEORETICAL ASSUMPTIONS AND HYPOTHESES

2.1 | The state of the art: Technical features, personal dispositions, and algorithmic literacy

A highly influential theoretical framework used to study of how people perceive and evaluate novel technologies is the technology acceptance model or TAM, for short (Davis, 1989). This model initially held that two main factors drive technology acceptance, the perceived ease of use and the perceived usefulness. Over time, though, the TAM has been complemented by other individual-level variables that seemed to also influence technology acceptance (King & He, 2006; Marangunić & Granić, 2015). These factors relate to individual experience with and confidence in technology, to specificities of the technology as perceived risk or trust, but also to individual characteristics such as gender or cultural background. The TAM and its focus on how users experience the technology itself have also informed research on algorithm acceptance (Shin et al., 2020). A quickly growing literature shows that features of algorithmic systems, such as performance-related aspects or transparency affect how much people trust these systems and support their adoption (for an overview, see Glikson & Woolley, 2020).

This strong focus on design features is understandable in light of a broad literature interested particularly in dealing with the challenges of safeguarding the FAccT of algorithmic systems (Lepri et al., 2018). However, it has also been shown that personal dispositions, such as expertise in the task that an algorithm performs and algorithmic literacy, matter for the acceptance of algorithms (Logg et al., 2019). Evidence also clearly indicates that the context in which an algorithm is adopted is important. Notably, it is a recurring finding that trust in algorithmic systems is higher in areas involving technical tasks as opposed to tasks that involve human judgment (Araujo et al., 2020; Juravle et al., 2020; Lee, 2018; Logg et al., 2019; Schepman & Rodway, 2020).

Research on the evaluation of algorithms, so far, deals mainly with commercial applications in various areas like medicine or recruiting, though. Studies on applications specifically in the public sector are still scarce and they, too, show a focus on aspects of user experience and features of the algorithmic systems themselves. Similar to research on algorithms adopted outside the public sector, studies on public sector applications have shown that biases and lack of transparency (Grimmelikhuijsen, 2022; Schiff et al., 2022; Waggoner et al., 2019) as well as the complete automation of the system (Kennedy et al., 2022) reduce trust. Certain groups in society may also show a stronger appreciation of a given feature: Citizens with discrimination experience prefer algorithmic decision-making out of an expected higher fairness (Miller & Keiser, 2021).

Looking beyond the importance of features of the applications themselves, a few contributions have also covered different domains of application. While support for algorithms can be similar in even highly different settings, such as child welfare and criminal justice (Schiff et al., 2022), existing evidence does indicate that, just like with commercial applications, citizens support algorithms more if these inform or carry out technical and routine tasks as opposed to tasks that involve human judgment (Aoki, 2020; Chatterjee et al., 2022; Ingrams et al., 2022; Miller & Keiser, 2021; Starke & Lünich, 2020). At the same time, little is known about the role of attitudes toward aspects of the context in which an algorithm is employed. It has been shown specifically for self-service technologies in the public sector that trust in government leads to a more positive experience with the applications (Chen et al., 2021). In a similar vein, existing evidence suggests that the reputation of the developer, which too forms a part of the socio-technical system in which an algorithm is embedded, affects algorithm acceptance (Kennedy et al., 2022).

Overall, the user experience perspective of the TAM and a focus on characteristics of the technology itself also loom large in research on the perception of algorithms in the public sector. This focus is well aligned with where the public administration literature has located possible benefits and challenges of algorithm used. Their scalability, cost-effectiveness, and above-human performance in some areas are all rooted in technical properties and together make "artificial discretion" a possible means to improve decision-making (Young et al., 2019). Equally, the technology as such introduces new information asymmetries and problems of transparency that can compound accountability challenges in the public sector. These may easily conceal the value-laden choices and trade-offs inherent to the design of an algorithmic system that determine its performance and built-in biases (Busuioc, 2021; König & Wenzelburger, 2021; Veale & Brass, 2019; Vogl et al., 2020). In this sense, technical aspects have direct consequences for questions of bureaucratic legitimacy, and the centrality of algorithm design choices means shifting discretion from frontline decision-makers to those developing and designing an algorithmic system (Levy et al., 2021). Since their predominant concern may often be with the efficiency that delegation to an algorithm promises to realize, there is a risk that algorithm adoption will harm important public values, such as transparency and fairness, that bureaucracy is supposed to realize (Schiff et al., 2022).

However, while properties of algorithmic systems themselves can be presumed to be of special importance for their acceptance where people directly interact with the algorithmic systems, this is much less the case for applications that people do not use themselves. Many algorithms used in the public sector fall into this second category. It is public sector employees who are in direct contact with these systems, whereas citizens have no direct experience from interactions with them and can therefore hardly evaluate the technology based on ease of use, for instance. As the next section argues, in situations where a direct contact does not exist, algorithmic acceptance will be strongly driven by evaluations that concern not so much the algorithms and their design, but aspects of the context in which they are adopted.

2.2 | On the need to consider dispositions toward the context of algorithms

Technical design features of an algorithmic system can be expected to influence to what extent people endorse the adoption of these systems. However, from a perspective that sees this technology as a part of a larger socio-technical system (e.g., Ananny & Crawford, 2018; Johnson & Verdicchio, 2017; Meijer, 2018; Shin & Park, 2019), an algorithm's evaluation cannot be detached from the concrete social context in which they are to be used. Clearly, one could examine a range of features that are tied to this social context. In the following, we focus on two factors that can be presumed to be particularly important.

First, one should note that essentially the same technology in the form of an algorithmic system could be used in different domains and serve widely different purposes. For instance, an image recognition system could be trained and used to spot suspicious behavior in a crowd or to help a deaf person with lip reading. Therefore, the expected value offered by an algorithm and the risks it entails will depend on the specific purpose it is supposed to fulfill in a given social context. This notion harks back to the original TAM and the variable of "perceived usefulness," which is one of the key predictors of technology acceptance among consumers. However, in our understanding, the perceived use of an algorithm should primarily depend on how important an individual deems the specific life area, in which the algorithm will be applied. More concretely, if a person places a high value on security, for example, we would expect her to be, ceteris paribus, more supportive of predictive policing algorithms. Moreover, the extent to which people are personally affected by the problem that an algorithm is supposed to address should also be relevant—and this perception may be based on a cognitive evaluation (e.g., probability of crime) of the situation or on a more affective component (e.g., fear of crime). We therefore formulate the following two hypotheses:

> **HYPOTHESIS 1A.** The greater the personal importance of the problem that the algorithm's application is supposed to deal with, the more a person supports the use of an algorithm in the public sector.

> **HYPOTHESIS 1B.** The more a person is subjectively affected by the problem that the algorithm is supposed to deal with, the more a person supports the use of an algorithm in the public sector.

Second, it is important to consider that the algorithm will not be used by citizens themselves but deployed by a public authority in a given context. Hence, people's acceptance of the algorithm may not so much be driven by the technology, but rather by their relation to the organization running the algorithm. In the management literature, research on "organizational reputation" has shown that reputation is key for firms to generate trust with consumers and, eventually, the decision to purchase a product (Kim et al., 2008, p. 551). As reputation has also been shown to be important for public organizations (Wæraas & Ma'or, 2015), it is reasonable to expect that citizens will be more favorable to accept the use of algorithms if they are implemented by public authorities that are looked upon positively and are seen as trustworthy—as has been shown in other life areas (Kehr et al., 2015; Kim et al., 2008). More concretely: People who think highly about the police may therefore show support for the use of predictive policing algorithms merely because they have a favorable orientation toward the responsible organization (i.e., the police or the state in general). We therefore test the following hypothesis:

> **HYPOTHESIS 2.** The more favorable a person's attitude toward the organization operating the algorithm is, the more this person supports the use of the algorithm in the public sector.

Figure 1 summarizes the expected relationships between people's dispositions and their evaluation of an algorithm. It illustrates our three main hypotheses positing that individuals' dispositions toward an organizational (responsible actor operating the system) and a functional (realized purpose) context-specific dimension will affect their acceptance of algorithms—while controlling for the key variables discussed in the literature about algorithmic acceptance. We also include algorithm performance as a central feature of an algorithm. While this feature is of interest in itself, it also useful for a comparison with the influence of the context-specific evaluations. Since algorithm performance or effectiveness is not an individual-level factor but an objective characteristic of an algorithm, we probe the influence of algorithm effectiveness on people's algorithm support with an experimental treatment in our survey, randomizing respondents into two conditions with differing algorithm performance levels. We describe these conditions together with data collection and measure in the next section.

3 | DATA, MEASURES, AND METHOD

3.1 | Survey design and sample

To study our theoretical claims, we set up an online survey covering two real-world applications of algorithms used in the public sector. The first application was in the context of predictive policing. It included information about an algorithm used to predict the risk of burglary based on past crime patterns. The respondents were informed that in case of high risk, the area would be monitored more closely by the police to prevent burglaries. The second application was skin cancer risk prediction. Here, we presented an algorithm developed with health data to assess future skin cancer risk. In case of high risk diagnosed by the algorithm, a patient would be selected for intensified screening for skin cancer. These two contexts have been chosen for two reasons. First, they are comparatively prominent use cases and they also both concern fundamental values, namely security and health. As we want to probe how much the problem that an algorithm addresses—and thus the value that it realizes—matters for algorithm support, it is important to have applications involving generally relevant values. Otherwise, if the values in play are marginally

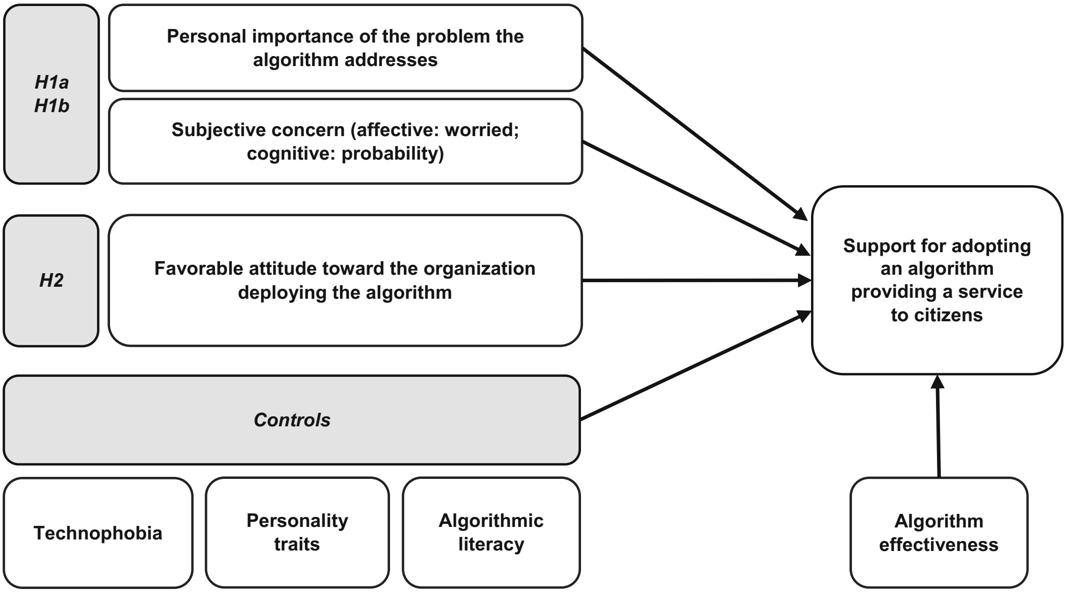

FIGURE 1 Model of the analyzed relationships

relevant to most people, this would make the use cases less relevant for our argument and analysis. Second, the two contexts are comparable in the consequences of their predictions: Both algorithms' recommendations potentially result in intensified monitoring—more policing and more health screening, respectively—due to risk classifications. The survey questions have been formulated to keep the two studied contexts maximally parallel and enable comparability (see Annex A1 and A2 for the survey questions).

The survey was set up online and 2661 respondents were recruited and paid via the professional polling firm *respondi* AG using an online panel representative of the German population from 18 to 74 years. In order to have a valid comparison of contexts, respondents were randomly assigned to either the predictive policing or the skin cancer context. Having over 1000 respondents per context, is suitable for detecting the posited relationships (for details on the sample composition, see Annex A6).[1] To reduce noise in the data, we used a speeding check as well as an attention check and a control question at the end of the survey.[2] Answers from respondents who responded overly quickly or took particularly long (beyond two standard deviations above and below the mean) were excluded from the analysis.[3]

It is important to note the participants were asked to evaluate something with which they are presumably not familiar. Algorithms are not something that most people are confronted with in their daily lives and our questions partly involve some technical aspects. This introduces a certain trade-off when gauging citizens' acceptance of algorithmic systems. On the one hand, we are interested in how the general population perceives algorithms and evaluates them as laypersons. On the other hand, a minimal understanding of and familiarity with what the survey is about when responding to the questions seems desirable. We have therefore included introductory material at the beginning of the survey with a basic description of the algorithms used in the two domains.[4] Participants furthermore saw a series of exemplary algorithms, which varied along dimensions such as performance, cost, and transparency, and which they were asked to compare and rate. They thus also had to actively engage with these materials.

As we did not want respondents to have one specific configuration in mind, they saw various different designs before asking them about their general acceptance of algorithms in the two studied domains (for details on the introductory materials, see Annex A3–A5). These materials served the purpose of familiarizing participants with algorithms. Yet, seeing algorithms as a way of dealing with given problems may already lead to a certain bias toward more favorable perceptions of the technology. The materials are designed to counter such a bias with a balanced account of algorithms in the introductory texts and through the presentation of different algorithm designs with partly desirable, partly undesirable features. We can, however, not exclude the possibility that citizens, when asked to provide entirely spontaneous evaluations about algorithms, would show on average a different overall evaluation of algorithms. We also acknowledge while participants did indeed spend much time getting to know the exemplary settings of different algorithms, a substantial part of the participants did not spend enough time on the introductory text and therefore did probably not fully process the initial information. As this goal of the survey has therefore not entirely been reached, we performed additional tests with a subgroup of our sample where only those respondents were included that took more time than the median to read the introductory material.[5]

3.2 | Measures

3.2.1 | Dependent variable

Our dependent variable reflects how much respondents support the adoption of the studied algorithmic systems based on four items. These items are substantively identical for the predictive policing and for the skin cancer prediction setting. They only differ in wording to adapt them to the specific context. The items cover both a more general support of algorithms and a more specific support in the respective domain and also include two negative items (see Annex A1). The items are all scaled from -2 to $+2$ and capture the extent to which respondents agree with the adoption of such algorithmic systems in policing (Cronbach's $\alpha = 0.86$) and medicine (Cronbach's $\alpha = 0.80$), respectively, or instead prefer to leave decision-making entirely to humans. These two dependent variables have been rescaled to the range of 0–1 before using them in the analysis.

3.2.2 | Independent variables

To measure the *importance of the problem* that an algorithm is designed to deal with, we draw on the Schwartz values inventory (Schwartz, 2012) and the six items that measure peoples' importance of the value of security (Cronbach's $\alpha = 0.88$). As the Schwartz values items for security also include one item that asks about how much people try to avoid sickness, we have also asked this item in the medical context (i.e., skin cancer prediction). For this medical context, we furthermore draw on a validated scale for the importance of health as a value (Lau et al., 1986). Cronbach's α of this scale (0.55) turns out to be the lowest of all measures we used in our study. As removing any of the items did not increase Cronbach's alpha, we have added the Schwartz value item mentioned above, thereby increasing α to 0.64. We therefore use this amended scale, but also perform additional analyses with alternative model specifications as robustness checks. Given the registered reliability score for this scale, our findings below are likely to underestimate the importance of this factor.

To capture subjective concern, we use measures tapping into the (I) affective and (II) cognitive dimensions behind concern for security and health risks. Concerns about insecurity are measured by several questions from the German victimization survey (Birkel et al., 2018) and include (I) the subjective worry about burglary as well as (II) two measures (combined into one variable) about the expected likelihood of burglaries in one's home and neighborhood. In the medical context, we asked questions on how much people are (I) worried about getting skin cancer in the next 5 years and asked them about what they think that (II) the probability of this diagnosis is for them, for members of their household, and for friends and relatives within the next 5 years (these three items on expected probabilities are again combined into one variable).

For measuring the *attitudes toward the organization operating an algorithmic system*, we used several items measuring trust in relevant institutions using common survey items in sociological and political science surveys. We have selected items that measure trust in parliament, the judicial system, the police, and the public administration in the policing context (Cronbach's $\alpha = 0.90$) and adapted a second set of items to the medical context: These items measure trust in the health ministry, the health system, doctors, and hospitals (Cronbach's $\alpha = 0.87$).

3.2.3 | Controls

To account for possible confounding variables, we introduced control variables based on the discussion of the literature further above. To measure demand for transparency in public administration, we included a variable that measures how much transparency and access a person wants in the management of government affairs and administration. Unlike the trust variable (H2), it is not a direct evaluation of actors using an algorithm, but about how citizens would ideally like public sector institutions to operate. Including this kind of expectation as a control variable seems important, as the introduction of algorithmic systems may be felt as increasing general opacity of decision-making and because transparency has been found to be significantly related to trust in algorithms in existing studies (see above). The measure for demand for transparency relies on the principled transparency scale used by Piotrowski and Van Ryzin (2007), which consists of four items scaled from -2 to $+2$ (Cronbach's $\alpha = 0.85$ and 0.87 in the policing and the medical setting, respectively).

To measure personal values, we used a validated short scale of the Five Factor scale of personal values, which includes two items to measure each of the following variables: agreeableness, conscientiousness, openness, extroversion, and neuroticism. We also measured participants' technophobic versus technophile orientation with a technophobia scale used by Sinkovics et al. (2002) and Nimrod (2018). As this scale is rather long, we included only items that belong to the dimension "human versus machine ambiguity," which comprises 10 items (Sinkovics et al., 2002, p. 486). Cronbach's α of the technophobia scale amounts to 0.89 in both conditions.

To capture respondents' knowledge about algorithms, we asked them (I) to state their self-assessed knowledge using a slider running from 0 (low knowledge) to 100 (high knowledge). We also asked respondents (II) to assess their understanding of computers and specific applications of algorithms by using a scale used by Cheng et al. (2019). The four items of this scale showed Cronbach's α's of 0.75 and 0.77 for the policing and the medical context, respectively. Both knowledge variables were scaled to the range from 0 to 1 and combined into a single variable measuring literacy of algorithmic systems.[6] While we draw on a self-assessment of algorithmic literacy—also for reasons of keeping the cognitive burden low in an overall longer survey[7]—this measure may show reduced reliability, especially for scores in its mid-range as many people might overestimate their algorithmic literacy. As an alternative measure, we have therefore also created a dichotomized variable that distinguishes respondents scoring among the highest 20% (coded 1) on the metric algorithmic literacy variable from all others (coded 0).[8] Finally, we have included variables for age, education (coded as 1 vs. 0, if a person has attained upper secondary tier education), and gender (1 = female, 0 = male and diverse). Like the dependent variables and all metric independent variables have been standardized to a range from 0 to 1 (taking the empirical maximum for age) for an easier interpretation of the results (for an overview, see Table 1).

Finally, we include a measure for the performance of the algorithm. However, since the performance of an algorithm is a property of the algorithm itself and can therefore not be asked via a survey question, we assessed the effect of algorithm performance on algorithm acceptance by means of a treatment introduced in the survey (as the only systematic difference in the algorithm design examples presented to the respondents). To do so, we randomly assigned respondents to a low- or a high-performance treatment in both the skin cancer and the predictive policing context using the true positive rate as a simple measure of algorithmic performance that was also explained to respondents in the introduction part of the survey (e.g., how many of all burglaries that took place have been predicted by the algorithm). One group saw examples of algorithm configurations with true positive rates varying between 5%, 10%, and 15% (low-effectiveness condition), for a second group they varied between 85%, 90%, and 95% (high-effectiveness condition).[9] Hence, with this treatment, we suggested to respondents that the achieved true positive rate of the algorithms was either very low or very high.[10]

Importantly, while this randomization means that respondents simply saw either a badly or a well-performing algorithm, we did not prompt them to further reflect on the performance of the algorithm they saw. This setting mirrors a natural evaluation situation in terms of how people may come to perceive algorithms in their daily lives—where they will hardly be confronted with different algorithm designs varying in their performance from which one could choose. Yet, given the very strong effectiveness contrast, we would expect that the treatment clearly affects support for the use of algorithms. It is important to note that the variable for algorithm performance via the treatment is the only experimental part of the survey. Including this variable also provides us with a benchmark to which we can compare the non-experimental context-related variables. All these other variables of main theoretical interest are measured post-treatment as part of the survey. It should be noted that only the experimental treatment lends itself to a causal interpretation, the statistical effects found for the other variables are of an associational nature and may thus pick up other underlying causal influences or partly go through other included variables.

4 | RESULTS

Before turning to the results from the regression analyses, Figure 2 illustrates the two dependent variables—the general acceptance of an algorithm in predictive policing and in skin cancer prediction. As the survey contained an experimental treatment introducing half of the respondents to an algorithm with low performance and the other half to an algorithm with much higher performance, we report the distributions split by the treatment. The description of the dependent variables thus also allows for inspecting the role of the performance treatment which will serve for a comparison with the main, non-experimental independent variables examined further below.

TABLE 1 Descriptive statistics

Construct	Predictive policing # items	Predictive policing Cronbach's alpha	Predictive policing M	Predictive policing SD	Skin cancer prediction # items	Skin cancer prediction Cronbach's alpha	Skin cancer prediction M	Skin cancer prediction SD
Dependent variables								
Support for algorithmic system in domain	4	0.86	0.62	0.24	4	0.80	0.66	0.23
Main predictors								
Personal importance of security/health (H1A)	6	0.88	0.72	0.19	5	0.64	0.75	0.16
Worried about burglary/sickness (H1B)	1	–	0.22	0.41	1	–	0.20	0.40
Perceived probability of burglary own home/own sickness (H1B)	1	–	0.12	0.32	1	–	0.16	0.36
Perceived probability of burglary neighborhood/sickness household (H1B)	1	–	0.38	0.49	1	–	0.22	0.42
Index perceived probability break-in/ Index perceived probability sickness (H1B)	2	0.79	0.40	0.23	3	0.86	0.39	0.23
Trust in government/healthcare institutions (H2)	4	0.90	0.60	0.23	4	0.87	0.66	0.20
Control variables								
Demand for transparency	4	0.85	0.71	0.22	4	0.83	0.75	0.20
Technophobia	10	0.89	0.36	0.15	10	0.89	0.33	0.15
Self-reported understanding of algorithmic systems	4	0.75	0.69	0.20	4	0.77	0.68	0.21
Self-reported knowledge about algorithms	1	–	0.39	0.27	1	–	0.31	0.25
Index algorithmic literacy	5	0.76	0.55	0.20	5	0.75	0.51	0.19
Big five: Agreeableness	2	–	0.46	0.19	2	–	0.46	0.20
Big five: Conscientiousness	2	–	0.31	0.20	2	–	0.31	0.20
Big five: Extroversion	2	–	0.51	0.25	2	–	0.50	0.24
Big five: Openness	2	–	0.60	0.24	2	–	0.60	0.25
Big five: Neuroticism	2	–	0.40	0.23	2	–	0.43	0.25
Female	1	–	0.51	0.50	1	–	0.53	0.50
Age in years	1	–	0.45	0.26	1	–	0.46	0.26
Upper secondary tier education	1	–	0.32	0.47	1	–	0.32	0.47

Note: Performance of an algorithm is included in the analysis as an experimental treatment. The dependent variables and all independent variables are normed to a range from 0 to 1. M = mean, SD = standard deviation.

Several points are noteworthy: First, in both contexts, the support for implementing algorithms is similarly skewed toward the acceptance of predictive policing (mean = 0.62, variable scaled from 0 to 1) and skin cancer (mean = 0.66, variable scaled from 0 to 1). Nevertheless, a *t*-test between the two groups reveals that the difference between the means of both groups, with higher acceptance rates for the healthcare context, is significant at a 99% level. Second, the graphs also illustrate the considerable variance in the data. Quite some people are highly supportive of the studied algorithms (almost more than a third with values of at least 0.75) whereas others, roughly 20%, fall below the mid-point of the scale (0.5). Third, this variation is not primarily due to the different information about the algorithms' performance that we provided as an experimental treatment (i.e., 5%, 10%, or 15% true positive rate vs. 85%, 90%, or 95% true positive rate). *T*-tests reveal significant but substantively small differences (difference $= 0.03$, $t = -1.89$, $p = 0.06$, in the case of predictive policing; difference $= 0.02$, $t = -1.76$, $p = 0.08$, in the case of skin cancer prediction). The substantively small differences are remarkable in view of the presented information about the algorithms' performance differing massively between the experimental treatments. Nonetheless, the data clearly indicate that whether respondents saw an algorithm with very high or instead with rather low performance in terms of its true positive rate only explains a very small part of the variation of the dependent variable.

Whereas the distribution of algorithmic support for different subsamples of our respondents allows a first insight into the relevance of performance for algorithmic acceptance, our variables of primary theoretical interest (as well as the control variables) have been measured as survey items identically for all subgroups. Therefore, to test the formulated hypotheses about the role of context-related evaluations, we use OLS regression models to estimate the strength of associations between the predictors and our two dependent variables, the algorithmic acceptance in the predictive policing or the skin cancer context.[11] The results are presented in Table 2 (unstandardized coefficient estimates, all variables scaled from 0 to 1). We report these findings following the structure of the table.

4.1 | Main predictors (H1A, H1B, and H2)

From our main predictors, the variable for personal importance is strongly and significantly correlated with the acceptance of algorithms for both cases, predictive policing and skin cancer prediction. Hence, in line with

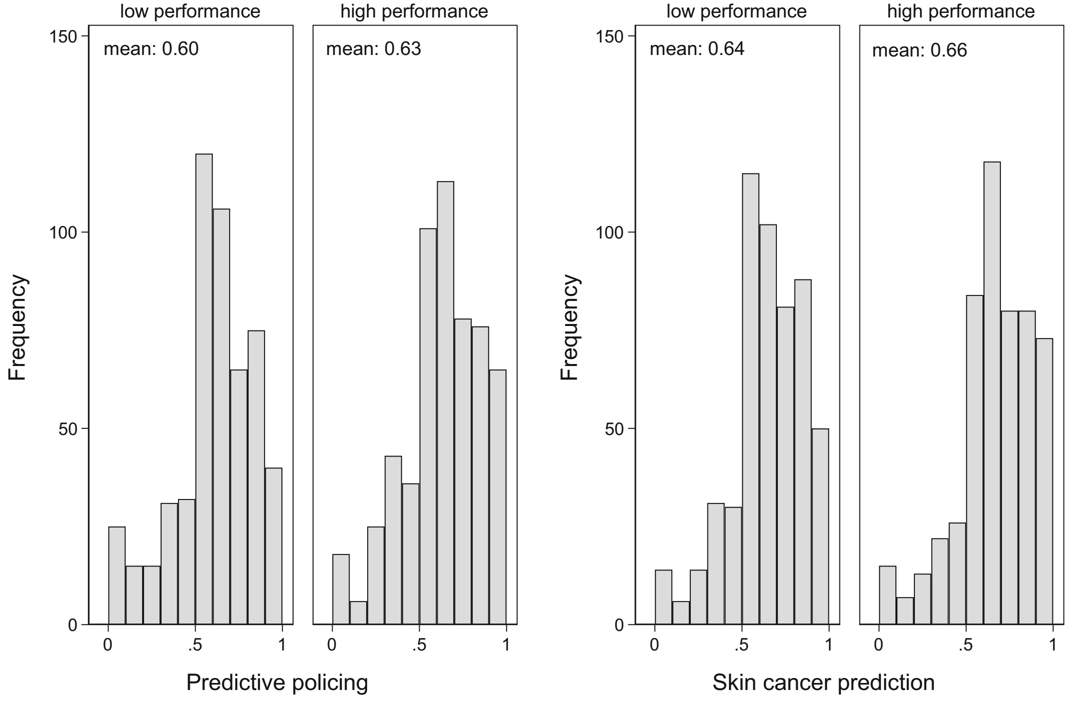

FIGURE 2 Distribution of the scores of algorithm support. The left- and right-hand figures for each condition (predictive policing and skin cancer prediction) represent different received experimental conditions. Respondents were randomly assigned to the condition of a low true positive rate of an algorithm ("low performance," configurations varied between 5%, 10%, and 15%) or a high true positive rate ("high performance," presented configurations varied between 85%, 90%, and 95%).

Hypothesis 1A, the importance a person gives to security (predictive policing) or health (skin cancer prediction) shows a highly significant and positive association with acceptance of algorithms.[12] As expected in the theory section, it seems indeed that citizens already show a more favorable stance toward adopting algorithms in the public sector if they care a lot about the value that is at stake.

The results for subjective concerns (H1B) are less clear-cut. For skin cancer prediction, affective worries about getting skin cancer are strongly correlated to algorithmic acceptance, whereas the cognitive indicator—the perceived risk—is not (although the coefficient points in a similar direction). For predictive policing, in contrast, the affective part seems to be somewhat less important than the cognitive part: While worries about burglary are not significantly related to algorithmic acceptance in all categories, the indicator tapping into the perceived risk is significant.[13] Nevertheless, as the coefficients for all variables subsumed under H1B point in the same direction, we can at least conclude that there is overall support for this hypothesis—with slight differences concerning the relevance of cognitive or affective aspects depending on the actual case presented to the respondents. These associations are, however, clearly less pronounced than for the personal importance of the realized values of security and health in predictive policing and skin cancer prediction, respectively.

We posited in H2 that respondents' attitudes toward the institutions responsible for deploying the algorithms in the respective context are correlated with the degree to which they support the use of these algorithms. The results from the regression analyses shown in Table 2 lend strong support to this hypothesis. Indeed, institutional trust is a highly significant and positive predictor of algorithm acceptance in both examined contexts: Respondents who trust government institutions or healthcare institutions are clearly more in favor of implementing an algorithm in the respective setting.[14]

Altogether, the associations reported in Table 2 therefore lend robust support to our hypotheses that the specific social context in which an algorithmic system is embedded is related to the acceptance of the algorithm. We can further compare the effect sizes to the experimental treatment with differing algorithmic performance. This dichotomous variable (taking the value of 1 for those respondents who saw the better performing algorithm, with mean and SD at 0.5 for both cases) can serve as a benchmark as it refers to a central technical feature of algorithms and a relevant predictor of technological acceptance according to the literature. However, the effect of this treatment, that is, of the provided information on an algorithm's performance, emerges as barely influential in Table 2. Corroborating the patterns of the *t*-tests above, the experimental contrast of the presented performance of an algorithm (i.e., a true positive rate of about 10% vs. about 90%) only slightly affects respondents' support of algorithms: Seeing a better performing algorithm leads to higher acceptance, but only by 0.03 units (on a dependent variable ranging from 0 to 1). In contrast, changing personal importance from 0 to 1 increases the predicted acceptance by roughly 0.2 (depending on the model).

To get a visual impression of the substantive size of the most important predictors in the regression models, that is, personal importance of the context in which the algorithm will be applied (H1A), and trust in the institutions deploying the algorithm (H2), Figure 3 visualizes the predictive margins for these variables using the same scale for the linear prediction. Moreover, we include technophobia as one of the major control variables that proves to be highly significant (see below). The results are clear-cut: While technophobia clearly shows the strongest association with algorithm acceptance, acceptance also considerably increases, ceteris paribus, when increasing the two context-dependent factors personal importance of the realized purpose and trust in the institutions deploying the algorithm. When looking at the standardized coefficients (not tabled), increasing technophobia by one standard deviation decreases the dependent variable by about one-third of a standard deviation. The standardized coefficients of the personal importance of security/health and institutional trust are about half that size.

4.2 | Control variables

In contrast to the clear associations found for the variables that are of main interest, most of the control variables are irrelevant, with a few major exceptions. First, the extent to which people demand transparency from governmental institutions is related to algorithm acceptance in one of the two studied contexts. Whereas for predictive policing,

TABLE 2 Coefficient estimates from OLS regression predicting support for algorithm use

	Predictive policing		Skin cancer prediction	
	Model 1	Model 2	Model 3	Model 4
Main predictors				
Personal importance of addressed problem (H1A)	0.17**	0.19**	0.17**	0.18**
	(4.35)	(4.55)	(3.80)	(3.63)
Worried burglary/skin cancer (H1B)				
Not at all	Reference category		Reference category	
A bit worried	0.04*		0.04*	
	(2.35)		(2.32)	
Quite worried	0.04		0.07**	
	(1.91)		(3.32)	
Very worried	0.05		0.09**	
	(1.72)		(2.86)	
Probability of a burglary/skin cancer (H1B)		0.08*		0.07
		(2.23)		(1.74)
Trust in institutions in charge (H2)	0.11**	0.12**	0.15**	0.17**
	(3.34)	(3.51)	(3.99)	(3.95)
Experimental treatment variable: algorithm performance				
High versus low performance of algorithm	0.03*	0.03*	0.03*	0.02
	(2.29)	(2.00)	(2.07)	(1.46)
Control variables				
Demand for transparency of institutions	−0.15**	−0.15**	0.00	−0.00
	(−5.18)	(−4.77)	(0.06)	(−0.13)
Technophobia	−0.61**	−0.59**	−0.40**	−0.45**
	(−13.55)	(−12.12)	(−8.76)	(−8.73)
Algorithmic literacy (self-assessed)	0.037	0.02	0.05	0.09*
	(1.03)	(0.47)	(1.22)	(2.03)
Extraversion	0.04	0.05	0.03	0.03
	(1.44)	(1.62)	(0.81)	(0.96)
Agreeableness	0.05	0.05	−0.02	−0.02
	(1.47)	(1.27)	(−0.51)	(−0.50)
Conscientiousness	−0.02	−0.04	0.07	0.07
	(−0.51)	(−0.92)	(1.95)	(1.62)
Openness	−0.07*	−0.07*	0.04	0.03
	(−2.33)	(−2.23)	(1.27)	(0.92)
Neuroticism	0.07*	0.08*	0.00	0.03
	(2.30)	(2.41)	(0.03)	(0.84)
Female	0.00	0.00	−0.02	0.00
	(0.36)	(0.65)	(−1.02)	(0.01)
Age	0.01	0.02	−0.09**	−0.07*
	(0.69)	(0.68)	(−3.16)	(−2.03)

TABLE 2 (Continued)

	Predictive policing		Skin cancer prediction	
Education	−0.03*	−0.03*	−0.00	−0.01
	(−2.48)	(−2.09)	(−0.44)	(−0.93)
Constant	0.67**	0.65**	0.50**	0.47**
	(10.80)	(9.99)	(7.55)	(6.16)
N	1070	971	1031	761
R^2	0.290	0.284	0.168	0.198

Note: Unstandardized coefficients, *t* statistics in parentheses. The dependent variables are 4-item indices measuring agreement with the adoption of algorithmic systems in policing and medicine, respectively. The dependent variables and all independent variables are normed to a range from 0 to 1. N is lower in Model 4 as fewer respondents have answered the question on the probability of getting skin cancer.
$^*p < 0.05$. $^{**}p < 0.01$.

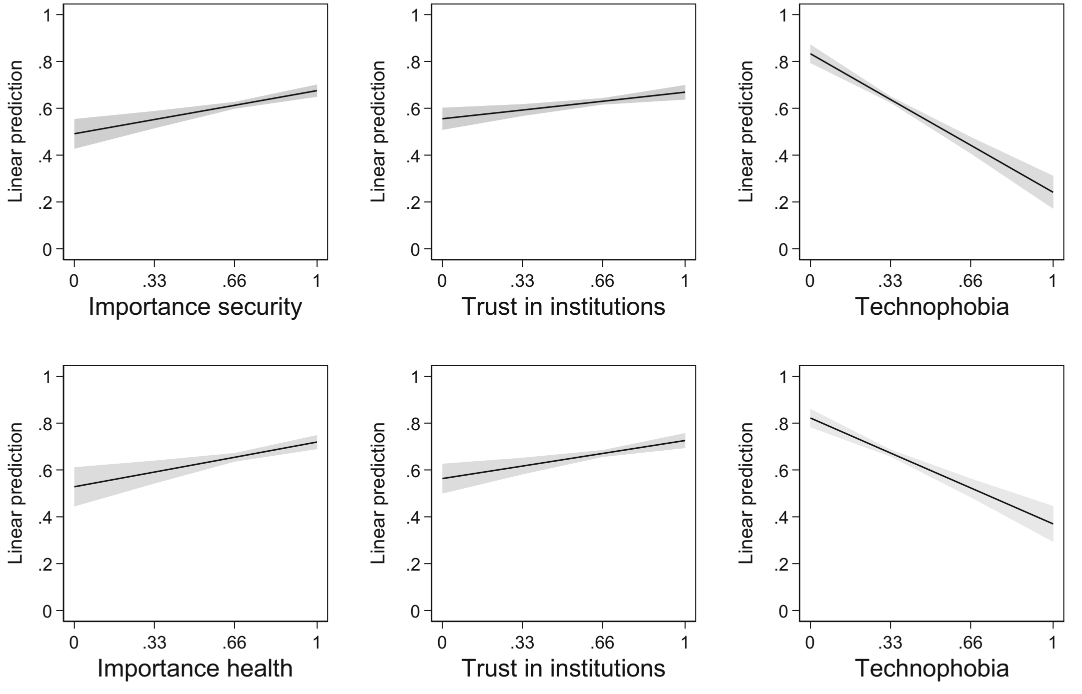

FIGURE 3 Predictive marginal effects with 95% confidence intervals for selected variables. Top panels: Predictive policing context; bottom panels: Skin cancer prediction context. Predictive margins are based on models 2 and 4, displayed in Table 2. The dependent variables and all independent variables are normed to a range from 0 to 1.

people with a strong demand for transparency are less inclined to accept algorithms, the corresponding coefficient for skin cancer prediction is far from significant. This is a striking finding which deserves further inspection. It is possible that it results from policing—but not health risk prediction—being seen as a strong intervention into people's liberty, thus evoking stronger demands for transparency that can prevent abuse of state power.

A second relevant control variable is the measure of technophobia, which is highly significant and shows a very strong negative association with algorithm acceptance both in predictive policing and skin cancer prediction (see also

Figure 3, above). While this finding is not surprising given the literature on technophobia, the large coefficient is still notable and provides an important point of reference. The estimated coefficients of around −0.6 in the predictive policing example means that changing from the minimum score to the maximum score on technophobia reduces the support for algorithms in predictive policing by about three-fifths. Finally, the regression results indicate that higher age is associated with lower algorithm support in the medical, but not in the policing setting. Why this is the case cannot be concluded from the data at hand, but it could be that older people are less ready to give up exclusively human expertise and advice in the medical context.

Turning finally to the model fit, the explained variance of the two main regression models amounts to R^2 scores of around 26% for the predictive policing setting and 17%–20% for skin cancer prediction. A likely reason for the markedly lower explanatory power in the skin cancer condition is the non-significant coefficient of the principled transparency stance. Also, technophobia seems to be less powerful as a predictor of algorithm acceptance in skin cancer prediction.

5 | CONCLUSION

The increasing adoption of algorithms that are supposed to provide services to citizens in various areas of everyday life has sparked much scholarly attention centered on questions about the ethically acceptable and accountable design of algorithmic systems. Existing research has emphasized the issue of possible—intended or unintended—algorithmic biases that might go against the interests of those they are supposed to serve and has discussed how algorithm design and especially transparency and accountability features can prevent these biases.

While research that focuses on such technical aspects of algorithms is clearly important, this focus risks missing an important facet when it comes to the public acceptance of algorithms: the social context, in which algorithms are employed. When directly prompted to evaluate the importance of an algorithm's design features, people may well deem certain technical features very important. Yet this does not mean that this is how people will commonly perceive and evaluate algorithms in their daily lives. This is especially true for public sector applications of algorithms with which citizens do not directly come into contact and that they do not use themselves—hence, where they lack user experience in the strict sense. As we have argued above, it is therefore important to adopt a perspective that considers the larger social context in which such algorithms are implemented because aspects of this context may already cue how much citizens support this technology. This perspective can therefore offer an important extension and corrective to a view that narrowly looks at the algorithms and their design features.

In line with that perspective, we have argued and shown empirically that the personal importance of the problem that an algorithm is supposed to deal with and the values at stake clearly matter for the extent to which citizens show general support of algorithms in policing and health care. Also, the extent to which people trust the organizations behind the implementation of an algorithm matters for their acceptance of the technology. If people trust the organizations deploying the algorithms, acceptance rates for the algorithms are clearly higher—a finding that is in line with what research on organizational reputation would suggest.

These associations are strong even in comparison to technophobia as an attitude that, as to be expected, shows a very strong relation with algorithm acceptance. In contrast, the algorithm's performance as a key technical feature of an ADM system only weakly affected algorithm acceptance in the two contexts that we studied. This effect is based on an experimental treatment that manipulated the algorithm performance respondents saw in our survey, but without pushing respondents to explicitly consider this performance in their general support of algorithms. A likely reason for this finding is that respondents do not have a clear frame of reference regarding what constitutes a good performance. Being unable to process and evaluate such technical information about an algorithm, it would seem only natural that they rely on the context in which the algorithm is deployed as they can more easily evaluate this setting. Features of the algorithms may, however, be much more important when directly confronting citizens with cases of failures and especially if values such as fairness or transparency are violated (Schiff et al., 2022).

In line with previous research (Schiff et al., 2022), our findings are rather similar between the two different settings—although it is important to note that this does not mean that context is irrelevant. In fact, as we measure context-specific attitudes, we can assess how much context-specific factors matter; and our results indeed indicate that the more citizens value public safety and their personal health, the more they support the use of an algorithm in the respective domain. However, we also found some notable differences. First, the perceived importance of transparency of government is relevant only in the policing setting, but not for the health care context. This finding may have to do with the fact that policing is a case of state interference that potentially affects one's basic liberties, which makes institutional transparency much more important for citizens than in the context of skin cancer prediction. Furthermore, the result could also be triggered by a discussion about police violence when the survey was in the field. Second, whereas age shows a negative association with algorithm acceptance in the medical setting, it does not in the policing setting. Possible reasons for this are that health-related risks are perceived differently depending on age, which is not the case for perceived security risks, or that older people want medical advice exclusively by humans. Future research could investigate more in detail how people's life situations affect support for algorithms, and how much these matter in comparison to features of the algorithm that they are confronted with.

The findings altogether suggest that it is important to consider attitudes toward aspects of the specific context in which algorithms as part of socio-technical systems are embedded. Citizens' evaluations of algorithms seem to be strongly related to the personal importance of the area of application, the attitudes toward the organization (e.g., the government) deploying an algorithm and toward the problem that the algorithm addresses. The personal stakes and the trust vis-à-vis the deployment organization appear to significantly shape whether people accept the use of algorithms—and this is true for both contexts. Hence, looking narrowly at how technical features of algorithms are related to their public acceptance may not entirely capture how people form opinions about algorithms in the public sector. Rather, citizens seem to be pragmatic in the sense that they base their assessments of algorithms on more familiar and accessible context-related factors—if not already on such general dispositions as their technophobia.

The findings thus imply that citizens may come to support the use of algorithms in the public sector even if they do not realize expected public values. A lack of characteristics, such as responsiveness or transparency, that guarantee public value may well matter for algorithm acceptance (Schiff et al., 2022). However, citizens may not even be aware of how and how well algorithmic systems in the public sector perform as they do not directly come into contact with them. If they rely on contextual cues, as described above, the realization of public values may ultimately not be what matters for algorithm acceptance among citizens. People could already accept algorithms simply because they have a generally positive attitude toward digital technologies, trust the organization responsible for implementing the algorithm, or care a lot about their expected benefits (e.g., preventing crimes) that the algorithm promises to deliver. Yet, the opposite may also be true. People could reject an algorithm that is well-designed (e.g., in terms of accountability and commonly accepted fairness standards) for reasons that are completely detached from the design of the algorithm itself.

If citizens are led to trust algorithms already based on context-related factors, this could also mean that accountability problems highlighted with the emergence of an algorithmic bureaucracy (Busuioc, 2021; Veale & Brass, 2019; Vogl et al., 2020) are of relevance mainly in academic discussions and within bureaucracies, but not necessarily something that is immediately problematized by citizens. Consequently, it seems unlikely that there will be strong pressure from the larger public demanding an open deliberation and justification of value-laden choices behind the design of algorithmic systems. Yet, such deliberation becomes especially important as discretion is transferred from frontline decision-makers to those designing algorithmic decision-making systems (Busuioc, 2021; Levy et al., 2021). Finally, our findings also have consequences for the possibilities of politicizing algorithms (on this, see Katzenbach & Ulbricht, 2019), as it seems possible to advocate or attack the use of algorithms without having to refer to any technical features of the algorithm itself. Based on our findings, it seems especially difficult to publicly defend the introduction of algorithms in the public sector that are of little palpable value to citizens and where trust in public administration is rather low.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable comments and suggestions. The manuscript also owes to the feedback by Adam Harkens, Tobias Krafft, Johannes Schmees, Wolfang Schulz, Karen Yeung, and Katharina Zweig. The authors disclose receipt of the following financial support for the research, authorship, and/or publication of this article: This research has been conducted within the project "Deciding about, by, and together with algorithmic decision-making systems," funded by the Volkswagen Foundation (Grant reference 19-0087). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/padm.12901.

OPEN RESEARCH BAGES

This article has been awarded Open Materials, Open Data, Preregistered Research Designs badges. All materials and data are publicty accessible via the Open Science Framework at [http://re3data.org/]. Learn more about the Open Practices badges from the Center for Open Science: http://re3data.org/.

DATA AVAILABILITY STATEMENT

Please find the files for replication https://doi.org/10.7910/DVN/4AACQE.

ORCID

Georg Wenzelburger https://orcid.org/0000-0001-6195-7664
Pascal D. König https://orcid.org/0000-0001-9466-4024
Julia Felfeli https://orcid.org/0000-0001-6837-4728
Anja Achtziger https://orcid.org/0000-0001-7207-9342

ENDNOTES

[1] The required sample size for a commonly posited predictive power of 0.8 can be determined using Cohen's f^2. This score reflects the change in explained variance when a set of predictors is included in a model and can thus be seen as a measure of the effect size. According to Cohen (1988, p. 413), an $f^2 = 0.02$, corresponding to an increase of 2% in R^2, constitutes a small effect. To calculate the required sample size in order to find an effect of this size under a given alpha and with a given predictive power, one also needs to specify the number of predictors in the reference model as a fourth parameter. To find a small effect of a single regression coefficient under alpha $= 0.05$ and with a predictive power of 0.8, the required sample size is 395 cases (see Cohen, 1988, p. 437; calculated using the R package pwr using one degree of freedom in the numerator for testing a single predictor). It should be noted that the highly significant main effects in the analysis presented further below are larger than the posited small effects with $f^2 = 0.02$. For instance, Cohen's f^2 is 0.09 for the personal importance of security and 0.05 for institutional trust in the predictive policing model (this has been calculated using the R package effect size). Besides a power analysis for the effect size of individual predictors, one can also perform the analysis for the entire model. To be able to find a small f^2-value of 0.02 for the entire model (containing 17 predictors), as statistically significant under an alpha of 0.05 with a predictive power of 0.8, a sample size of 999 is required.

[2] The attention check is one item within the technophobia item battery (order randomized) and reads "Attention is very important for this study. Please select the option 'do not agree at all' for this option." The control question is the very last question of the survey. It states that the quality of the study depends on careful and honest answers and asks respondents to answer truthfully whether their responses are usable, reassuring them that they would be receive remuneration in any case. Respondents could then choose "Yes, I have provided answers carefully and with best conscience.

My responses can be used for the analysis" or "No, I have not carefully and earnestly provided answers to the survey. My responses should rather not be used for the analysis."

[3] Using the control question as a filter means losing 27 cases, applying the attention check question means we additionally lose 197 of 2573 cases (73 are missing). Having applied these filters, 113 further cases do not pass the speeding check. Robustness checks have been performed without the speeding and additionally without the attention check, leading to substantively identical results and conclusions (see Annex A13).

[4] These introductory texts were also pre-tested the survey with a sample of persons from diverse backgrounds as well as students and adjusted some of the texts in response to this feedback before fielding the survey.

[5] The median time respondents spent on the instruction materials is 33 s for the predictive policing setting. Together with the series of presented algorithm designs, this time increases to 173 s. For the health care setting, these times are 21 and 162 s. While respondents have thus overall spent ample time getting to know algorithmic systems and different possible realizations, the time spent on the instructions with the basic information alone is rather low. We have therefore done additional analysis with cases that lie above the median (see Annex A14). Importantly, these analyses show that, despite a severe reduction of the sample, main results of our analysis still hold. More details on the time that respondents spent on the individual pages of the survey can be looked up in the Annex, Table A15. Similarly, even when we run the regressions with only those participants who have spent more than 3 min on the introductory texts reducing the samples to less than 200 observations for each of the cases (tables can be obtained from the authors), all results of theoretical interest (e.g., on H1A, H1B, and H2) remain nevertheless significant (with the exception of the variable covering "personal importance of addressed problem [H1A]" for health, which loses significance at conventional values).

[6] Results for the individual variables are shown in Annex A9.

[7] Existing objective measures of algorithmic literacy are cognitively rather taxing (see, e.g., Cheng et al., 2019).

[8] See Annex A10 for results with this variable.

[9] This contrast is empirically informed by actual performance: predictive policing algorithms achieve a true positive rate of below 10% (Mohler et al., 2015), whereas skin cancer prediction algorithms reach about 90% (Roffman et al., 2018). Hence, while a low or high performance is realistic only for one of the contexts respectively, using the experimental treatment within *both* the policing and health context allows for keeping the context constant. Although one can presume that citizens do not generally know what constitutes a realistic algorithm performance in these settings, we conducted a robustness check using a variable on self-reported expert domain and/or algorithm knowledge (see Annex A11).

[10] We inspected whether the treatment groups showed significant differences with regard to any of the associational independent variables used in the analysis. The results show a lack of balance with regard to three of overall 28 inspected variables (14 in each setting): being worried about a burglary, the perceived probability of a break-in, and extroversion (see Annex A7). As we control for these and other variables in the regression model, this reduces the likelihood of obtaining biased estimates of the treatment effect.

[11] As the emotional indicators of subjective worry (about burglaries or skin cancer) are highly correlated with the cognitive indicators (estimated probability of burglaries or skin cancer), we decided to estimate two separate regression models in order to alleviate multicollinearity. To probe whether any independent variables in the main model in Table 2 show unduly strong intercorrelations that would decrease the accuracy of the coefficient estimates, we have checked for multicollinearity. The results suggest that there are no problems of multicollinearity in the presented model as all scores are below the commonly used rule-of-thumb threshold of 4 (see Annex Table A8). Moreover, due to the presence of heteroscedasticity, we used robust standard errors (Huber-White).

[12] This effect is robust to using the other measures of personal importance described above in the section on data and measures (see Annex A12 for the regression models) and to excluding those respondents from the sample that have spent less than the median of time for the introductory part of the survey (Annex Table A14).

[13] The pattern is similar if we exclude those respondents who have spent less than the median time on the introductory part of the survey (see Annex Table A14).

[14] Results are less clear-cut for predictive policing if we reduce the sample size to those who have spent more than the median time on the introductory part (see Annex Table A14). The variable for institutional trust is not significant. However, this due to the simultaneous inclusion of still highly significant "principled transparency" variable and collinearity between the two variables becoming an issue in light of substantially lower numbers of observation. The trust variable does become significant when excluding demand for transparency.

REFERENCES

Ananny, M. & Crawford, K. (2018) Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. Available from: https://doi.org/10.1177/1461444816676645

Aoki, N. (2020) An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly*, *37*(4), 101490. Available from: https://doi.org/10.1016/j.giq.2020.101490

Araujo, T., Helberger, N., Kruikemeier, S. & de Vreese, C.H. (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. Available from: https://doi.org/10.1007/s00146-019-00931-w

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J. & Shadbolt, N. (2018) "It's reducing a human being to a percentage": perceptions of justice in algorithmic decisions. In: *Proceedings of the 2018 CHI conference on human factors in computing systems – CHI'18*. Montreal, QC: ACM Press, pp. 1–14. Available from: https://doi.org/10.1145/3173574.3173951

Birkel, C., Church, D., Hummelsheim-Doss, D., Leitgöb-Guzy, N. & Oberwittler, D. (2018) *Der Deutsche Viktimisierungssurvey 2017: Opfererfahrungen, kriminalitätsbezogene Einstellungen sowie die Wahrnehmung von Unsicherheit und Kriminalität in Deutschland*. Wiesbaden: Bundeskriminalamt.

Brown, S., Davidovic, J. & Hasan, A. (2021) The algorithm audit: scoring the algorithms that score us. *Big Data & Society*, *8*(1), 205395172098386. Available from: https://doi.org/10.1177/2053951720983865

Busuioc, M. (2021) Accountable artificial intelligence: holding algorithms to account. *Public Administration Review*, *81*, 825–836. Available from: https://doi.org/10.1111/puar.13293

Chatterjee, S., Khorana, S. & Kizgin, H. (2022) Harnessing the potential of artificial intelligence to foster citizens' satisfaction: an empirical study on India. *Government Information Quarterly*, *39*, 101621. Available from: https://doi.org/10.1016/j.giq.2021.101621

Chen, T., Guo, W., Gao, X. & Liang, Z. (2021) AI-based self-service technology in public service delivery: user experience and influencing factors. *Government Information Quarterly*, *38*(4), 101520. Available from: https://doi.org/10.1016/j.giq.2020.101520

Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through UI: strategies to help non-expert stakeholders. *Proceedings of the 2019 CHI conference on human factors in computing systems – CHI'19*, Paper No. 559,1–12. Glasgow, Scotland: ACM Press. https://doi.org/10.1145/3290605.3300789

Chiusi, F., Fischer, S., Kayser-Bril, N. & Spielkamp, M. (2020) *Automating Society Report 2020*. Berlin; Güterslog: AlgorithmWatch; Bertelsmann Foundation.

Coglianese, C. & Ben Dor, L.M. (2020) AI in adjudication and administration. *Brooklyn Law Review*, 86(3), 791-838.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*, 2nd edition. Hillsdale, NJ: L. Erlbaum Associates.

Davis, F.D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. Available from: https://doi.org/10.2307/249008

Glikson, E. & Woolley, A.W. (2020) Human trust in artificial intelligence: review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. Available from: https://doi.org/10.5465/annals.2018.0057

Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P. & Weller, A. (2018) Human perceptions of fairness in algorithmic decision making: a case study of criminal risk prediction. In: *Proceedings of the 2018 World Wide Web Conference on World Wide Web – WWW'18*. Lyon: ACM Press, pp. 903–912. Available from: https://doi.org/10.1145/3178876.3186138

Grimmelikhuijsen, S. (2022) Explaining why the computer says no: algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, online first, 1–22. Available from: https://doi.org/10.1111/puar.13483

Ingrams, A., Kaufmann, W. & Jacobs, D. (2022) In AI we trust? Citizen perceptions of AI in government decision making. *Policy & Internet*, *14*, 390–409. Available from: https://doi.org/10.1002/poi3.276

Johnson, D.G. & Verdicchio, M. (2017) Reframing AI discourse. *Minds and Machines*, *27*(4), 575–590. Available from: https://doi.org/10.1007/s11023-017-9417-6

Juravle, G., Boudouraki, A., Terziyska, M. & Rezlescu, C. (2020) Trust in artificial intelligence for medical diagnoses. In: Parkin, B.L. (Ed.) *Progress in brain research*, Vol. 253. Amsterdam: Elsevier, pp. 263–282. Available from: https://doi.org/10.1016/bs.pbr.2020.06.006

Katzenbach, C. & Ulbricht, L. (2019) Algorithmic governance. *Internet Policy Review*, *8*(4), 1-18. Available from: https://doi.org/10.14763/2019.4.1424

Kehr, F., Kowatsch, T., Wentzel, D. & Fleisch, E. (2015) Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus: privacy calculus: dispositions and affect. *Information Systems Journal*, *25*(6), 607–635. Available from: https://doi.org/10.1111/isj.12062

Kennedy, R.P., Waggoner, P.D. & Ward, M.M. (2022) Trust in public policy algorithms. *The Journal of Politics*, *84*, 1132–1148. Available from: https://doi.org/10.1086/716283

Kim, D.J., Ferrin, D.L. & Rao, H.R. (2008) A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. *Decision Support Systems*, *44*(2), 544–564. Available from: https://doi.org/10.1016/j.dss.2007.07.001

King, W.R. & He, J. (2006) A meta-analysis of the technology acceptance model. *Information & Management*, *43*(6), 740–755. Available from: https://doi.org/10.1016/j.im.2006.05.003

Krafft, T.D., Zweig, K.A. & König, P.D. (2022) How to regulate algorithmic decision-making: a framework of regulatory requirements for different applications. *Regulation & Governance*, *16*(1), 119–136. Available from: https://doi.org/10.1111/rego.12369

Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. et al. (2017) Accountable algorithms. *University of Pennsylvania Law Review*, *165*, 633–705.

König, P.D. & Wenzelburger, G. (2021) The legitimacy gap of algorithmic decision-making in the public sector: why it arises and how to address it. *Technology in Society*, *67*, 101688. Available from: https://doi.org/10.1016/j.techsoc.2021.101688

Langer, M., König, C.J. & Fitili, A. (2018) Information as a double-edged sword: the role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior*, *81*, 19–30. Available from: https://doi.org/10.1016/j.chb.2017.11.036

Lau, R.R., Hartman, K.A. & Ware, J.E. (1986) Health as a value: methodological and theoretical considerations. *Health Psychology*, *5*(1), 25–43. Available from: https://doi.org/10.1037/0278-6133.5.1.25

Lee, M.K. (2018) Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. Available from: https://doi.org/10.1177/2053951718756684

Lepri, B., Oliver, N., Letouzé, E., Pentland, A. & Vinck, P. (2018) Fair, transparent, and accountable algorithmic decision-making processes: the premise, the proposed solutions, and the open challenges. *Philosophy & Technology*, *31*(4), 611–627. Available from: https://doi.org/10.1007/s13347-017-0279-x

Levy, K., Chasalow, K.E. & Riley, S. (2021) Algorithms and decision-making in the public sector. *Annual Review of Law and Social Science*, *17*(1), 309–334. Available from: https://doi.org/10.1146/annurev-lawsocsci-041221-023808

Logg, J.M., Minson, J.A. & Moore, D.A. (2019) Algorithm appreciation: people prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. Available from: https://doi.org/10.1016/j.obhdp.2018.12.005

Marangunić, N. & Granić, A. (2015) Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81–95. Available from: https://doi.org/10.1007/s10209-014-0348-1

Meijer, A. (2018) Datapolis: a public governance perspective on "smart cities.". *Perspectives on Public Management and Governance*, *1*(3), 195–206. Available from: https://doi.org/10.1093/ppmgov/gvx017

Meijer, A., Lorenz, L. & Wessels, M. (2021) Algorithmization of bureaucratic organizations: using a practice lens to study how context shapes predictive policing systems. *Public Administration Review*, *81*(5), 837–846. Available from: https://doi.org/10.1111/puar.13391

Miller, S.M. & Keiser, L.R. (2021) Representative bureaucracy and attitudes toward automated decision making. *Journal of Public Administration Research and Theory*, *31*(1), 150–165. Available from: https://doi.org/10.1093/jopart/muaa019

Miller, S.M., Song, M. & Keiser, L.R. (2022) The effect of human versus automated interaction on willingness to participate in government programs: the role of representation. *Public Administration*, online first, 1–18. Available from: https://doi.org/10.1111/padm.12879

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016) The ethics of algorithms: mapping the debate. *Big Data & Society*, *3*(2), 1–21. Available from: https://doi.org/10.1177/2053951716679679

Mohler, G.O., Short, M.B., Malinowski, S., Johnson, M., Tita, G.E., Bertozzi, A.L. et al. (2015) Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, *110*(512), 1399–1411. Available from: https://doi.org/10.1080/01621459.2015.1077710

New Zealand Government. (2018) *Algorithm assessment report*. Wellington: New Zealand Government. Available from: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf

Nimrod, G. (2018) Technophobia among older internet users. *Educational Gerontology*, *44*(2–3), 148–162. Available from: https://doi.org/10.1080/03601277.2018.1428145

Piotrowski, S.J. & Van Ryzin, G.G. (2007) Citizen attitudes toward transparency in local government. *The American Review of Public Administration*, *37*(3), 306–323. Available from: https://doi.org/10.1177/0275074006296777

Prokop, C. & Tepe, M. (2022) Talk or type? The effect of digital interfaces on citizens' satisfaction with standardized public services. *Public Administration*, *100*(2), 427–443. Available from: https://doi.org/10.1111/padm.12739

Roffman, D., Hart, G., Girardi, M., Ko, C.J. & Deng, J. (2018) Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Scientific Reports*, *8*(1), 1701. Available from: https://doi.org/10.1038/s41598-018-19907-9

Schepman, A. & Rodway, P. (2020) Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, *1*, 100014. Available from: https://doi.org/10.1016/j.chbr.2020.100014

Schiff, D.S., Schiff, K.J. & Pierson, P. (2022) Assessing public value failure in government adoption of artificial intelligence. *Public Administration*, *100*, 653–673. Available from: https://doi.org/10.1111/padm.12742

Schwartz, S.H. (2012) An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, *2*(1), 1-20. Available from: https://doi.org/10.9707/2307-0919.1116

Shin, D. & Park, Y.J. (2019) Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284. Available from: https://doi.org/10.1016/j.chb.2019.04.019

Shin, D. (2020) How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, *109*, 106344. Available from: https://doi.org/10.1016/j.chb.2020.106344

Shin, D. (2021) The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. Available from: https://doi.org/10.1016/j.ijhcs.2020.102551

Shin, D., Zhong, B. & Biocca, F.A. (2020) Beyond user experience: what constitutes algorithmic experiences? *International Journal of Information Management*, *52*, 102061. Available from: https://doi.org/10.1016/j.ijinfomgt.2019.102061

Sinkovics, R.R., Stöttinger, B., Schlegelmilch, B.B. & Ram, S. (2002) Reluctance to use technology-related products: development of a technophobia scale: reluctance to use technology-related products. *Thunderbird International Business Review*, *44*(4), 477–494. Available from: https://doi.org/10.1002/tie.10033

Starke, C. & Lünich, M. (2020) Artificial intelligence for political decision-making in the European Union: effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*, *2*, e16. Available from: https://doi.org/10.1017/dap.2020.19

Veale, M. & Brass, I. (2019) Administration by algorithm? Public management meets public sector machine learning. In: Yeung, K. & Lodge, M. (Eds.) *Algorithmic regulation*. Oxford: Oxford University Press, pp. 121–149.

Vogl, T.M., Seidelin, C., Ganesh, B. & Bright, J. (2020) Smart technology and the emergence of algorithmic bureaucracy: artificial intelligence in UK local authorities. *Public Administration Review*, *80*(6), 946–961. Available from: https://doi.org/10.1111/puar.13286

Wachter, S., Mittelstadt, B. & Floridi, L. (2017) Transparent, explainable, and accountable AI for robotics. *Science Robotics*, *2*(6), eaan6080. Available from: https://doi.org/10.1126/scirobotics.aan6080

Waggoner, P.D., Kennedy, R., Le, H. & Shiran, M. (2019) Big data and trust in public policy automation. *Statistics, Politics and Policy*, *10*(2), 115–136. Available from: https://doi.org/10.1515/spp-2019-0005

Wæraas, A. & Ma'or, M. (2015) Understanding organizational reputation in a public sector context. In: Wæraas, A. & Ma'or, M. (Eds.) *Organizational reputation in the public sector*. New York, NY: Routledge, pp. 1–14.

Young, M.M., Bullock, J.B. & Lecy, J.D. (2019) Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration, perspectives on public management and governance, 2 (4), 301–313. Available from: https://doi.org/10.1093/ppmgov/gvz014

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wenzelburger, G., König, P. D., Felfeli, J., & Achtziger, A. (2024). Algorithms in the public sector. Why context matters. *Public Administration*, *102*(1), 40–60. https://doi.org/10.1111/padm.12901