

ARTICLE

# Motivated reasoning about artificial intelligence in public policy: comparative evidence from Germany

Sebastian Hemesath<sup>1</sup>  and Markus Tepe<sup>2</sup> 

<sup>1</sup>Department of European Social Research, Institute of Political Science, Saarland University, Saarbrücken, Germany and <sup>2</sup>Department of Political Science, SOCIUM Research Center Inequality and Social Policy, University of Bremen, Bremen, Germany

**Corresponding author:** Markus Tepe; Email: [markus.tepe@uni-bremen.de](mailto:markus.tepe@uni-bremen.de)

(Received 7 May 2024; revised 22 December 2024; accepted 10 February 2025)

## Abstract

This study tests whether citizens' evaluations of the performance of artificial intelligence (AI) in public policies are subject to motivated reasoning. Specifically, we test whether respondents' preferences for AI regulation or their subjective attitudes toward AI are sources of motivated reasoning across varying use cases, differing in nature, complexity, safety-criticality and normative considerations: AI in municipal services, self-driving cars and recidivism prediction. Experimental results from two preregistered studies conducted among German citizens reveal that subjective attitudes toward AI cause substantial and robust motivated reasoning across all three policy domains. Regulatory preferences are only a selective source for motivated reasoning about AI in public policy. Overall, the results point to the cognitive limitations of strategies that attempt to objectify the benefits of AI without considering the context of the application domain. Politicians and policymakers need to consider these limitations in their attempts to increase citizens' appreciation of AI in public policy.

**Keywords:** algorithm aversion; artificial intelligence; bounded rationality; motivated reasoning; survey experiment

## Introduction

Artificial intelligence (AI) is believed to have significant potential to enhance public services' effectiveness, efficiency and convenience. However, citizens often meet the use of AI in the public domain with skepticism (König and Wenzelburger, 2021), as AI-based technologies raise important normative questions about the quality of decision-making in public policy (Bullock, 2019; Busuioc, 2021; Grimmelikhuisen and Meijer, 2022). A popular approach to counter citizens' skepticism is objectively evaluating AI's benefits (Cartwright and Hardie, 2012; Hortal, 2023). According to evidence-based

policymaking (EBPM), such objective evidence can be provided through a rigorous evaluation of novel technology, for example, in pilot projects. The evaluation studies that emerge from such pilot projects are supposed to provide hard statistical evidence for the advantages and disadvantages of AI. This article examines how effective such objectification can be concerning the performance assessment of AI in three policy domains with varying levels of safety-criticality (Krafft *et al.*, 2022).

A fundamental premise of EBPM is that citizens, politicians and administrators can draw inferences from scientific evidence (Cartwright and Hardie, 2012; Oliver, 2013; Dhami and Sunstein, 2022). The validity of the data presented in a single report and the policy implications drawn from such an evaluation study may be open to debate, and citizens' cognitive abilities to understand and process information may also vary. Nevertheless, any objectification strategy must presume that citizens can – in principle – correctly infer the information from simple statistical evidence presented in evaluation studies. Motivated reasoning challenges this fundamental premise of evidence-based public policy (e.g., Kunda, 1990; Mintz *et al.*, 2021) and describes the phenomenon that peoples' existing beliefs hinder their ability to interpret scientific evidence correctly (Kahan *et al.*, 2017: 56). Instead of updating their beliefs according to new evidence, they make biased evaluations in defense of their desired conclusions, even when scientific evidence suggests the opposite (Christensen and Moynihan, 2020: 3–4).

The presence of motivated reasoning has already been documented in various policy domains: peoples' beliefs about the causes of climate change, for example, biases their evaluation of climate data (Druckman and McGrath, 2019), peoples' ideological orientation biases their assessment of gun control policies (Redlawsk, 2002; Leeper and Slothuus, 2014; Martin *et al.*, 2020), as well as their performance evaluation of public organizations (Baekgaard and Serritzlew, 2016). This study tests whether citizens' assessment of the performance of AI in three use cases is influenced by motivated reasoning. Specifically, we test whether motivates reasoning of processing information about the performance of AI is present and whether such biased information processing stems from citizens' preference for state regulation of AI-based systems or their subjective attitudes toward AI.

Current research suggests that citizens' skepticism toward AI-based technologies in the public domain tends to be conditional on the specific task delegated to or supported by AI (Wenzelburger *et al.*, 2024). Experimental studies presenting citizens' design options and regulatory frameworks on AI across multiple policy domains (Miller and Keiser, 2020; Kennedy *et al.*, 2022; König *et al.*, 2024; Grimmelikhuijsen, 2023) confirm that technology-associated risk perceptions play a crucial role in AI acceptance. On the other hand, recent evidence on the moderating effect of ideological predispositions on regulatory preferences indicates a robust effect across contexts (Hemesath and Tepe, 2023, 2024a; König *et al.*, 2023). This literature shows that several preexisting beliefs about AI, such as its perceived threat to personal values and ethics (Kleizen *et al.*, 2023), about companies building or promoting AI for the public sector, as well as general beliefs about the role of technology in society, influence citizens' acceptance of AI in public policy (Hemesath and Tepe, 2024b).

This study contributes to this literature on AI in public policy (Gesck and Leyer, 2022; Kleizen *et al.*, 2023; Wenzelburger *et al.*, 2024) by examining citizens' cognition of

information about the performance of AI systems, testing whether citizens' regulation preferences and attitudes toward AI are a relevant source of motivated reasoning about AI-based systems across varying public policy domains. Drawing on an established experimental design to detect motivated reasoning (Christensen and Moynihan, 2020), we conducted two online survey experiments among German citizens and examined motivated reasoning in three distinct public administration tasks that vary in complexity, safety criticality and normative considerations: routine public administration of allocating parking permits, the public safety approval of self-driving cars and the forecast of recidivism of incarcerated individuals.

## Theoretical framework

Motivated reasoning builds on psychological insights (e.g., cognitive dissonance theory, Festinger, 1957), and describes the phenomenon that peoples' preexisting beliefs shape their ability to process and interpret information, often in biased ways to protect these beliefs (Kunda, 1990; Taber and Lodge, 2006). The core assumption is that individuals' reasoning is goal-driven, influencing their cognitive processing and judgment of information (Druckman and McGrath, 2019). Rather than acknowledging novel information and updating their beliefs, people choose to optimize a tradeoff between accuracy and directional motives (Little, 2024). This study considers a narrow conceptualization of motivated reasoning as a biased evaluation of policy information.<sup>1</sup>

Motivated reasoning about AI occurs if citizens are more likely to correctly infer statistical evidence about the performance of an AI in public policy if the evidence is congenial to their AI regulation preferences and less likely to correctly infer statistical evidence if it is uncongenial to these preferences. This mechanism offers an important addendum to existing research on public perceptions of AI, which often assumes rational evaluation of AI systems. Existing research indicates that individuals tend to base evaluations of AI on contextual cues, heuristics and prior dispositions (Schiff *et al.*, 2022; Kleizen *et al.*, 2023; Wenzelburger *et al.*, 2024). By directly targeting the cognitive mechanism of motivated reasoning, this study goes beyond prior research by investigating how citizens' biases influence their interpretation of evidence about AI systems, offering a new lens to understand public perceptions of AI in policy contexts.

As a controversially discussed technology, AI intersects with key public values, such as fairness, privacy and redistribution, which are tied to individuals' ideological orientations, trust in institutions, emotional investment and perceived risks (König and Wenzelburger, 2021; Horvath *et al.*, 2023; Hemesath and Tepe, 2024b). These values make AI evaluations personally and politically salient, even in contexts where direct experiences with AI may be limited. Therefore, citizens may interpret information on the performance of an AI-based system in public policy primarily to affirm their existing beliefs about AI. Hence, motivated reasoning diminishes the likelihood that scientific evidence about the performance and reliability of AI in public policy will be assessed based on its true value. We subsequently hypothesize that citizens engage in motivated reasoning when they evaluate the performance of AI-based systems and

---

<sup>1</sup>Little (2024) models the link between motivated reasoning and Bayesian updating. We will pick up on these considerations in the discussion.

interpret information on the performance of AI in public policy to reaffirm their existing beliefs. The baseline hypothesis states:

**H1.** Citizens engage in motivated reasoning when evaluating an AI-based system's performance in public policymaking.

### *Two sources of directional motives toward AI*

Motivated reasoning is a universal cognitive distortion that has been observed in various areas (Mintz *et al.*, 2021). In this respect, one might say that the challenge is not whether motivated reasoning occurs but which motives are responsible for this bias in a particular context. This study tests two types of beliefs that might cause motivated reasoning toward AI in public policy.

First, citizens might have political beliefs about the general need to regulate AI and the role of the state. The public use of technology is connected to questions of resource distribution, public safety, public trust and state intervention. The state and corresponding regulations significantly influence when, how and in which shape technology is and can be used. Moreover, the evaluation of AI performance in the context of EBPM specifically occurs to make inferences on political regulation. Existing research on citizens' preferences toward the regulation of AI, however, suggests that respondents' self-reported ideological position has little explanatory power (Hemesath and Tepe, 2023, 2024a; König *et al.*, 2023). This sharply contrasts with many other policy fields (e.g., welfare politics, fiscal policy, moral policies), where citizens' ideological orientation is crucial for policy preferences. Likewise, previous research on motivated reasoning by Christensen and Moynihan (2020) and others (Baekgaard and Serritzlew, 2016; Martin *et al.*, 2020) shows that political views are responsible for motivated reasoning about the performance of public vs private organizations.

A possible explanation for these findings might be a lack of elite discourses (Druckman, 2012) on the regulation of AI in the past, reducing citizens' ability to use partisan cues and heuristics for reasoning. With recent debates about how society should deal with AI and substantial legislation, such as the European Union's AI act, passed, this issue should be more politically salient for individuals (Laux *et al.*, 2024). Explicitly, we anticipate that citizens' regulatory preference toward AI, their preference for more or less state regulation of AI-based technologies, is a source of motivated reasoning toward AI-based systems in various policy domains.

**H2.** Citizens' preference toward regulating AI-based technologies is a source of motivated reasoning biasing the evaluation of AI in public policymaking.

Second, citizens might have preexisting beliefs about AI on a more deeply rooted psychological level rather than regulation preferences that stem from ideological beliefs. Citizens' emotions and subjective beliefs toward AI could represent a second source of motivated reasoning. Schepman and Rodway (Schepman and Rodway, 2023: 2725–2776) conceptualize citizens' general attitudes toward AI alongside a positive and a negative dimension: positive attitudes correspond to positive perceptions of the utility of AI (e.g., economic opportunities, improved performance), individual attitudes toward using such technologies (e.g., at work or in private life) and overly positive

emotions (e.g., excitement, being impressed). Negative attitudes correspond to significant concerns about AI's safety (e.g., unethical use, making errors) and strong negative emotions regarding AI (e.g., discomfort, finding AI sinister). In contrast to AI regulation preferences (more or less state regulation of AI-based systems), subjective AI attitudes are rooted on subjective, emotional and psychological levels. The technology acceptance model (TAM) is a popular approach to explaining consumers' willingness to adopt technology (Davis, 1989). However, when it comes to AI in the public sector, citizens are not consumers in the sense that they have the freedom to choose, nor does TAM capture the psychological aspects of attitudes toward AI. Schepman and Rodway (2023) show that their measure of AI attitudes correlates with respondents' personality traits and generalized social trust. In light of this research pointing toward the importance of technology attitudes, we expect that citizens are more likely to correctly infer statistical evidence about the performance of an AI in public policy if the evidence is congenial to their subjective AI attitudes and less likely to correctly infer statistical evidence if it is uncongenial to these attitudes.

**H3.** Citizens' subjective AI attitude is a source of motivated reasoning biasing the evaluation of AI in public policymaking.

While both regulation preferences and general attitudes toward AI are expected to bias citizens' evaluations of AI systems, the literature on motivated reasoning suggests that emotionally stronger beliefs or attitudes may exert a greater influence (Kunda, 1990; Taber and Lodge, 2006). General attitudes toward AI are often deeply tied to subjective psychological and emotional factors, e.g., perceptions of fairness, risk and trust (Schepman and Rodway, 2023), which are known to amplify motivated reasoning. In contrast, regulation preferences, which are connected to broader ideological beliefs about state intervention, may be weaker sources of motivated reasoning in this context, given the relatively low levels of politicization observed in related studies (Hemesath and Tepe, 2023; König *et al.*, 2023). Nonetheless, given the relative novelty of AI as a policy issue, this comparison remains empirical.

## Research design

### *Policy domains*

There is no comparative evidence on whether or how the policy domain influences motivated reasoning, or whether it is stable across varying contexts. Previous research suggests that respondents' perceived issue salience moderates the magnitude of related cognitive biases such as framing (Lecheler *et al.*, 2009). Existing research on AI in public policy suggests that the area of application matters for citizens' willingness to entrust AI-based systems. Some evidence points toward a general appreciation of AI-based services, and in some instances, even preferred over human decision-making (e.g., Logg *et al.*, 2019). Other studies revealed significant apprehensions about the use of AI in general (Dietvorst *et al.*, 2015) and in the public domain in particular (Wenzelburger *et al.*, 2024). Castelo *et al.* (2019) suggest that algorithm apprehension is contingent on the nature of the task. They found that respondents were particularly reluctant to assign tasks to AI-based systems when considering them subjective, e.g.,

selecting clothing or writing a poem. The reason, they argue, is that algorithms are believed to lack human-like emotional characteristics. In contrast, allegedly objective tasks, quantifiable by figures, were more likely to be entrusted to AI-based systems. This is because algorithms are often perceived as more accurate and rational than human judgment. Wenzelburger *et al.* (Wenzelburger *et al.*, 2024: 42) provide further evidence that citizens' acceptance of AI-based systems is contingent upon the concrete area of application. Studying the use of AI in policing and health care, they find that personal importance is an essential determinant of citizens' acceptance of AI (Wenzelburger *et al.*, 2024: 54). In contrast, evidence on the moderating effect of beliefs and ideological predispositions on regulatory preferences for AI suggest those beliefs to have a relatively stable effect across various domains (Hemesath and Tepe, 2023, 2024a; König *et al.*, 2023). Subsequently, there is a theoretical argument for both, with the sources for motivated reasoning being context-dependent or stable across contexts.

Disentangling the various context factors (such as stakes, actor roles, task complexity and normative considerations) and potential case-specific sources presents a methodological challenge. This study starts from the opposite end and explores whether more general sources of motivated reasoning possess substantial variation across substantially different use cases for AI. Specifically, we test citizens' motivated reasoning based on general preferences for regulation and subjective attitudes toward AI to evaluate AI-based systems in three distinct public policy contexts. They differ in the nature of the tasks, task complexity, the state's role, their safety-criticality (Krafft *et al.*, 2022) and normative considerations. In all three cases, an AI-based system is used to fulfill a task that a human individual has conducted so far.

First, we investigate motivated reasoning in routine public service in municipal administration. Allocating parking permits represents a low-stakes, low-complexity case of state-citizen service encounters (Lipsky, 2010). Previous research on citizens' acceptance of AI in public services also considered standardized routine public services as a low-stake venue for digitalizing public services (Prokop and Tepe, 2022; Grimmelikhuijsen, 2023; Horvath *et al.*, 2023), particularly since some of these systems are already in trial use (e.g., semi-automated self-service terminals<sup>2</sup>). While technically relatively simple, such tasks can still provoke concerns about transparency and impartiality when citizens perceive unequal treatment.

Second, we investigate motivated reasoning in the context of self-driving cars. The approval of self-driving cars reflects a technically complex, regulatory function of the state that has considerable safety implications for individuals. Compared to administrative processes, the state no longer serves as a provider of services but is tasked with oversight. The safety implications of self-driving cars and how citizens' risk perceptions affect their acceptance have been the topic of extensive research (Liu *et al.*, 2019). On a normative level, evidence (Bonnefon *et al.*, 2016; Awad *et al.*, 2018) suggests citizens'

---

<sup>2</sup>See for example: <https://www.bundesdruckerei-gmbh.de/en/solutions/self-service-terminal>.

evaluation is highly emotional, as they prefer others to buy a self-driving car that sacrifices the passengers in favor of other traffic participants but would like to ride in a self-driving car that protects the passengers at all costs. The lack of trust in self-driving cars is supported by the occupants' strong preferences for permanent supervision of self-driving cars (Hemesath and Tepe, 2024a).

Third, we investigate the evaluation of AI for predicting the recidivism of incarcerated individuals. This scenario captures a less tangible and more abstract risk for personal safety relating to a fear of crime (Hart *et al.*, 2022). Already in use in some parts of the US juridical system (Dressel and Farid, 2018), various studies have tested under which circumstances citizens are willing to rely on AI for predicting recidivism (Meijer and Wessels 2019; Hartmann and Wenzelburger, 2021; Meijer *et al.*, 2021; Kennedy *et al.*, 2022; Wenzelburger *et al.*, 2024), suggesting that citizens are more likely to approve using such systems if they adhere to normative standards, such as transparency and when human decision-makers have room to adjust and interpret the machine forecast instead of directly implementing it. Predicting recidivism using AI introduces profound ethical and normative concerns that extend beyond technical considerations (Hartmann and Wenzelburger, 2021) and raise questions about fairness, bias and legitimacy, particularly when outcomes impact marginalized communities disproportionately. In this context, the dual nature of harm (potentially depriving defendants of liberty through false positives or jeopardizing public safety through false negatives) highlights its salience as a morally and politically ambivalent case. Competing concerns could shape citizens' evaluations: protecting individual rights vs safeguarding public safety. Evaluations may thus align with individuals' broader values and normative beliefs about justice and security. When the justice system fails to deliver impartial enforcement of laws, the state risks eroding its legitimacy, as citizens may perceive it as incapable of maintaining order or serving the common good (Tyler, 1990).

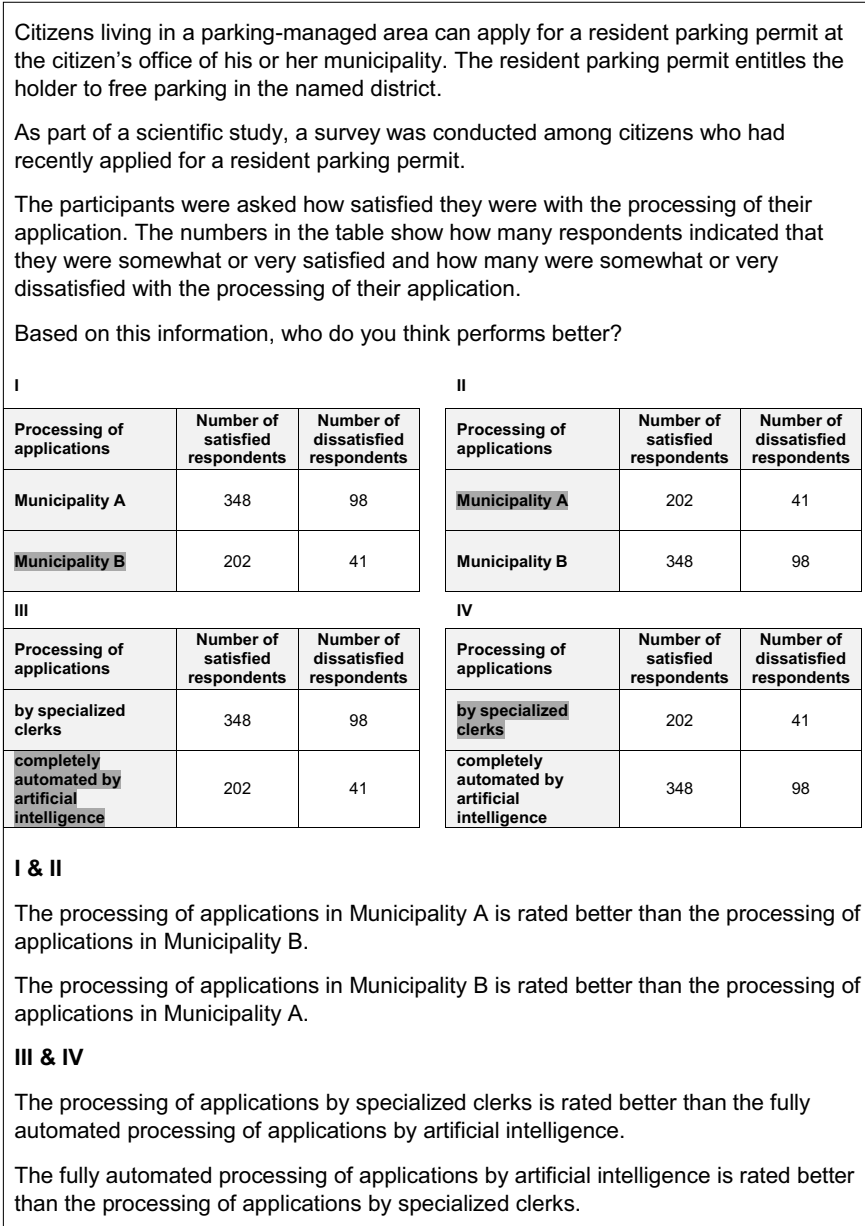
The three use cases are expected to provide sufficient variation to examine the stability of motivated reasoning across policy domains.

### **Experimental instrument**

We use an established experimental instrument (Baekgaard and Serritzlew, 2016; Baker *et al.*, 2020; Christensen and Moynihan, 2020). In the following, we will explain how it was adjusted for the three use cases (see Figures 1–3).

Respondents are randomly assigned to one of four (I–IV) contingency tables for each use case (between-subjects design). The tables illustrate the results of a supposed evaluation study of either (1) citizens' satisfaction with public service delivery at a municipal office, (2) the results of a safety reliability test of new types of vehicles or (3) the reliability of predictions made for determining recidivism of criminal offenders: For each use case, respondents evaluate the performance of two types of procedures, where the placebo group (Tables I and II) uses neutral labeling of the procedures as Type A vs Type B, and the treatment groups (Tables III and IV) use the same results, but respondents evaluate a human decision-maker and an AI-based system. The key to correctly interpreting the presented data is to calculate the difference in ratio between the two rows (Baker *et al.*, 2020: 204). The information in the contingency tables is





**Figure 1.** Satisfaction with municipal service.

Note: Darker gray marks the numerically correct answer. Text marked in gray has not been reported to respondents.

unambiguous; the answers to which municipal office receives a better satisfaction rating, which type of vehicles, or which recidivism prediction is more reliable can be identified as either correct or incorrect.



Before new vehicle types are approved for sale on the automotive market, they must meet strict safety requirements (type approval testing).

A scientific study investigated the reliability of two new vehicle types. To this end, the vehicle types were tested for their susceptibility to faults in standardized test drives on public roads.

The number of test drives where a safety-critical incident occurred was evaluated. Safety-critical incidents are technical system failures that risk the lives of occupants and other road users.

The numbers in the table indicate the number of test drives in which a safety-critical incident occurred.

Based on this information, which type of vehicle do you think is more reliable?

I

Vehicle type	Rides without safety-critical incident	Rides with safety-critical incident
Automobile A	223	75
Automobile B	107	21

II

Vehicle type	Rides without safety-critical incident	Rides with safety-critical incident
Automobile A	107	21
Automobile B	223	75

III

Vehicle type	Rides without safety-critical incident	Rides with safety-critical incident
Conventional automobile	223	75
Self-driving cars	107	21

IV

Vehicle type	Rides without safety-critical incident	Rides with safety-critical incident
Conventional automobile	107	21
Self-driving cars	223	75

**I & II**

Type A vehicles are more reliable than type B vehicles.

Type B vehicles are more reliable than type A vehicles.

**III & IV**

Conventional automobiles are more reliable than self-driving cars.

Self-driving cars are more reliable than conventional automobiles.

**Figure 2.** Safety approval of new cars.

Note: Darker gray marks the numerically correct answer. Text marked in gray has not been reported to respondents.

In Tables I and III, procedure Type B, respectively, the AI-based system performs better. In Tables II and IV, the numbers are reversed, making procedure Type A, respectively, the human decision-maker perform better. In Figures 1–3, the

In criminal law, there is the possibility of early release from prison. For this purpose, a forecast must be made as to how likely the incarcerated person will recidivate.

Within the framework of a scientific study, it was examined how reliable two prognosis procedures are for determining recidivism. For this purpose, the accuracy of the prognosis procedure was tested using a training set. Of the offenders included in the training set, whether they recidivate after early release is known.

The cases in which the forecast was favorable, i.e., no recidivism was expected, but the person then nevertheless committed another severe crime after release in which other people were harmed, are problematic.

The numbers in the table indicate how often a favorable forecast was correct or incorrect.

Based on this information, which prognosis method do you think is more reliable?

**I**

Forecast	Correct (no recidivism)	False (offender relapsed)
Procedure A	348	98
Procedure B	202	41

**II**

Forecast	Correct (no recidivism)	False (offender relapsed)
Procedure A	202	41
Procedure B	348	98

**III**

Forecast	Correct (no recidivism)	False (offender relapsed)
by specialized psychiatric experts	348	98
by a computer program using artificial intelligence	202	41

**IV**

Forecast	Correct (no recidivism)	False (offender relapsed)
by specialized psychiatric experts	202	41
by a computer program using artificial intelligence	348	98

**I & II**

The forecasts of procedure A are more reliable than the forecasts of procedure B.

The forecasts of procedure B are more reliable than the forecasts of procedure A.

**III & IV**

The forecasts by specialized psychiatric experts are more reliable than those from a computer program that uses artificial intelligence.

The forecasts from a computer program that uses artificial intelligence are more reliable than those from specialized psychiatric experts.

**Figure 3.** Predicting recidivism in incarcerated person.

Note: Darker gray marks the numerically correct answer. Text marked in gray has not been reported to respondents.

numerically correct answer is marked in dark gray. In Tables I and II, using the neutral frame, respondents' ability to correctly identify the correct answer should only depend on respondents' numeracy. Hence, we should not observe motivated reasoning in those

conditions. We utilize Tables I and II as a control or placebo test, offering a baseline against which the occurrence of motivated reasoning can be tested (Christensen and Moynihan, 2020: 8–9).<sup>3</sup> For the outcome variable, we ask respondents to select the correct statement on which system performs better: Type A or B in Tables I and II, or the AI system vs the human decision-maker in Tables III and IV of the contingency tables. In each use case, the two answer options below the table were presented in random order to minimize a potential bias in the dependent variable due to order effects (e.g., always pick the first option).

Respondents' technology attitudes are measured with two constructs (see Appendix Tables 3 and 4): First, we measure citizens' preference toward the political regulation of AI using four statements: (1) *Artificial intelligence should be more strictly regulated by the government.* (2) *Companies that produce artificial intelligence should be able to act on the market without restriction.* (3) *The government should take responsibility for ensuring that artificial intelligence benefits all people.* (4) *Each person should be responsible for judging artificial intelligence's benefits and drawbacks.* Second, we measure respondents' attitudes toward AI using the General Attitudes toward Artificial Intelligence Scale (GAAIS) developed by Schepman and Rodway (Schepman and Rodway, 2020, 2023). Both item batteries are measured before respondents are asked to evaluate the information in the contingency table. At the end of the survey, all respondents were informed that the evidence presented in the contingency table was hypothetical examples (debriefing).

### **Samples and data analysis**

The experiment was conducted on two samples of the German population drawn from the online access panel of 'Bilendi'. The AI-based allocation of municipal parking permits was tested in Study 1. Prior power analysis assuming a small effect of  $f = 0.2$  ( $\alpha = 0.05$ ; power = 0.8) resulted in a required sample size of  $n = 788$  for both studies. For both studies, we recruited adult respondents through the online access panel of 'Bilendi', employing quotas for age, gender, state of residence and level of education. Bilendi reimbursed participants as per their standing agreements. We included an attention check in the GAAIS item battery to increase sample quality. Respondents failing to pass the attention check were excluded from the sample. After that, Study 1 (municipal parking permits) included 1692 respondents, and Study 2 (self-driving cars and prediction of recidivism) comprised 2464 respondents.<sup>4</sup>

<sup>3</sup>Study 2 also included a use case in which citizens were asked to evaluate the performance of an AI-based system that detects tumors from images. However, we had to remove this use case from the analysis due to an implementation error. In the health use case, the values and labels in two tables were inadvertently reversed, making the answer button inconsistent. Since these errors negate any meaningful analysis, we exclude all observations from the health use case from the analysis. Including the non-sensical data from the health use case as a control in the main analysis does not change the results for the other three use cases.

<sup>4</sup>The OSF preregistration for Study 1 can be found here: <https://doi.org/10.17605/OSF.IO/SBEPV>. The experimental material for Study 1 can be found here: <https://osf.io/aygtm/>. The OSF preregistration for Study 2 can be found here: <https://doi.org/10.17605/OSF.IO/7M3AJ>. The experimental material for Study 2 can be found here: <https://osf.io/3nbe8/>. The dataset and syntax are available on the Harvard Dataverse: <https://doi.org/10.7910/DVN/JC6F9O>.

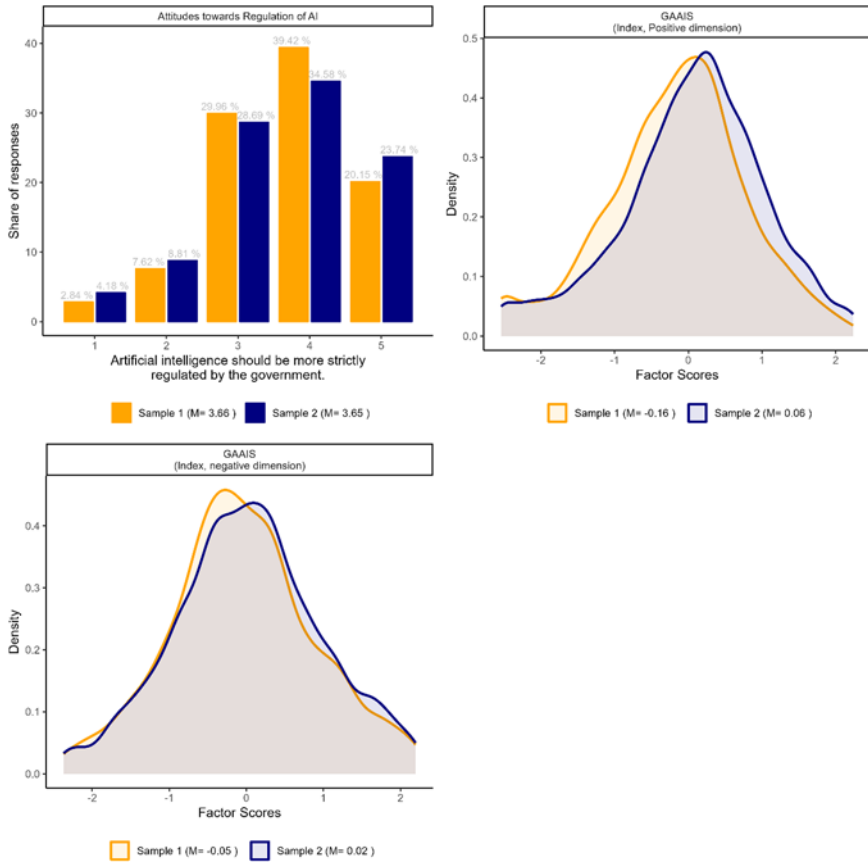


Figure 4. AI regulation preference and AI attitudes.

The dependent variable measures whether the respondent chose the correct answer for a given contingency table (yes/no). The five items on AI regulation preferences show an alpha of 0.34. Since this is indicative of low internal consistency and little correlation between the individual items, we elected to use only the first item of that battery ('Artificial intelligence should be more strictly regulated by the government.' 1 strongly disagree, ..., 5 = strongly agree) for further analysis, as it best captures our intended belief dimension. The GAAIS items (Schepman and Rodway, 2023) were transformed into two indexes measuring positive and negative attitudes toward AI. For both subdimensions, we find strong internal consistency with an alpha of 0.94 for the positive scale and an alpha of 0.88 for the negative scale. We ran a factor analysis for each subdimension and extracted factor scores. The distribution of respondents' preferences for AI regulation and the factor scores of the GAAIS subindexes are depicted in Figure 4.<sup>5</sup> Descriptive statistics on the distribution of socio-demographic

<sup>5</sup>The correlation between the scale for negative attitudes toward AI and the preference for state regulation of AI is 0.31. The correlation between the scale for positive attitudes toward AI and the preference for state regulation of AI is -0.15.

variables (gender, age, education and region) from the two samples are summarized in Appendix Figure 1.

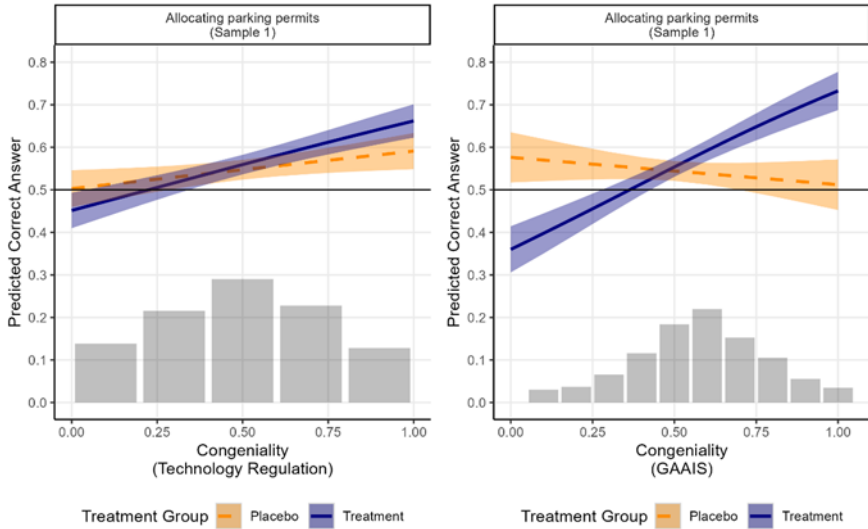
To identify motivated reasoning, we estimate the marginal effect (ME), conditioned by congeniality. The ME is calculated based on two logistic interaction models that predict respondents' correct choices. The binary logistic model for estimating the baseline results is estimated in the following form:  $Correct\ Choice = Congeniality \times Treatment + Covariates + Error$ . *Correct Choice* is a dummy variable that equals one if the respondent picked the correct answer and zero if the answer was incorrect. We coded two different *congeniality* variables, one based on respondents' AI regulation preference and the other based on the two GAAIS scales. For each measure, congeniality reports how strongly (rescaled on a scale from 0 to 1) individuals' attitudes are in accordance with the correct choice in each condition (e.g., the AI-based system is the correct answer, and respondents report a positive attitude toward AI). If respondents evaluated Tables I and III, where AI performs better, the congeniality measure is the invert of 'stricter regulation of AI' and the index of the GAAIS scale for the positive subdimension. For Tables II and IV, where the human performs better, congeniality captures attitudes toward 'stricter regulation of AI', respectively, the negative GAAIS subdimension index measure. *Treatment* is a nominal variable indicating whether the respondent evaluated versions I and II (placebo group) or III and IV (treatment group) of the contingency tables. Since selecting the correct choice could also depend on individuals' numeracy, we control for respondents' level of education, age and gender (see Appendix Table 1).<sup>6</sup>

## Results

### Baseline analysis

Findings on motivated reasoning in the three use cases are summarized in Figures 5–7. The left panel of each figure reports the results for the congeniality measure based on respondents' preferences for stricter AI regulation. In comparison, the right panel reports the results based on the congeniality measure for respondents' subjective attitudes toward AI (GAAIS). The  $y$ -axis denotes the probability that respondents chose the correct answer, while the  $x$ -axis represents the congeniality measure ranging from 0, indicating complete uncongeniality (respondents' prior beliefs strongly contradict the objectively correct result), and 1, indicating high congeniality (respondents' prior beliefs strongly aligning with the objectively correct result). The lines report the effect of congeniality on the probability respondents chose the correct answer. The dotted orange line illustrates the effect on the placebo groups, and the solid blue line illustrates the treatment effect on the treatment groups. To illustrate the distribution of varying levels of congeniality among participants, we included histograms detailing the frequency of observations.

<sup>6</sup>The estimation results show that higher levels of education increase the probability of choosing the correct answer (see Appendix Tables 1 and 2). The probability of choosing the correct answer is also higher for males than for females and increases with age. Excluding these control variables from the statistical model does not alter any of the main results.

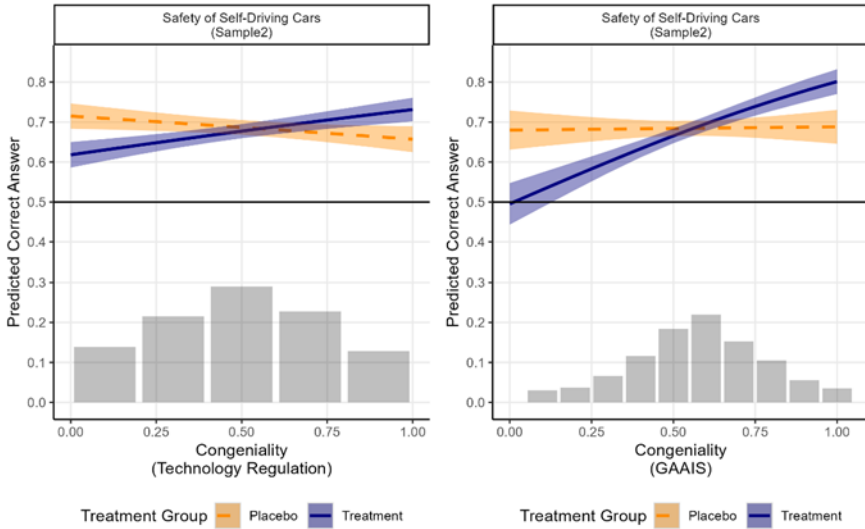


**Figure 5.** Evaluation of an AI in allocating parking permits (Study 1).  
 Note: GLM estimates with robust SE clustered at the respondent level (Appendix Table 1).

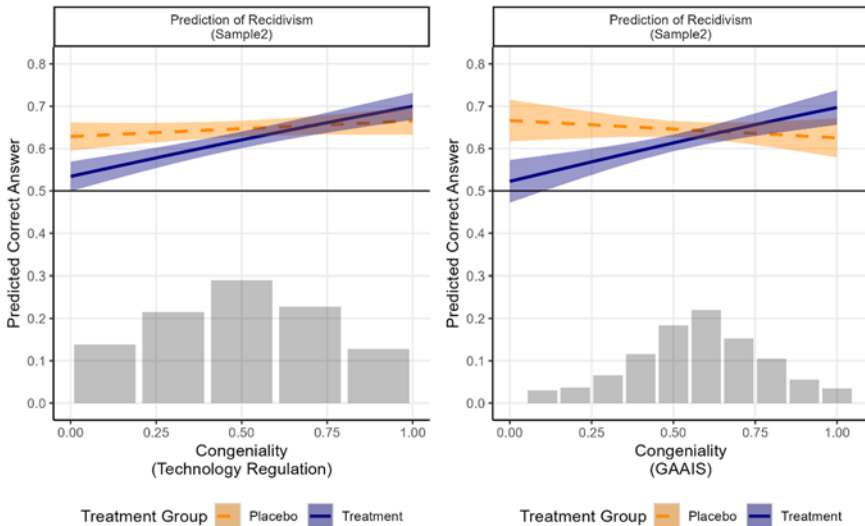
For the case of the allocation of municipal parking permits (Figure 5), we find that information that is in accordance with citizens’ AI regulation preferences hardly affects respondents’ likelihood of choosing the correct answer. While we see a positive slope in the treatment group, we find no significant differences from the placebo group. In contrast, concerning respondents’ subjective attitudes toward AI (GAAIS), we find a solid symmetric effect of congeniality on respondents’ likelihood of choosing the correct answer. In contrast to the placebo group, respondents are significantly more likely to select the correct answer if the correct answer aligns with their prior subjective AI attitudes. Respondents in the treatment groups are about 20 percentage points more likely to choose the correct answer if it is highly congenial to their initial beliefs (compared to respondents in the placebo group). At the same time, they are more than 20 percentage points less likely to select the correct choice if said choice is strongly uncongenial to their subjective AI attitudes. For both measures, we find no effect of congeniality in the placebo group, strengthening the robustness of the findings.

For the case of self-driving cars (Figure 6), we find a similarly strong effect of congeniality based on subjective attitudes toward AI as observed in the first scenario and a mild positive slope of congeniality based on attitudes toward technology regulation. While we observe no treatment effect in the placebo group for subjective attitudes toward AI, we observe a slight negative slope in the placebo group for congeniality based on preferences for AI regulation. While this strengthens the robustness of the finding on the impact of subjective attitudes, the positive effect of regulatory preferences should be interpreted with caution.

Last, for the case of predicting recidivism (Figure 7), we find only uncongeniality of information based on regulatory preferences to significantly affect respondents’ choices, reducing their likelihood to correctly evaluate the scenario by up to



**Figure 6.** Evaluation of AI in self-driving cars' safety (Study 2).  
 Note: GLM estimates with robust SE clustered at the respondent level (Appendix Table 1).



**Figure 7.** Evaluation of an AI in predicting recidivism (Study 2).  
 Note: GLM estimates with robust SE clustered at the respondent level (Appendix Table 1).

15 percentage points compared to the placebo group. We find no significant differences when the information provided is congenial to regulatory preferences. For subjective attitudes toward AI, we find a similarly strong asymmetrical effect that is strong and significant if information is uncongenial to the evaluated data, but we observe no



significant differences between treatment groups and placebo groups if the information is highly congenial to subjective beliefs.

Overall, subjective attitudes toward AI (GAAIS) produce a robust pattern of motivated reasoning across all three use cases. When the information provided in the contingency table is uncongenial to individuals' attitudes, individuals perform significantly worse than in the placebo groups. When the information is in accordance with respondents' AI attitudes, they perform significantly better. In comparison, regulatory preferences are found to be only a selective source of motivated reasoning.

### **Secondary analysis**

Our results indicate considerable differences in absolute correct choices between samples. While respondents in the first sample, the use case for assessing AI in the allocation of parking permits, gave the correct answer on average 52% of the time (across treatment groups), we find this rate to increase to 66% (self-driving cars) and 61% (prediction of recidivism) in the second sample. To test whether this was caused by different sample quality, we tested whether our results were significantly moderated by the response time of individuals. Figures A2 and A3 report the results for both measures of congeniality, conditional on whether respondents answered in a low, medium or high response time (relative to the median response time). The lines represent the baseline effects for reference, while the points illustrate the binned estimates for low, medium and high levels of congeniality and the shapes the low, medium and high groups of response time. For both measures, we find considerably stronger heterogeneity in the first sample, while results in the second sample are more robust across varying levels of response time. Disregarding individual outliers potentially subject to the limited power of this additional analysis (three-way interaction), the general trend in the first sample, however, supports the conclusions drawn from the baseline results, as we find similar slopes across the different response time groups.

Second, we repeated the baseline analysis on binned subgroups to strengthen the robustness of this finding and rule out that the reported effects are due to overfitting (Hainmueller *et al.*, 2019). We binned the congeniality measure based on AI regulation preferences and subject attitudes toward AI into three terciles and estimated the same regression models used in the main analysis (see Appendix Table 2). The estimation results using the binned congeniality measures are summarized in Appendix Figure 5. The robustness analysis strengthens our findings (also see Appendix Table 3).

Third, selecting only one of the items to measure preferences for regulation might induce unintended confounding by wording or comprehension. To strengthen the robustness of this measure, we repeated the analysis with the fourth item ('Each person should be responsible themselves for judging artificial intelligence's benefits and drawbacks'), which should yield an identical effect if congeniality is reversed. Appendix Figure 4 reports the results for this analysis, indicating results that are similar in direction yet significantly more pronounced than with item 1. Here, we can observe a strong symmetrical effect for the prediction of recidivism, meaning that individuals who strongly support this statement are significantly more likely to correctly evaluate the AI as better if the AI is better and significantly less likely to correctly evaluate the human as better if the human performs better.

Last, previous studies (Christensen and Moynihan, 2020) also tested for potential variation of motivated reasoning among subgroups (e.g., age, gender, level of education or political ideology). While the results from such analyses should be treated with caution, considering the issue of multiple comparisons and lack of statistical power induced by a limited sample size, we likewise tested for the moderating effect of age, gender, level of education and political ideology. The results of those tests (Appendix Figures 6a–9b) indicate no substantial variation among subgroups.

## Discussion and conclusions

Objectifying the benefits of AI-based systems in the public sector through systematic evaluations is a critical element in policymakers' strategies to overcome skepticism and increase the legitimacy of AI in public service provision (König and Wenzelburger, 2021; Grimmeliikhuijsen and Meijer, 2022). This study addressed how motivated reasoning about AI might limit the objectivation of the pros and cons of AI in public policies. In a series of experiments, we tested whether citizens' AI regulation preferences or their subjective attitudes toward AI cause motivated reasoning when they are asked to evaluate the performance of AI-based technology in three distinctive policy domains.

Experimental results from two preregistered studies conducted among German citizens confirm the existence of motivational reasoning about AI in public policy (H1) with two refinements: First, apart from the case of predicting recidivism, citizens' regulatory preferences toward AI are not a vital source of motivating reasoning (H2). In contrast to other studies on motivated reasoning in public policies, e.g., public vs private provision of education and health services (Baekgaard and Serritzlew, 2016; Martin *et al.*, 2020), we do not find preferences for state regulation to be a significant source of motivated reason about AI in public services. This result suggests that explicit political or ideological preferences may not strongly influence citizens' assessments of AI performance in the three policy contexts studied here. Nevertheless, this interpretation should be approached with caution. The analysis does not take into account specific political attitudes, such as beliefs about crime or trust in government, that might influence citizens' reasoning. The findings for predicting recidivism draw attention to the possible influence of institutional trust or the salience of normative concerns, particularly in contexts where issues such as public safety and justice are closely linked to political attitudes (Harcourt, 2007). Probing deeper, one could examine how specific political attitudes, such as beliefs about crime, public safety or government authority, interact with citizens' evaluations of AI, as they represent an important dimension of trust and legitimacy in public policy.

Second, citizens' subjective attitudes toward AI are a robust source of motivating reasoning about the performance of AI in public policies (H3). In contrast to AI regulation preferences, these attitudes are more deeply rooted on an emotional and psychological level, as GAAIS has been shown to correlate with interpersonal trust and personality traits (Schepman and Rodway, 2023). This might explain why AI attitudes' role in motivating reasoning about AI in public policies does not differ substantially between the three use cases utilized in this study. Considering that the AI attitudes measured with GAAIS are related to respondents' personality and other

psychological parameters, AI attitudes as a source of motivated reasoning toward using AI in public policies may be hard to overcome. Research on debiasing interventions to mitigate motivated reasoning based on prior political or ideological beliefs offers a fruitful theoretical perspective but with mixed empirical results (Christensen and Moynihan, 2020; Boissin *et al.*, 2021). A follow-up consideration is how to design simple and effective interventions to mitigate the distorting effects of motivated reasoning about AI in public policy.

All of these results are based on a narrow conceptualization of motivated reasoning, asking respondents to identify a numerically correct answer from a contingency table. However, the data source itself may be perceived as a signal rather than an objective evaluation. In this case, motivated reasoning, understood as ‘once-motivated reasoning’, is less irrational (Little, 2024); instead, it becomes a process in which contradictory evidence is given less weight in respondents’ belief updating. Bayesian updating can be regarded as a complementary perspective on how individuals process information. It suggests that prior beliefs are not fixed but can be gradually updated in light of new evidence. Little (2024) proposed a model combining Bayesian updating and ‘once-motivated reasoning’ about information signals, showing that subjects sometimes completely reject signals that lead to less pleasant beliefs. This model could better account for the increasing skepticism, even hostility, that can be observed toward scientific evaluations. To validate this promising model of the relationship between motivated reasoning and Bayesian updating, one could manipulate the source of the data in the contingency tables (e.g., evaluations by an independent scientific body or a partisan think tank).

This study is not without methodological limitations. First, this study is certainly limited by its measurement of regulatory preferences and attitudes toward AI. Future research could explore the robustness of the findings presented here by using alternative measures of respondents’ demand for regulation of AI (Heinrich and Witko, 2024) and psychological resonance toward AI, such as the technophilia scale (Martínez-Córcoles *et al.*, 2017). Second, this study does not investigate how motivated reasoning about AI influences real-world behaviors, such as voting or technology adoption. While cognitive measures provide important insights into the cognitive processes underlying decision-making, future research should examine the behavioral implications of motivated reasoning on political decision-making in the case of AI. Finally, this study is based on survey data from a single country. While this limits the generalizability of findings to other institutional contexts and policy environments, Germany represents an interesting and appropriate case for examining motivated reasoning. Germans are known for their relatively high-risk aversion, paired with comparatively high levels of trust in government and administration. These characteristics may amplify motivated reasoning about AI in public policy. A follow-up study could adopt a cross-country framework to explore whether our findings hold in countries with differing levels of trust in government, cultural attitudes toward technology and administrative traditions.

What policy implications can be drawn from these findings? This study should not be considered an argument for abandoning the ambition of EBPM despite the robust bias in citizens’ evaluations of AI performance (H1) caused by subjective attitudes toward AI. Instead, we believe these findings point to the limitations of policy

objectification (Sunstein, 2002; Oliver, 2013; Barak-Corren and Kariv-Teitelbaum, 2021; Dhimi and Sunstein, 2022). Our study suggests that the evaluation of scientific data about the pros and cons of AI alone will not be sufficient to improve public acceptance among those with solid pre-perceptions of AI. In addition to identifying effective debiasing measures, future studies should also address the question of how to target debiasing, i.e., how to provide objective information to groups that are receptive to this information and support the evaluation of policy by merit, whereas highly biased individuals are unlikely to be swayed by additional information in their evaluation of AI in the public sector. Finally, we have no reason to believe that motivated reasoning about AI based on subjective attitudes toward AI is limited to public sector employees. Thus, if the phenomenon is general and applies to public employees, their subjective attitudes toward AI in public policy should be addressed and carefully considered in strategies to digitize public services.

**Supplementary material.** To view supplementary material for this article, please visit <https://doi.org/10.1017/bpp.2025.2>.

**Acknowledgements.** This study was funded by the Niedersächsisches Vorab VWVN 1466 (Niedersächsischer Ministerium für Wissenschaft und Kultur). Previous versions of this study have been presented, for example, at the Tripartite Political Science Conference of the Swiss (SVPW), Austrian (ÖGPW), and German (DVPW) Political Science Association at Linz University in 2023. We want to thank all participants for their valuable comments, which contributed to the revision of the study.

## References

- Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff and I. Rahwan (2018), 'The moral machine experiment', *Nature*, **563**(7729): 59–64.
- Baekgaard, M. and S. Serritzlew (2016), 'Interpreting performance information: motivated reasoning or unbiased comprehension', *Public Administration Review*, **76**: 73–82.
- Baker, S. G., N. Patel, C. Von Gunten, K. D. Valentine and L. D. Scherer (2020), 'Interpreting politically-charged numerical information: the influence of numeracy and problem difficulty on response accuracy', *Judgment and Decision Making*, **15**: 203–213.
- Barak-Corren, N. and Y. Kariv-Teitelbaum (2021), 'Behavioral responsive regulation: bringing together responsive regulation and behavioral public policy', *Regulation & Governance*, **15**: 163–S82.
- Boissin, E., S. Caparos, M. Raelison and W. De Neys (2021), 'From bias to sound intuiting: boosting correct intuitive reasoning', *Cognition*, **211**: 104645.
- Bonnefon, J. F., A. Shariff and I. Rahwan (2016), 'The social dilemma of autonomous vehicles', *Science*, **352**(6293): 1573–1576.
- Bullock, J. B. (2019), 'Artificial intelligence, discretion, and bureaucracy', *The American Review of Public Administration*, **49**: 751–761.
- Busuioac, M. (2021), 'Accountable artificial intelligence: holding algorithms to account', *Public Administration Review*, **81**: 825–836.
- Cartwright, N. and J. Hardie (2012), *Evidence-Based Policy: A Practical Guide to Doing It Better*, Oxford: Oxford University Press.
- Castelo, N., M. W. Bos and D. R. Lehmann (2019), 'Task-dependent algorithm aversion', *Journal of marketing research*, **56**(5): 809–825.
- Christensen, J. and D. P. Moynihan (2020), 'Motivated reasoning and policy information: politicians are more resistant to debiasing interventions than the general public', *Behavioural Public Policy* **8**(1): 47–68.
- Davis, F. D. (1989), 'Perceived usefulness, perceived ease of use, and user acceptance of information technology', *MIS Quarterly*, **13**(3): 319–340.
- Dhimi, S. and C. R. Sunstein (2022), *Bounded Rationality: Heuristics, Judgment, and Public Policy*, Cambridge, MA: MIT Press.

- Dietvorst, B. J., 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.
- Dressel, J. and H. Farid (2018), 'The accuracy, fairness, and limits of predicting recidivism', *Science advances*, **4**(1): eaao5580.
- Druckman, J. N. (2012), 'The politics of motivation', *Critical Review*, **24**(2): 199–216.
- Druckman, J. N. and M. C. McGrath (2019), 'The evidence for motivated reasoning in climate change preference formation', *Nature Climate Change*, **9**: 111–119.
- Festinger, L. (1957), *A Theory of Cognitive Dissonance*. Redwood City: Stanford University Press. <https://doi.org/10.1515/9781503620766>
- Gesk, T. S. and M. Leyer (2022), 'Artificial intelligence in public services: when and why citizens accept its usage', *Government Information Quarterly*, **39**: 101704.
- Grimmelikhuijsen, S. (2023), 'Explaining why the computer says no: algorithmic transparency affects the perceived trustworthiness of automated decision-making', *Public Administration Review*, **83**: 241–262.
- Grimmelikhuijsen, S. and A. Meijer (2022), 'Legitimacy of algorithmic decision-making: six threats and the need for a calibrated institutional response', *Perspectives on Public Management and Governance*, **5**: 232–242.
- Hainmueller, J., J. Mummolo and Y. Xu (2019), 'How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice', *Political Analysis*, **27**(2): 163–192.
- Harcourt, B. E. (2007), *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*, Chicago: The University of Chicago Press.
- Hart, T., M. Chataway and J. Mellberg (2022), 'Measuring fear of crime during the past 25 years: A systematic quantitative literature review', *Journal of Criminal Justice*, **82**: 101988, <https://doi.org/10.1016/j.jcrimjus.2022.101988>.
- Hartmann, K. and G. Wenzelburger (2021), 'Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA', *Policy Sciences*, **54**: 269–287.
- Horrich, T. and C. Witko (2024), 'Self-interest and preferences for the regulation of artificial intelligence', *Journal of Information Technology & Politics*, 1–16.
- Hemesath, S. and M. Tepe (2023), 'Framing the approval to test self-driving cars on public roads. the effect of safety and competitiveness on citizens' agreement', *Technology in Society*, **72**: 102177.
- Hemesath, S. and M. Tepe (2024a), 'Multidimensional preference for technology risk regulation: the role of political beliefs, technology attitudes, and national innovation cultures', *Regulation & Governance*, **18**(4): 1264–1283.
- Hemesath, S. and M. Tepe (2024b), 'Public value positions and design preferences toward AI-based chatbots in e-government. Evidence from a conjoint experiment with citizens and municipal front desk officers', *Government Information Quarterly*, **41**: 101985.
- Hortal, A. (2023), 'Evidence-based policies, nudge theory and Nancy Cartwright: a search for causal principles', *Behavioural Public Policy*, **7**: 333–352.
- Horvath, L., O. James, S. Banducci and A. Beduschi (2023), 'Citizens' acceptance of artificial intelligence in public services: evidence from a conjoint experiment about processing permit applications', *Government Information Quarterly*, **40**: 101876.
- Kahan, D. M., E. Peters, E. Cantrell Dawson and P. Slovic (2017), 'Motivated numeracy and enlightened self-government', *Behavioural Public Policy*, **1**: 54–86.
- Kennedy, R. P., P. D. Waggoner and M. M. Ward (2022), 'Trust in public policy algorithms', *The Journal of Politics*, **84**: 1132–1148.
- Kleizen, B., W. Van Dooren, K. Verhoest and E. Tan (2023), 'Do citizens trust trustworthy artificial intelligence? Experimental evidence on the limits of ethical AI measures in government', *Government Information Quarterly*, **40**(4): 101834.
- König, P. D., J. Felfeli, A. Achtziger and G. Wenzelburger, (2024), 'The importance of effectiveness versus transparency and stakeholder involvement in citizens' perception of public sector algorithms', *Public Management Review*, **26**(4): 1061–1082.
- König, P. D. and G. Wenzelburger (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.
- König, P. D., S. Wurster and M. B. Siewert (2023), 'Sustainability challenges of artificial intelligence and citizens' regulatory preferences', *Government Information Quarterly*, **40**: 101863.

- Krafft, T. D., K. A. Zweig and P. D. König (2022), 'How to regulate algorithmic decision-making: a framework of regulatory requirements for different applications', *Regulation & Governance*, **16**: 119–136.
- Kunda, Z. (1990), 'The case for motivated reasoning', *Psychological Bulletin*, **108**: 480.
- Laux, J., S. Wachter and B. Mittelstadt (2024), 'Trustworthy artificial intelligence and the European Union AI act: on the conflation of trustworthiness and acceptability of risk', *Regulation & Governance*, **18**: 3–32.
- Lecheler, S., C. de Vreese and R. Slothuus (2009), 'Issue Importance as a Moderator of Framing Effects', *Communication Research*, **36**(3): 400–425. <https://doi.org/10.1177/0093650209333028>.
- Leeper, T. J. and R. Slothuus (2014), 'Political parties, motivated reasoning, and public opinion formation', *Political Psychology*, **35**: 129–156.
- Lipsky, M. (2010), *Street-level bureaucracy: Dilemmas of the individual in public service*. Russell sage foundation.
- Little, A. T. (2024), 'How to distinguish motivated reasoning from Bayesian updating', *Political Behavior*, forthcoming. <https://doi.org/10.31219/osf.io/b8tvk>.
- Liu, P., R. Yang and Z. Xu (2019), 'How safe is safe enough for self-driving vehicles?', *Risk analysis*, **39**(2): 315–325.
- 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.
- Martin, B., O. James, S. Serritzlew and G. G. Van Ryzin (2020), 'Citizens' motivated reasoning about public performance: experimental findings from the US and Denmark', *International Public Management Journal*, **23**: 186–204.
- Martínez-Córcoles, M., M. Teichmann and M. Murdvee (2017), 'Assessing technophobia and technophilia: development and validation of a questionnaire', *Technology in Society*, **51**: 183–188.
- Meijer, A., L. Lorenz and M. Wessels (2021), 'Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems', *Public Administration Review*, **81**(5): 837–846.
- Meijer, A. and M. Wessels (2019), 'Predictive policing: Review of benefits and drawbacks', *International journal of public administration*, **42**(12): 1031–1039.
- Miller, S. M. and L. R. Keiser (2020), 'Representative bureaucracy and attitudes toward automated decision making', *Journal of Public Administration Research and Theory*, **31**: 150–165.
- Mintz, A., N. A. Valentino and C. Wayne (2021), *Beyond Rationality: Behavioral Political Science in the 21st Century*, Cambridge: Cambridge University Press.
- Oliver, A. (2013), *Behavioural Public Policy*, Cambridge: Cambridge University Press.
- Prokop, C. and M. Tepe (2022), 'Talk or type? The effect of digital interfaces on citizens' satisfaction with standardized public services', *Public Administration*, **100**: 427–443.
- Redlawsk, D. P. (2002), 'Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making', *The Journal of Politics*, **64**: 1021–1044.
- Schepman, A. and P. Rodway (2020), 'Initial validation of the general attitudes towards artificial intelligence scale', *Computers in Human Behavior Reports*, **1**: 100014.
- Schepman, A. and P. Rodway (2023), 'The General Attitudes Towards Artificial Intelligence Scale (GA AIS): confirmatory validation and associations with personality, corporate distrust, and general trust', *International Journal of Human-Computer Interaction*, **39**: 2724–2741.
- Schiff, D. S., K. Jackson Schiff and P. Pierson (2022), 'Assessing public value failure in government adoption of artificial intelligence', *Public Administration*, **100**: 653–673.
- Sunstein, C. R. (2002), *Risk and Reason: Safety, Law, and the Environment*, Cambridge: Cambridge University Press.
- Taber, C. S. and M. Lodge (2006), 'Motivated skepticism in the evaluation of political beliefs', *American Journal of Political Science*, **50**: 755–769.
- Tyler, T. R. (1990), *Why people obey the law*. Yale University Press.
- Wenzelburger, G., P. D. König, J. Felfeli and A. Achtziger (2024), 'Algorithms in the public sector. Why context matters', *Public Administration*, **102**: 40–60.

---

**Cite this article:** Hemesath S and Tepe M (2025), 'Motivated reasoning about artificial intelligence in public policy: comparative evidence from Germany', *Behavioural Public Policy*, 1–21. <https://doi.org/10.1017/bpp.2025.2>