How Ends Motivate Means: Unpacking How Motivated Reasoning Shapes Legitimacy Judgments Of Automated Travel Authorisation Systems*

David Karpa[†1] and Daria Gritsenko[1]

[1]Technical University of Munich

November 2026

Abstract

How do citizens evaluate the legitimacy of automated decision-making (ADM) in public administration? Existing research highlights the importance of procedural fairness, performance, and institutional design, often assuming that citizens assess these features in a neutral and consistent way. Drawing on theories of motivated reasoning from political psychology and behavioral public administration, we challenge this assumption by arguing that outcome favorability systematically shapes how citizens attend to and evaluate procedural information when evaluating the legitimacy of ADM systems. We test this argument using a preregistered survey experiment conducted in Finland (N = 2,040) that combines an outcome-based vignette with a conjoint experiment manipulating key procedural features of an ADM system, including human oversight, accountability, cost, and data practices, in the context of electronic travel authorization. We show that citizens' evaluations of these procedural features depend systematically on whether the ADM system produces politically preferred outcomes. We theorize and empirically assess two mechanisms underlying this pattern: selective attention to procedural information and biased evaluation of procedural safeguards. Evidence from both the conjoint estimates and open-ended survey responses indicates that outcome favorability activates motivated reasoning, shaping which procedural features citizens attend to and how they weight them. The observed effects are concentrated among ideologically committed respondents, even in the low-polarization context of Finland. Rather than treating biased responses to ADM as isolated cognitive errors, we conceptualize them as part of a broader motivated reasoning process that conditions legitimacy judgments of administrative authority. The study integrates institutional and cognitive perspectives on legitimacy and highlights why procedural improvements alone may be insufficient to secure public acceptance of ADM in public administration.

Keywords: legitimacy, automated decision-making, motivated reasoning, public administration, conjoint experiment

*This research was supported by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (Grant agreement No. 101116772, project AGAPP – Algorithmic Governance: A Public Perspective). Prior versions of this paper were presented at the Helsinki Conference on Emotions, Populism and Polarization in 2025 and the Biennial Conference of the ECPR Standing Group on Regulatory Governance in 2025. The authors thank Madalina Busuioc and Martino Maggetti for their helpful comments and suggestions. We also thank Vaibhav Agarwal and Shobhana Sharma for excellent research assistance.

[†]corresponding author: david.karpa@tum.de

1 Introduction

How do citizens evaluate the legitimacy of public administration? Information about procedures, performance, and institutional design form the core of standard accounts of legitimacy. Citizens are assumed to weigh how fairly decisions are made, how effectively institutions perform, and how appropriately they are structured (Tyler 2006; Grimmelikhuijsen & Meijer 2022). However, a large body of political science and behavioral public policy research shows that political evaluations rarely emerge from neutral or rational information processing (Taber & Lodge 2006; Grimmelikhuijsen et al. 2016; Alon-Barkat & Busuioc 2023). Instead, individuals frequently interpret political information and evaluate performance of public administration in ways that support their political attitudes, engaging in so-called motivated reasoning (Baekgaard & Serritzlew 2015; James & Van Ryzin 2017; Brogaard et al. 2024; Strömbäck et al. 2024; Damgaard & James 2024). It is currently unclear whether legitimacy judgments about public administration are also shaped by motivated reasoning – including the applied case of this study: the legitimacy of automated decision-making (ADM) in public administration. This study addresses this gap by examining how political attitudes influence the information citizens recruit and how they evaluate key features of automated decision-making when forming legitimacy judgments.

Automated decision-making systems are evaluated by citizens through multiple, overlapping dimensions of legitimacy, and research shows that outcomes, processes, and institutional context jointly structure these judgments. Outcome favorability has emerged as a particularly powerful determinant of legitimacy: citizens consistently view ADM as more legitimate when the results it produces align with their moral values, distributive preferences, or group interests, often outweighing formal considerations of fairness or accuracy (Christensen et al. 2020; Skitka et al. 2009; Ford 2012; Schreck et al. 2025). At the same time, procedural considerations – such as transparency, explainability, discretion, human oversight, and opportunities for appeal – remain central anchors for judgments about fairness and due process, especially in high-stakes or morally charged domains (Tyler 2006; Grimmelikhuijsen &

Meijer 2022; Grimmelikhuijsen 2023; Busch 2023). Citizens frequently express discomfort with fully automated processes and prefer hybrid arrangements, even when human decision-makers are objectively less accurate, suggesting that perceived procedural justice is partly tied to normative expectations about how public authority ought to be exercised (Martin & Waldman 2022; Yalcin et al. 2023). Finally, legitimacy evaluations are shaped by the broader institutional and contextual setting within which ADM is embedded: trust in the deploying authority, domain-specific expectations, issue salience, and the perceived deservingness of target groups all condition how citizens interpret both processes and outcomes (Schiff et al. 2025; Wenzelburger et al. 2024; Haesevoets et al. 2024; Lünich & Kieslich 2024). These three pillars – outcomes, processes, and institutional context – do not operate independently. Rather, they interact to create a complex evaluative environment in which outcomes are interpreted in light of features of the decision-making process and the broader institutional context (Sunshine & Tyler 2003; Helberger et al. 2020; Martin & Waldman 2023). As a result, legitimacy judgments of ADM are currently understood as the product of what the system does, how it does it, and where it is situated.

Our applied case is the European Travel Information and Authorisation System (ETIAS), a forthcoming EU-wide automated travel authorisation that will require visa-exempt nationals to apply online before short stays in the Schengen area; applications are screened against EU security and migration databases and flagged cases referred to national ETIAS units for manual review.[1] The official EU description states that "some 1.4 billion people from over 60 visa-exempt countries [...] are required to have a travel authorisation to enter most European countries," and the European Council confirmed in March 2025 that ETIAS "will start operations in the last quarter of 2026."[2] ETIAS is therefore an actual large-scale automated decision-making system in public administration, not a hypothetical scenario — yet data collection in October 2024 preceded rollout and any sustained public debate, so Finnish respondents could not draw on direct experience or settled media framing. The study thus

[1] https://travel-europe.europa.eu/etias_en
[2] https://travel-europe.europa.eu/etias/about-etias/who-should-apply_en

captures legitimacy formation for a real ADM system ahead of deployment.

Despite extensive work on how citizens evaluate administrative performance, it remains unclear how motivated reasoning enters citizens' legitimacy judgments. Behavioral public administration research has long shown that citizens process administrative information in biased and preference-consistent ways, suggesting a plausible link between motivated reasoning and legitimacy assessments (Grimmelikhuijsen et al. 2016). Recent work on citizen–algorithm encounters further underscores that cognitive biases shape how individuals interpret and respond to automated decision-making, indicating that legitimacy evaluations of ADM may likewise be filtered through motivated reasoning (Alon-Barkat & Busuioc 2024). Motivated reasoning is widely documented in political psychology, where individuals selectively recruit, evaluate, and recall information in ways that are consistent with their prior attitudes or desired conclusions (Kunda 1990; Taber & Lodge 2006; Epley & Gilovich 2016). This mechanism has been observed across diverse domains, including political polarization on social media (Su 2022), climate change beliefs (Druckman & McGrath 2019; Bayes & Druckman 2021; Jost et al. 2022), health care politics (Strickland et al. 2011), and broader patterns of public opinion (Leeper & Slothuus 2014). Parallel findings in behavioral public administration demonstrate that citizens evaluate the performance, fairness, and trustworthiness of public agencies in preference-consistent ways (Baekgaard et al. 2019; James & Van Ryzin 2017; Alon-Barkat & Busuioc 2023; Strömbäck et al. 2024). Taken together, this research identifies motivated reasoning as a central mechanism shaping evaluations of public administration, yet its role remains underexamined in the context of *legitimacy* judgments of ADM.

We argue that motivated reasoning enters legitimacy judgments when citizens use outcome favorability as a motivational cue for interpreting the procedural features of ADM systems. Outcome favorability – already a well-established determinant of legitimacy – can thus additionally affect how citizens judge identical procedures as more or less legitimate depending on whether the outcome aligns with their preferences. Applied to ADM, this

implies that citizens may evaluate the same decision rules differently depending on whether the system produces (dis)preferred outcomes, selectively attending to and weighing procedural information in ways consistent with their preferences. We test this proposition using a preregistered survey experiment in Finland (N = 2,040) that links a vignette manipulation of system defaults for asylum seekers to a conjoint experiment on procedural attributes of EU travel authorization. This design allows us to identify whether respondents' evaluations of key ADM features reflect motivated reasoning rather than stable preferences, and open-ended responses provide additional evidence of selective recruitment and justification. Across both components, we find that citizens systematically downplay or discount procedural safeguards when outcomes conflict with their political preferences, and highlight them when outcomes align with those preferences. These patterns show that legitimacy judgments of ADM systems are shaped not only by their design but by preference-consistent information processing triggered by outcome favorability. Hence, we identify motivated reasoning as a core mechanism through which legitimacy judgments are formed.

This study makes four contributions to research on administrative legitimacy and automated decision-making. First, we show that outcome favorability – apart from being an independent driver of perceived legitimacy – interacts with throughput legitimacy by moderating how citizens interpret procedural features (Sunshine & Tyler 2003; Tyler 2006; Schmidt 2013; Doherty & Wolak 2012; Arnesen 2017; Esaiasson et al. 2019; Grimmelikhuijsen & Meijer 2022). Second, we demonstrate that all major procedural attributes of ADM systems are susceptible to biased attention and selective weighting, providing an attribute-level account of when and how citizens reinterpret institutional design features (Helberger et al. 2020; Ingrams et al. 2021; Martin & Waldman 2022; Grimmelikhuijsen & Meijer 2022; Yalcin et al. 2023; Horvath et al. 2023; Lünich & Kieslich 2024; Haesevoets et al. 2024; Wenzelburger et al. 2024; Araujo et al. 2020; Schiff et al. 2025). Third, by documenting motivated reasoning in a low-polarization, consensus-oriented context, we clarify that the mechanism operates robustly across populations but is activated only under strong prior preferences

within them, thereby refining boundary conditions for motivated reasoning in public administration (Kunda 1990; Redlawsk 2002; Taber & Lodge 2006; Slothuus & De Vreese 2010; Kahan 2016; Leeper & Slothuus 2014; Baekgaard et al. 2019; Már & Gastil 2020; Boyer et al. 2022). Fourth, we study legitimacy formation for an actual forthcoming ADM system in public administration ahead of its rollout in the last quarter of 2026.[3] Capturing citizen attitudes before deployment and before sustained public debate avoids the contamination of priors by direct experience or media coverage of a live system, and yields evidence relevant to policy-makers. In sum, our findings speak to broader debates on how citizens make sense of automated decision-making and, more generally, on the cognitive foundations of legitimacy in contemporary governance.

We proceed as follows. Section 2 reviews the literature and develops hypotheses. Section 3 describes the research design and data. Section 4 presents the results. Section 5 discusses their implications, and Section 6 concludes.

2 Theory and Hypotheses

2.1 Legitimacy: Process, output, context

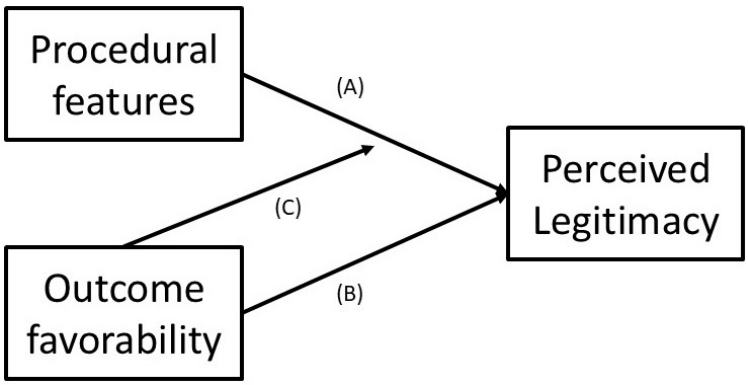
Legitimacy fundamentally refers to the belief that an authority or system is entitled to be obeyed and deferred to, not merely because of coercion or performance, but because it is seen as appropriate, proper, and just (Sunshine & Tyler 2003; Tyler 2006). Legitimacy in contemporary governance does not rest primarily on majoritarian control, but on whether institutions follow robust procedural safeguards and confine their authority to technically complex, non-redistributive tasks (Majone 1997). Hence, institutions are seen as legitimate when their decisions are grounded in expertise, transparency, and clear accountability mechanisms. Moreover, their acceptance depends not only on inputs and outputs but also on the quality of internal decision-making processes, including efficacy, openness, and procedural fairness (Schmidt 2013). In sum, the extant literature theorizes three sources of perceived

[3] https://travel-europe.europa.eu/etias/about-etias/who-should-apply_en

legitimacy: input, throughput, and output. These dimensions of legitimacy are analytically separable yet empirically entangled – arguments ostensibly about outputs are often also about the throughput processes that produced them. We follow Schmidt (2013) and use the distinction out of necessity for analytic purposes while theorizing how the dimensions interact.

The notion that fair procedures shape acceptance in their own right, beyond the outcomes they produce, traces back to early procedural-justice research showing that people are more willing to accept a decision when they feel they had control over how their case was presented (Thibaut & Walker 1978). Failures in these "throughput" processes – such as opacity, bias, or incompetence – can undermine public perceptions of legitimacy even when policy performance or formal participatory mechanisms appear strong. Figure 1 depicts a simple causal model of legitimacy judgments in the spirit of Esaiasson et al. (2019). The corresponding link to procedural aspects of legitimacy judgments is (A).

Figure 1: Conceptual model



Besides the importance of procedural goodness, outcomes also play a decisive role in shaping citizens' legitimacy judgments (Scharpf 1999) (link (B) in Figure 1). A large body of research shows that citizens evaluate elections more positively when the outcome aligns with their preferences, whereas unfavorable outcomes sharply reduce perceived legitimacy (Anderson 2005; Pierce et al. 2016; Dahlberg & Linde 2016; Arnesen 2017; Esaiasson et al. 2023). Christensen et al. (2020) show that citizens evaluated parliamentary decisions on

euthanasia as more legitimate when the policy outcomes aligned with their own preferences, regardless of the deliberative process used. Their study demonstrates that in moral and ethically sensitive domains, outcome favorability can outweigh procedural fairness as a basis for legitimacy judgments. Similarly, legitimacy decreases when individuals' moral or religious convictions conflict with judicial decisions, even when those decisions are made through highly authoritative procedures (Skitka et al. 2009). Comparable patterns are found in international governance, where citizens perceive institutions as more legitimate when decisions align with their group-based preferences, despite the intention of these bodies to operate impartially (Ford 2012). Outcome considerations have likewise been shown to shape evaluations of administrative and organizational legitimacy (Schreck et al. 2025), echoing organizational justice literature showing that the favorability of an outcome strongly conditions how people respond to a decision (Brockner & Wiesenfeld 1996). Indeed, outcome favorability has been argued to constitute one of the most decisive determinants of legitimacy judgments across a wide range of institutional settings (Esaiasson et al. 2019).

Yet the role of outcomes extends beyond simple winner–loser effects (Anderson 2005): outcome favorability can also shape how citizens interpret procedural qualities themselves. A growing literature shows that people often judge the fairness of procedures in light of whether these procedures deliver their preferred results (link (C) in Figure 1). This interdependence has long been documented in organizational justice, where procedural fairness and outcome favorability do not act in isolation but interact: fair procedures matter most for how people react precisely when outcomes are unfavorable (Brockner & Wiesenfeld 1996, 2013). One reason is that unfair procedures lead people to hold the deciding authority responsible for an unfavorable outcome, whereas fair procedures make the same outcome easier to accept as justified (Folger & Cropanzano 1998; Brockner et al. 2007). We turn this relationship around: where fair procedures can shape how citizens respond to an outcome (Brockner & Wiesenfeld 1996; Brockner et al. 2007), we ask whether the favorability of an outcome instead shapes how citizens perceive the very same procedure. People are known to attend more closely

to how a decision was made once its outcome matters to them; our argument is that this attention is not neutral but motivated, so that citizens may read identical procedural features as more or less legitimate depending on whether the outcome matches their preferences. When procedural fairness is ambiguous, citizens rely more heavily on their prior attitudes, evaluating the same process as more or less fair depending on whether it produces a desirable policy (Doherty & Wolak 2012). Similarly, experimental evidence shows that favorable outcomes substantially increase acceptance of decisions, whereas having influence over the process itself contributes comparatively little to perceived legitimacy (Arnesen 2017). Taken together, this research suggests that outcome favorability not only directly affects legitimacy judgments but also colors citizens' views of procedural fairness, reinforcing the conclusion that ends can shape how the means are perceived (Esaiasson et al. 2019). We build on this idea and explore not only the direct influence of throughput (link (A)) and output (link (B)), but also how outcome favorability influences perceptions of procedural legitimacy indirectly (link (C)).

2.2 ADM in public administration

Work on public perceptions of ADM in public administration shows that citizens form legitimacy judgments by drawing on information about decision *processes*, *outcomes*, and how these interact. First, consistent with link (A) in Figure 1, citizens place substantial weight on procedural features. They generally prefer human-in-the-loop arrangements and view fully automated systems as less legitimate across multiple dimensions of process quality, including transparency, fairness, discretion, and opportunities for appeal (Grimmelikhuijsen & Meijer 2022; Martin & Waldman 2022; Ingrams et al. 2021; Kern et al. 2022; Starke & Lünich 2020; Hillo et al. 2025). Human involvement is especially valued in emotionally charged or consequential cases (Yalcin et al. 2023). Even when ADM reduces red tape or avoids human error, its procedural profile tends to be judged less trustworthy (Ingrams et al. 2021, p. 390–391). This sensitivity to procedural aspects is echoed in public-sector recruitment, where disclosure of AI use triggers algorithm aversion and reduces perceived fairness de-

spite identical substantive evaluations (Keppeler 2024). Citizens also show appreciation of procedural safeguards, such as external monitoring and public sector control over algorithm development (Aula et al. 2026). A second body of work examines the data-intensity of ADM systems: citizens differentiate between systems on the basis of what personal information is collected, with more intrusive regimes (biometrics, health data) eroding perceived legitimacy even when accuracy is held constant (Helberger et al. 2020). Data protection is also related to the area of implementation, with citizens' concerns exacerbated by law enforcement context vs distributional services (Zheng et al. 2026). A third procedural feature is the compliance cost imposed on applicants, which the administrative-burden literature treats as a non-trivial determinant of citizen-state interactions (Moynihan et al. 2015).

Second, research highlights the importance of *outcomes* for acceptance, consistent with link (B). ADM becomes more legitimate when it produces outcomes citizens find substantively desirable, whether because a domain is framed as technical rather than ideological or because algorithmic decisions align with group-based deservingness judgments (Haesevoets et al. 2024; Lünich & Kieslich 2024). Some citizens view ADM as more neutral or fair than human officials, especially when they expect algorithms to counteract political bias (Helberger et al. 2020). Others support ADM because they believe it yields distributive or moral outcomes closer to their own preferences (Araujo et al. 2020). These findings echo broader work showing that outcome favorability can outweigh abstract considerations of procedural quality.

A large body of research in public administration has shown that contextual factors shape how citizens evaluate the procedural quality of ADM systems, with perceptions of fairness varying substantially across decision domains, institutional settings, implementing organisation, and issue characteristics (Horvath et al. 2023; Haesevoets et al. 2024; Wenzelburger et al. 2024; Lünich & Kieslich 2024; Schiff et al. 2025). Research further shows that citizens' evaluations depend on the institutional and social context: for example, Black respondents express greater support for automated enforcement when human decision-makers

are perceived as unrepresentative or biased (Miller & Keiser 2021); conversely, racially minoritized citizens do not have clear-cut preferences between AI and out-group bureaucrats (Gaozhao et al. 2024). Citizens generally continue to hold the government responsible when an automated system produces unfair decisions rather than a human, but they assign less responsibility when the system was developed by a private company (Alon-Barkat et al. 2025). This literature typically conceptualizes *context* as the primary moderator of process evaluations, examining how features such as decision importance, institutional trust, or issue salience condition citizens' judgments of procedural fairness (Martin & Waldman 2022; Schiff et al. 2025; Lünich & Kieslich 2024).

Legitimacy research acknowledges that outcome favorability can sometimes influence perceptions of procedural justice, with citizens judging the same procedures differently depending on whether they approve of the resulting decision (Doherty & Wolak 2012; Arnesen 2017; Esaiasson et al. 2019). Our study extends this literature by showing how outcome favorability moderates the way citizens evaluate the fairness and quality of ADM processes (link (C)). In doing so, we link recent legitimacy insights to the study of ADM and highlight motivated reasoning as a mechanism through which outcomes shape perceptions of procedural legitimacy. Experimental evidence shows that people sometimes defer excessively to algorithmic recommendations or selectively follow them when these reinforce existing stereotypes, revealing systematic biases in human–AI decision environments (Alon-Barkat & Busuioc 2023). These biases indicate that ADM-related judgments are shaped not only by contextual factors but also by cognitive shortcuts, providing a foundation for our argument that legitimacy evaluations themselves may be filtered through motivated reasoning. We now turn to reviewing the literature on motivated reasoning and deriving testable hypotheses about the mechanisms through which legitimacy judgments are affected.

2.3 Motivated Reasoning

Motivated reasoning refers to the tendency for individuals to process information in ways that serve directional goals (arriving at a preferred conclusion), rather than purely accuracy

goals of objective evaluation (Kunda 1990; Taber & Lodge 2006). In practice, this means that judgments are often driven by the need to protect prior attitudes, partisan attachments, or group identities (Redlawsk 2002; Kahan 2016; Leeper & Slothuus 2014). Theoretically, motivated reasoning operates as a form of heuristic processing, where individuals rely on cognitive shortcuts to navigate search and choice to arrive at a particular conclusion (Kunda 1990). Three distinct cognitive shortcuts enable effort reduction: a prior belief effect, confirmation bias, and disconfirmation bias (Taber & Lodge 2006; Strickland et al. 2011). First, the prior belief effect facilitates attribute substitution by anchoring judgments in existing attitudes. Second, confirmation bias – the uncritical acceptance of congruent information – lowers cognitive load via satisficing, where individuals restrict information search to easily accessible, congruent cues. Finally, disconfirmation bias – the heightened scrutiny and rejection of incongruent information – activates selective hyper-scrutiny, raising the evidentiary bar exclusively for incongruent data to disqualify it before it urges a belief update.

While we cannot directly observe motivated reasoning, it can be mapped out as a product of two main mechanisms that guide the reasoning process motivated to form coherence between preferences and beliefs: (1) *recruiting* and (2) *evaluating* information (Epley & Gilovich 2016). The first mechanism concerns how individuals *recruit* or select information when forming judgments. Subject to the prior belief effect, motivated reasoners attend more to cues that are consistent with their existing political attitudes and dismiss – or simply ignore – incongruent evidence (Slothuus & De Vreese 2010; Boyer et al. 2022).

Hypothesis 1 (Motivated selection of attributes): Respondents selectively attend to information about procedural features contingent on outcome favorability. In technical terms, this translates into attribute (non-)attendance, our estimand introduced in detail below in Section 3.

The second mechanism involves how individuals *evaluate* the available information. Rather than assigning equal weight to competing considerations, citizens selectively emphasize features that support their preferred stance while discounting opposing attributes (Baekgaard

et al. 2019; Slothuus & De Vreese 2010). At the *evaluation* stage, confirmation and disconfirmation biases drive motivated skepticism: congruent arguments are judged stronger, incongruent ones are scrutinized and rejected (Taber & Lodge 2006). Identity-protective cognition formalizes this as weighting the same evidence by its implications for in-group commitments (Kahan 2016; Leeper & Slothuus 2014). In the context of ADM, individuals may highlight costs or data-intensity of systems only when these attributes can be aligned with their preferred policy position, downplaying them otherwise. **Hypothesis 2 (Selective weighting of attributes):** Respondents evaluate procedural features differently contingent on outcome favorability.

Figure 2: Motivated reasoning and hypotheses

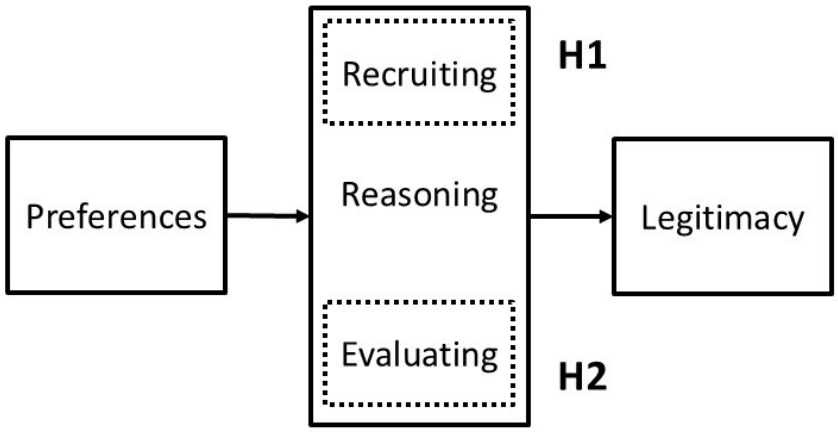


Figure 2 illustrates the process of motivated reasoning. Preferences shape cognition by guiding how individuals recruit and evaluate information, which in turn form the legitimacy judgments they hold about automated systems. After the fact, individuals engage in selective recall to justify their choices, reinforcing legitimacy judgments that are congruent with their prior preferences. Together, motivated reasoning operates across three stages of recruiting, evaluating, and recalling, linking preferences to perceptions of legitimacy.

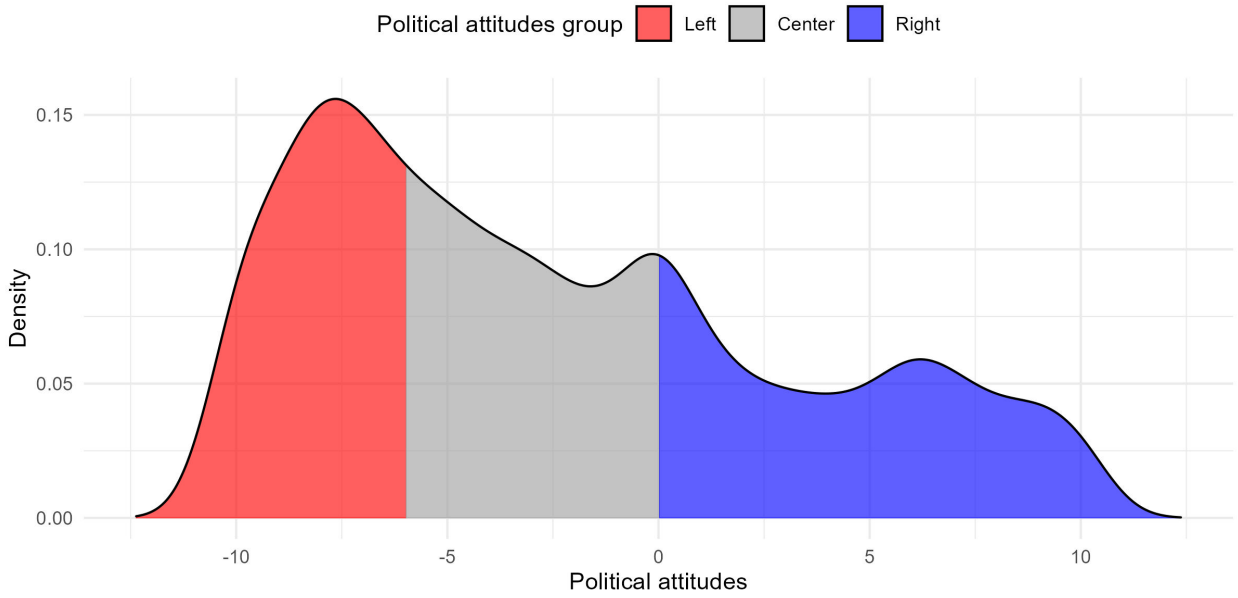
3 Research design

We rely on a preregistered survey experiment conducted in October 2024 by the University of Helsinki Citizen Barometer (*Kansalaisbarometri*).[4] The Citizen Barometer is a non-commercial academic online panel run by the University of Helsinki and the Helsinki Institute for Social Sciences and Humanities (HSSH); no commercial survey vendor was used. The sample (N = 2,040 Finnish citizens aged 18 and older) was drawn from the panel's opt-in subpanel, and no sampling quotas were imposed at fielding. Respondents provided information about their sociodemographic background, knowledge of automated travel systems, and general attitudes towards automation. Representativeness weights were applied to align the sample with population benchmarks on language, region, age, gender, and education. We also collected political attitudes and voting behavior. Political attitudes were operationalized as the distance between feeling thermometer scores for political parties, measured on a scale from 0 (*Strongly dislike*) to 10 (*Strongly like*). This measure is commonly used to identify political identification in contemporary Western societies characterized by polarization (Iyengar & Westwood 2015; Iyengar et al. 2019; Druckman & Levendusky 2019). Political attitudes, in turn, is the most commonly used measure to study motivated reasoning (Bolsen et al. 2014; Druckman et al. 2018; Dias & Lelkes 2022).[5] For the purposes of analysis, we focused on the two most polarized representative parties: Perussuomalaiset (right-wing) and Vasemmistoliitto (left-wing), following Stoetzer et al. (2023). Respondents were categorized into tertiles based on the absolute distance in their thermometer scores, corresponding to the lower (*Left*), middle (*Center*), and upper (*Right*) thirds of the distribution (Figure 3).

[4] https://aspredicted.org/48v2-9qk5.pdf
https://www.helsinki.fi/en/projects/kansalaisbarometri

[5] In multiparty systems, as Finland is, political attitudes are not about being in favor of *the* left/right party, but *a* left/right party.

Figure 3: Distribution of political attitudes



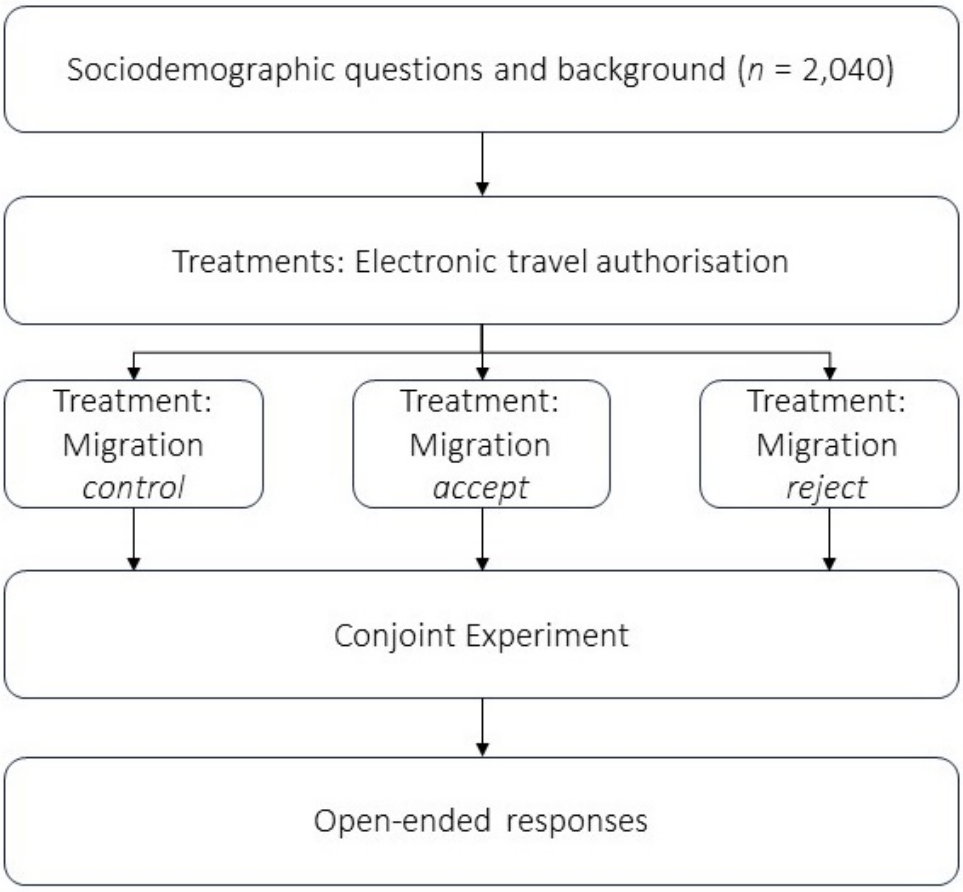
Note: This Figure displays the distribution of political attitudes, operationalized as the absolute difference in feeling thermometer scores (0 to 10) between respondents' ratings of Perussuomalaiset (right-wing party) and Vasemmistoliitto (left-wing party). Higher values indicate stronger political identification.

The survey concluded with an experimental setup in which respondents were randomly assigned to one of three conditions. Table A3 in the appendix shows that assignment to the three groups was balanced respective political attitudes, age, and gender. In the *control* condition, no default was mentioned. In the *accept* condition, respondents were told that the default system setting is to accept all asylum seekers. In the *reject* condition, they were told that the default system setting is to reject all asylum seekers. These conditions link system outcomes to outcome favorability by giving each political group a preferred default outcome of the system. In the spirit of "refugees welcome", the *accept* condition poses a preferable outcome for those on the left, whereas those on the *right* much prefer closed borders in the form of the *reject* condition. We are able to assume a strong link between general political attitudes and attitudes regarding migration as previous research highlights the connection (Dias & Lelkes 2022). We discuss the case selection further below in section 3.1.

Following this manipulation, respondents completed a conjoint experiment on automated

border-control systems (Hainmueller et al. 2014). Each system was described by five attributes with three levels each (Table A2). In each task, respondents compared two systems, indicated which they preferred, and rated both systems on a five-point scale. Each respondent completed three tasks, producing six system evaluations in total. The flow of the experiment is shown in Figure 4, and a reconstruction of the Qualtrics screen as seen by respondents is provided in Appendix Figure A1. Since we conduct multiple comparisons – multiplying groups and treatment conditions – we correct our estimations for false discovery rates using the Benjamini-Hochberg procedure (Benjamini & Hochberg 1995). Predictions and comparisons are calculated using the marginaleffects R package (Arel-Bundock et al. 2024).

Figure 4: Experiment Flow



We operationalize our hypotheses by translating the two theoretical mechanisms in Figure

2 into two empirical quantities we can observe. The first mechanism, selective recruitment, concerns whether people even use a piece of information when forming a judgment. In the conjoint experiment, we measure this using *attribute non-attendance* (ANA), an established estimand in discrete-choice analysis that asks whether a given attribute makes any difference to respondents' choices within a specific subgroup and treatment context (Hensher & Greene 2010; Campbell et al. 2011), and where non-attendance is debated as reflecting a decision heuristic rather than a genuine preference (Heidenreich et al. 2018). To test this, we apply the Conditional Randomization Test (CRT) (Ham et al. 2024), which checks whether choices change when an attribute varies, holding all other randomized features constant. If the CRT rejects the null, the attribute is attended to; if it does not, the attribute is effectively ignored. Consistent patterns of non-attendance tied to political attitudes and to congruent versus incongruent system defaults provide evidence for our first motivated reasoning mechanism: selective recruitment (Hypothesis 1). In other words, we hypothesize that respondents may ignore some information if they (dis)like the system default setting.

The second mechanism, biased evaluation, concerns how much weight respondents give to the attributes they attend to. We measure this using marginal means contrasts by political attitudes and treatment condition (Leeper et al. 2020). In other words, we estimate whether stated preferences for attributes vary between systems with (not) liked migration policies. If respondents were motivated purely by accuracy, the effects of each attribute would look similar across groups and across experimental contexts. In other words, preferences for features such as cost, data collection, or human oversight would not depend on whether the system produced a preferred or dispreferred outcome. By contrast, we assess differences in stated preferences for system attributes because of (not) liked system outcomes as evidence of bias evaluation (Hypothesis 2). Taken together, attribute non-attendance tells us whether an attribute enters the decision rule at all, while marginal means show how much importance it receives once it is considered. Both mechanisms correspond to link (C) in Figure 1, where outcome favorability shapes how procedural information is processed. Unweighted replicates

of both the attribute-attendance and rating-contrast results are reported in Online Appendix B.2.2. This framework also provides a natural falsification test: if attributes are attended to and weighted consistently across treatments and political attitudes, then motivated reasoning would not explain the observed legitimacy judgments.

3.1 Boundary conditions and case selection

Kunda (1990) distinguished between accuracy goals, which aim for unbiased evaluation, and directional goals, which aim for preferred conclusions. Motivated reasoning arises especially in contested and identity-relevant policy domains, where individuals are likely to hold strong prior preferences. It is conditional, intensifying with stronger priors, salient identity stakes, and higher cognitive sophistication (Taber & Lodge 2006; Már & Gastil 2020). It also depends on identity and status dynamics, as groups interpret the same issue through divergent motivational lenses, producing asymmetric patterns of reasoning (Boyer et al. 2022). The information environment further shapes these processes, since additional evidence does not necessarily reduce bias and can even amplify prior-attitude effects under directional goals (Baekgaard et al. 2019). Migration in Finland illustrates such dynamics: debates on border closures with Russia,[6] heightened salience due to increased border pressure,[7] and stricter citizenship policies[8] create an environment in which issue salience is high. In this context, citizens' legitimacy judgments of automated travel systems provide a strong test of motivated reasoning, as directional priors regarding the outcomes of such systems are strongly determined and increasing by political attitude intensity. In other words, people at the ideological extremes bring stronger migration-related priors to the task than those located near the center. Our research design leverages this context.

The conjoint task asks respondents to compare two configurations of ETIAS along five attributes, each chosen to correspond to a theoretically motivated procedural dimension

[6] https://valtioneuvosto.fi/en/situation-at-finlands-eastern-border

[7] https://www.reuters.com/world/europe/finland-says-thousands-migrants-seek-enter-via-russia-2024-02-20/

[8] https://migri.fi/en/-/longer-period-of-residence-required-for-citizenship-as-of-1-october-2024

of legitimacy and to an actual design choice in the forthcoming system. The *Decision* attribute and the *Appeal* attribute operationalise the human-oversight and process-quality dimensions citizens weigh most heavily, generally preferring human-in-the-loop arrangements and opportunities for appeal (Grimmelikhuijsen & Meijer 2022; Martin & Waldman 2022; Ingrams et al. 2021; Yalcin et al. 2023). The *Data Collection* attribute reflects evidence that more intrusive data regimes erode perceived legitimacy even when accuracy is held constant (Helberger et al. 2020), and mirrors the explicit decision to exclude biometric and health data from ETIAS.[9] The *Cost* attribute follows the administrative-burden literature, which treats compliance costs as a non-trivial determinant of citizen-state interactions (Moynihan et al. 2015), and tracks the recent ETIAS fee revision from EUR 7 to EUR 20.[10] The *Responsibility* attribute draws on evidence that citizens hold the government responsible for unfair automated decisions and assign accountability differently depending on the actor behind the system (Alon-Barkat et al. 2025). Respondents are told the systems they evaluate are hypothetical variants of ETIAS. Ex-ante public awareness of ETIAS was limited in Finland during fielding in October 2024, so respondents could not rely on prior experience or media frames when forming legitimacy judgments which, arguably, made them more sensitive to the presented information.

4 Results

4.1 Procedural drivers

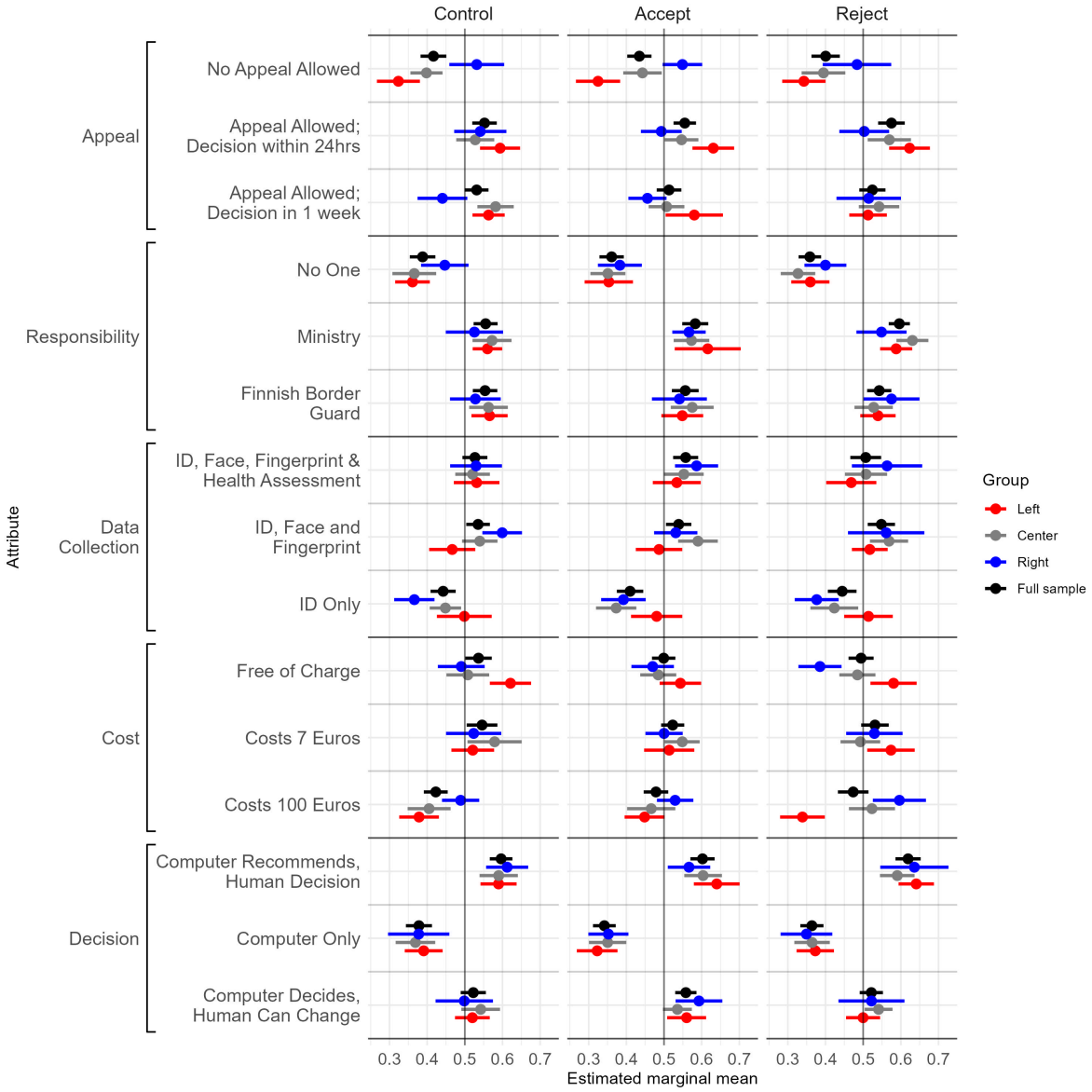
Preferences regarding procedural features of travel authorization are captured by the attribute evaluations in the conjoint experiment. To examine how these preferences vary both overall and across political orientations, we estimate marginal means (Leeper et al. 2020) separately for the full sample and for respondents on the left, in the political center, and on the right. For each of the three system outcomes (*Control*, *Accept*, and *Reject*), we compute

[9] https://travel-europe.europa.eu/etias/about-etias/news-corner/applying-for-etias-no-health-information-or-biomertics-needed_en

[10] https://travel-europe.europa.eu/etias/about-etias/news-corner/ETIAS-will-cost-EUR-20_en

predicted support for systems possessing specific attribute levels. Figure 5 presents these predictions. A deviation from 0.5 in the marginal mean indicates how much more (above 0.5) or less (below 0.5) likely respondents are to prefer a system with a given attribute level relative to being indifferent between the two options.

Figure 5: Predicted Attribute Preferences Across Treatment Conditions and Political Attitudes



Note: Predictions are based on multinomial logit models estimated via `cregg::mm()` separately for the full sample and for ideological subgroups (left, center, right). Estimates represent the marginal mean likelihood of preferring an average system with a given attribute level, weighted for population representativeness. Values above (below) 0.5 indicate positive (negative) evaluations relative to indifference. Confidence intervals represent 95% uncertainty. An unweighted replicate is reported in Online Appendix Figure B2. Full details are provided in Appendix B.1

To interpret these patterns, we examine the attributes one by one. We begin with the two attributes that display the highest degree of consensus across ideological groups: *Responsi-*

bility and *Decision*. For both attributes, the estimates for the left, center, right, and full sample are nearly identical, indicating broad agreement about their procedural importance. Respondents clearly prefer systems in which some identifiable public authority is responsible, with a slight preference for ministerial responsibility over frontline decision-makers, while the absence of any accountable actor sharply reduces support. Similarly, across all political groups, systems in which the computer makes the decision alone are strongly disliked, whereas systems in which the computer recommends and a human decides are consistently favored. These two attributes thus constitute core pillars of procedural legitimacy on which citizens across the political spectrum converge.

For the remaining attributes – *Appeal*, *Data Collection*, and *Cost* – the Figure reveals meaningful ideological differences. First, regarding *Appeal*, the absence of an appeal mechanism is, on average, strongly disliked, but this pattern is driven predominantly by respondents on the left and in the center. Right-leaning respondents appear comparatively indifferent to whether appeal rights are available, indicating that this procedural safeguard matters substantially more to those with left-of-center political orientations. Second, for *Data Collection*, the left displays relatively flat preferences across the three data regimes, suggesting that intensity of data collection does not meaningfully shape their evaluations. By contrast, the center and especially the right exhibit a pronounced dislike for the "ID only" level and higher support for systems collecting biometric or biometric-plus-health information. This pattern suggests that respondents with more right-leaning identities require a higher threshold of data collection for a system to appear legitimate. Finally, the *Cost* attribute shows the clearest divergence. Left-leaning respondents strongly prefer free or low-cost systems and react negatively to high fees, whereas right-leaning respondents hold almost the opposite view, expressing lower support for cheap systems and somewhat higher support for more costly ones. These countervailing preferences indicate that cost operates not simply as a burden but as a cue that different ideological groups interpret in distinct ways.

Taken together, these results show that while some procedural features generate broad

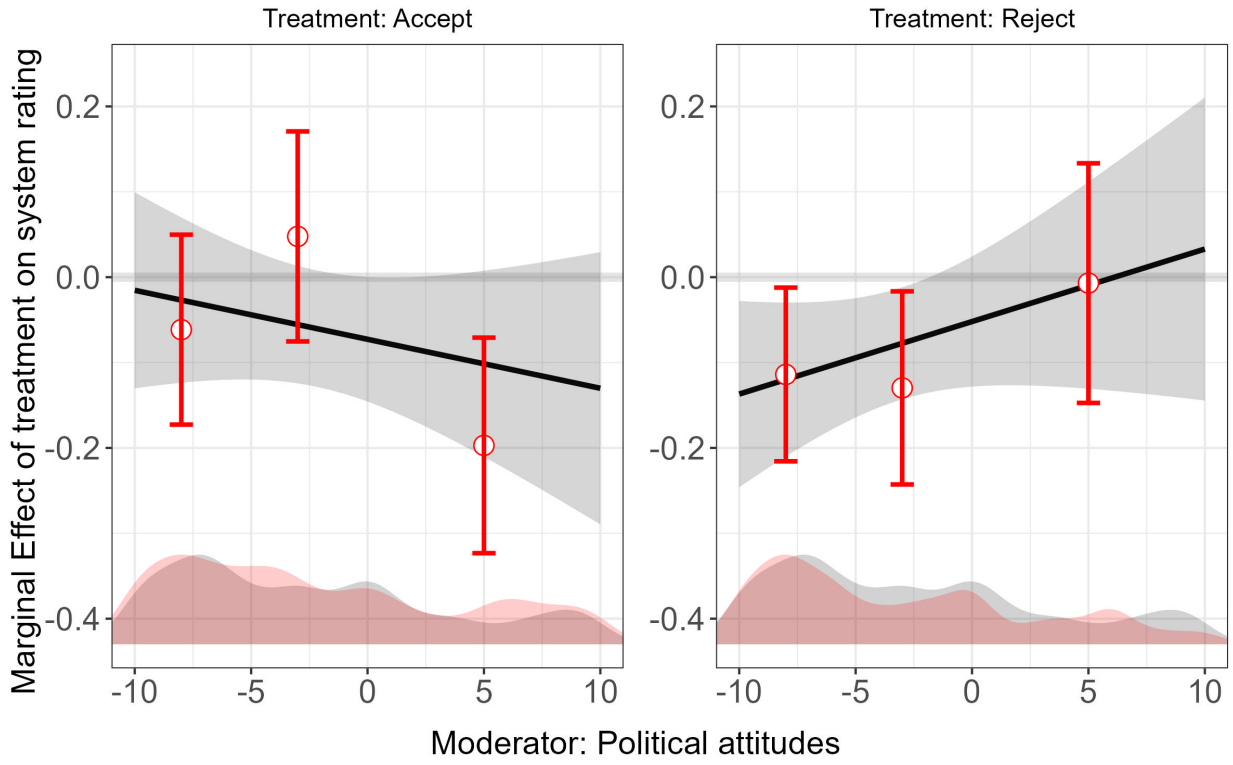
agreement, others become politically charged. Across attributes, the largest effects in marginal-mean predictions arise for decision-making and responsibility, with human involvement and clear accountability yielding strong legitimacy gains while full automation and the absence of a responsible authority produce the sharpest penalties. Appeal rights exert the next strongest influence, whereas cost and data collection generate more moderate shifts, indicating that human oversight, identifiable responsibility, and opportunities for appeal are the dominant procedural drivers of ADM legitimacy.

4.2 Outcome favorability

To assess how favorable or unfavorable outcomes influence legitimacy judgments, we now examine regression results that estimate how system ratings differ across treatment conditions, taking into account the political attitudes of respondents. In other words, we test whether systems are judged as more legitimate given a certain outcome for our three ideological groups (left, right, center). We use the system ratings here instead of choices between systems, because both systems always share the same outcome, by design. Figure 6 shows point estimates and confidence intervals of interactions terms (treatment $\times$ ideological group).

In the *Accept–Control* comparison (left-hand display), respondents on the political right rate systems about -0.2 points lower, whereas those on the left and in the center show no meaningful difference. In the *Reject–Control* comparison, respondents on the left and center rate systems lower, whereas the right is indifferent. These patterns indicate that legitimacy judgments are conditioned by preferred outcomes: systems are penalized when their defaults violate group preferences. Conversely, they do *not* receive a legitimacy dividend when they align with them. This confirms that ADM legitimacy is politically contested and further motivates our analysis of the underlying motivated reasoning mechanisms.

Figure 6: Conditional Average Treatment Effects by Political Attitudes



Note: This Figure presents the marginal effects of treatment conditions (accept and reject) on legitimacy ratings, conditional on levels of political attitudes. The attitudes are categorized into tertiles: left (≤33rd percentile), center (34th–66th percentile), and right (≥67th percentile), see Figure 3. Estimates are obtained using the `interflex` package in R, employing a binning estimator with covariates for system attributes, party preference variables, and demographic controls. The shaded areas represent 95% confidence intervals. Estimations via the `interflex` package in R (Hainmueller et al. 2019). Full details are provided in Appendix B.1

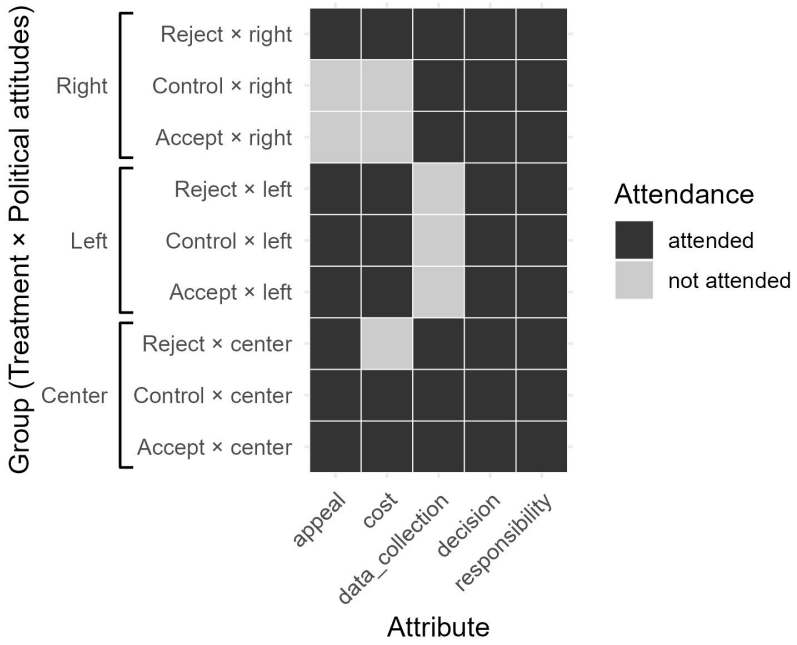
4.3 Motivated reasoning

4.3.1 Information recruitment

We now turn to evidence that is related to our hypotheses. First, we test hypothesis 1, which posits that respondents strategically divert attention to evidence according to their outcome preferences. Applied to the context of our conjoint experiment, we test whether all attributes are attended by respondents, or if there is (strategic) non-attendance. Figure 7 shows the results of these tests of attribute attendance across ideological groups and treatment conditions in a heatmap. The core institutional attributes *Responsibility* and *Decision*

are consistently attended to. By contrast, *Cost*, *Data Collection*, and *Appeal* show greater variability: the center tends to ignore *Cost* under the Reject condition, the left exhibits selective non-attendance to *Data Collection* regardless of treatment, and the right largely disregards *Appeal* and *Cost* under the Control and Accept conditions, with attention to these attributes reappearing under Reject. Overall, the heatmap suggests that while some features are universally salient, others are selectively attended depending on political attitude and system outcome.

Figure 7: Attribute attendance



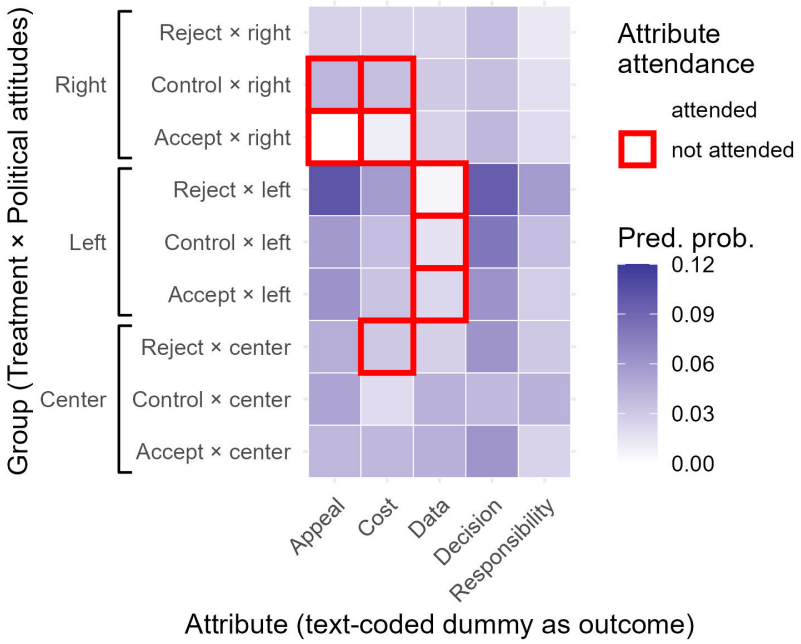
Attribute-specific attendance across experimental groups. Each row represents a subgroup defined by the interaction of participants' political affiliation (left, right, center) and treatment condition (Accept, Control, Reject), while each column corresponds to a policy attribute (Appeal, Responsibility, Cost, Data Collection, Decision). Estimation via the Conditional Randomization Test (CRT) implemented via the CRTConjoint R package (Ham et al. 2024). The CRT evaluates whether a given attribute affects respondents' choices in any way, conditional on the other randomized factors. The y–axis stacks groups defined by political attitudes and treatment (treatment × political attitudes). Cells are shaded according to whether the attribute was statistically attended to within the group: black indicates that the attribute was attended to, while grey indicates that it was not attended to. Brackets on the y-axis group rows by treatment condition. To account for multiple comparisons across attributes and subgroups, we adjust the raw CRT p-values using Benjamini-Hochberg (Benjamini & Hochberg 1995). The CRT is unweighted; a weighted-CRT replicate is reported in Online Appendix Figure B5. Full details are provided in Appendix B.1

The results show that attention to procedural attributes depends both on respondents'

political attitudes and on the system outcome. *Responsibility* and *Decision* are attended to across all groups and outcomes. By contrast, *Cost* and *Appeal* are not always attended among right-leaning respondents. In the Control and Accept conditions, these respondents largely ignore both attributes, whereas in the Reject condition they take them into account. This pattern does not suggest that cost and appeal are unimportant to right-leaning respondents. Open-ended responses indicate that many of them view low costs and appeal mechanisms negatively, often describing them as not secure enough in managing access to the country. Consistent with this, cost and appeal appear to matter little when respondents already dislike the system outcome, but become relevant when the system produces an outcome they favor. In that case, these features may be seen as limiting or qualifying an otherwise desirable policy. Overall, the findings show that which procedural features respondents consider is not fixed, but varies with the policy outcome and with respondents' political views. These findings speak affirmatively toward hypothesis 1.

A different kind of selective inattention emerges with the group of left leaning respondents. They show consistent non-attendance to *Data Collection* across all treatment conditions. Here, we do not find non-attention strategically employed in order to fit legitimacy judgments and preferences. Instead, the data collection dimension of a system is generally ignored. We suggest that this is a symptom on focusing on other attributes. Non-attendance may thus reflect a neglect of incongruent (or "unhelpful") attributes. Correspondingly, attributes that reinforce respondents' preferred stance can receive more attention. A different interpretation of this finding is that the left ignores *Data Collection* altogether because they do not (want to) consider the security of the system.

Figure 8: Predicted mention probabilities with attribute–attendance overlay



Note: Each tile shows the model–predicted probability that respondents mention a given attribute in their open–ended justification. The x–axis lists attribute buckets (Appeal, Responsibility, Cost, Data, Decision). The y–axis stacks groups defined by partisan identity and treatment (Treatment × Partisanship). Tile fill encodes the predicted probability from an unweighted logistic regression with cluster–robust standard errors. White grid lines delineate tiles to aid visual separation. A red outline marks cells where the attribute failed the attendance test and is classified as *not attended* according to estimations from Figure 7. The absence of an outline indicates *attended*. The attendance overlay is drawn from the ANA summary and aligned by common attribute, identity, and treatment keys. Together, fill and outlines distinguish *how much* an attribute is mentioned (intensity) from *whether* it is statistically attended (attendance). A weighted-CRT replicate of the attendance overlay is reported in Online Appendix Figure B6. Full details are provided in Appendix B.1

We triangulate attribute attendance with the open-ended responses after the conjoint task. We asked participants what the most important reason for their decision was and counted instances of mentions of attributes.[11] We estimate whether a respondent mentioned an attribute in their open–ended justification. Figure 8 shows a heatmap of predicted probabilities and attribute attendance (same as in Figure 7). These predicted probabilities are displayed as the colored tiles in the heatmap. In addition, red outlines are overlaid from the earlier introduced attribute–attendance test, indicating where an attribute is statistically classified as being attended to. Together, the fills capture the intensity of attribute men-

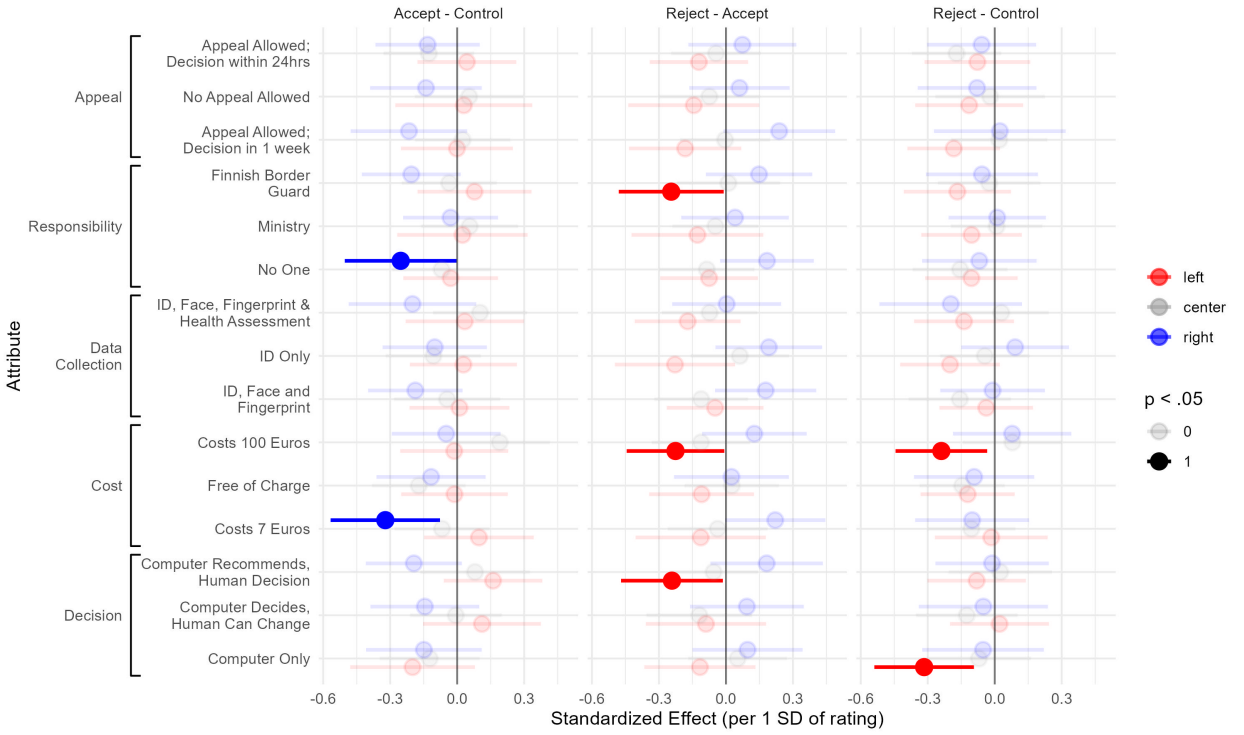
[11]Here, mentioning an attribute or a corresponding level was assessed as an instance of mentioning.

tions, while the outlines capture selection, allowing us to compare how groups emphasize or disregard different features across contexts. We add robustness to our findings of attribute (non-)attendance by showing that respondents also did not *mention* attributes that they neglected.

4.3.2 Information evaluation

Our second hypothesis suggests that system outcomes will impact the evaluation of system features across treatments. We expect individuals to rate the same procedural characteristics differently depending on the treatment condition they were assigned to. In other words, identical procedural characteristics will be evaluated differently depending on whether a system is expected to produce a favorable outcome. To this end, we estimate between treatment contrasts of triple interactions of treatment condition × attribute × partisanship (full details are provided in Appendix B.1). Figure 9 shows the results of estimations. Interpretation is as follows: a negative/positive coefficient in red/grey/blue shows a difference in ratings for seeing a given level in a system evaluation task for a given group left/center/right, between treatments, *ceteris paribus*. Put simply, if a coefficient is significant, it means that some attribute level for some subgroup was evaluated differently given the outcome. Thus, these estimations speak directly to our second hypothesis.

Figure 9: Context contrasts in legitimacy ratings by attribute level and partisan identity



Note: Points show estimated differences in legitimacy ratings for *Accept - Control*, *Reject - Accept*, and *Reject - Control*, with 95% confidence intervals. Color denotes partisan identity (left, center, right). Alpha reflects nominal significance as used in the plot legend. Estimates come from a linear model with weights. Uncertainty uses cluster-robust standard errors. Unweighted replicates are reported in Online Appendix Figures B3 and B4. Full details are provided in Appendix B.1

Figure 9 reveals several distinct patterns in how respondents adjust their evaluations of procedural features depending on outcome favorability. First, neither the *Appeal* attribute nor the *Data Collection* attribute show statistically significant differences across treatment conditions for any political group, indicating that these features are not subject to (detectable) motivated reasoning. Second, the *Responsibility* attribute does exhibit systematic shifts: right-leaning respondents penalize systems lacking a responsible authority, particularly when moving from the Control to the Accept condition, whereas left-leaning respondents react negatively when the Finnish Border Guard is responsible, especially when comparing Reject to Accept. Third, for the *Cost* attribute, left-leaning respondents most strongly downgrade systems that cost 100€ in the Reject condition relative to both other

conditions, while right-leaning respondents show heightened aversion to the 7€ option in the Accept condition compared to Control. Finally, the *Decision* attribute shows asymmetric responses: left-leaning respondents evaluate human decision-making more negatively when moving from Accept to Reject, yet they also react negatively to full automation in the Reject condition compared to Control. Notably, respondents from the center group *did not* evaluate any attribute level differently given changes in system outcomes.

These findings mirror fairness differences between political groups (as in section 4, see Figure 9) but also *within* groups given changes in system outcomes. That is, political groups show differences in cost perceptions because they deem it unfair (left dislikes too costly systems, right dislikes too cheap systems). Here, crucially, what makes systems more or less unfair is the outcome of the system (link (C)) *in addition to* baseline fairness perceptions of the process itself (link (A)). In the same vein, citizens dislike systems where "No One" or the "Finnish Border Guard" is responsible – contingent on system outcomes – *in addition to* baseline aversion to these levels (see Figure 9). Somewhat surprisingly, we find that systems with different levels of human-involvement can receive a penalty from the left-leaning respondents. We find that both "Human Decisions" and "Computer Only" can be disliked by those on the left when the outcome of the system is unfavorable.

Taken together, our findings provide clear evidence for hypothesis 2: outcome favorability moderates how citizens evaluate procedural features of ADM systems. At the same time, two important non-findings delimit the scope of this motivated reasoning. First, all significant effects are *negative*: no attribute level becomes more positively evaluated when outcomes align with respondents' preferences, and we observe no legitimacy dividend analogous to the systematic legitimacy penalties documented above. This pattern mirrors results reported earlier (Figure 6), where favorable outcomes did not generate higher overall system ratings but unfavorable outcomes reliably reduced them. Second, respondents in the political center do not exhibit strategic evaluation with respect to any procedural attribute, suggesting that reinterpretation of procedural fairness is concentrated among the political poles rather than

a general tendency across the electorate. Motivated reasoning in this context is therefore both asymmetrical – manifesting only as downward adjustment – and politically bounded, arising only among strongly identified left- and right-leaning respondents. Within these boundaries, however, the estimated effects are substantively meaningful, falling between small and medium in size (Cohen's d $\approx$ 0.26–0.37) and thus indicating non-trivial shifts in procedural evaluations when outcomes contradict group preferences. Overall, the results suggest that procedural legitimacy is vulnerable to motivated reinterpretation, but only under specific political and evaluative conditions.

5 Discussion

What can we say about the extent of motivated reasoning at play? Across conditions, outcome favorability shifted legitimacy ratings by 0.26 to 0.37 standard deviations. These small to medium effect sizes are noteworthy because legitimacy judgments are typically dominated by procedural design features (link A in Figure 1) and by the direct effect of outcome favorability (link B), both of which are well-established in the literature and are identifiable in our data (see sections 4.1 and 4.2). Our results show that the mechanism we isolate – the moderation of procedural evaluations by outcome favorability (link C) – produces effects comparable to many canonical predictors identified in prior work. Among left-leaning respondents, unfavorable outcomes reduced support for computer-based decision makers and higher-cost systems by d = –0.26 to –0.36, while right-leaning respondents reacted equally strongly to pro-migration defaults, with declines up to d = –0.37. In the responsibility domain, both groups registered meaningful shifts (d > –0.2), further underscoring that motivated reasoning permeates multiple procedural dimensions. This is our contribution to the literature on perceived legitimacy (Sunshine & Tyler 2003; Tyler 2006; Schmidt 2013; Doherty & Wolak 2012; Arnesen 2017; Esaiasson et al. 2019; Grimmelikhuijsen & Meijer 2022). That is, we add empirical evidence to the study of perceived legitimacy highlighting the role of outcome favorability as a moderator of process perceptions (Esaiasson et al. 2019). In our

experiment, outcome favorability – a well-established determinant of legitimacy – induced motivated reasoning by leading individuals to view identical decision-making processes as less legitimate when the outcomes did not align with their underlying preferences, or to ignore some features altogether. This pattern runs counter to the fair process effect: where the fair process effect holds that fair procedures help people accept an unfavorable outcome (Brockner & Wiesenfeld 1996), here an unfavorable outcome leads citizens to weigh down the perceived legitimacy of the very same procedures. These findings demonstrate that motivated reasoning is not merely a secondary distortion layered on top of baseline preferences but a measurable and consequential mechanism that shapes how citizens evaluate procedural features. The evidence suggests that the cognitive pathway through which outcomes color procedural judgments deserves a central place in explanations of ADM legitimacy, alongside institutional design and performance.

We furthermore contribute to the recent literature on ADM in public administration (Helberger et al. 2020; Ingrams et al. 2021; Martin & Waldman 2022; Grimmelikhuijsen & Meijer 2022; Yalcin et al. 2023; Horvath et al. 2023; Lünich & Kieslich 2024; Haesevoets et al. 2024; Wenzelburger et al. 2024; Araujo et al. 2020; Schiff et al. 2025). Our findings speak to this literature in multiple ways. First, we document citizens preferences regarding procedural features based on political attitudes. Responsibility and human oversight function as widely shared foundations of procedural legitimacy, whereas appeal rights, data collection, and cost are evaluated through the lens of citizens' political identities. We observe how procedural preferences can shift once political cues or policy-relevant considerations enter the evaluation context. These patterns also illuminate how underlying value orientations shape how different groups perceive procedural features of travel authorization systems. Left-leaning respondents exhibit broadly universalistic principles: systems that restrict appeal rights or impose high financial burdens are judged less legitimate because they threaten equal access and corrective opportunities. For right-leaning respondents, by contrast, particularistic considerations are central: systems that are free or inexpensive may be deemed inappropriate

because they imply a higher fiscal burden for taxpayers or insufficient stringency in border governance. In this sense, procedural legitimacy is filtered through divergent normative commitments, such that the same attribute – cost, appeal rights, or data collection – signals different concerns for different ideological groups.

Second, we show that citizens may evaluate the same algorithmic procedures differently depending on whether the system produces policy outcomes they prefer. We are able to demonstrate through which channels – i.e., attributes – this biased attention and evaluation operates. That is, all procedural attributes of our systems were in some way or another subject to motivated reasoning. Some system features were (strategically) ignored – *Appeal*, *Cost*, and *Data Collection*. Other system features were implicated in strategic devaluation due to system outcomes – *Responsibility*, *Cost*, and even *Decision*, which stands for the extent of human involvement. Human involvement in the decision-making process is the most robust predictor of ADM legitimacy, corresponding to link (A) in Figure 1. For example, human judges appear to be favored in terms of legitimacy (Yalcin et al. 2023), for decisions that are ideologically charged (Haesevoets et al. 2024), when reasons need to be explained (Grimmelikhuijsen & Meijer 2022), or whenever the decision requires trust, moral judgment, accountability, or relational qualities (Ingrams et al. 2021). This penalty does not reflect doubts about accuracy: Hermstrüwer & Langenbach (2023) show that citizens rate fully machine-based decisions as least fair while judging them no less accurate than human ones, and accept extensive delegation as long as a human retains the discretion to override. What citizens reject is not the algorithm, but the loss of human discretion. On the other hand, the literature also documents exceptions to this rule, for example when allocating COVID-19 vaccinations (Lünich & Kieslich 2024), more generally because algorithms are deemed impartial (Helberger et al. 2020), or better suited for some contexts than humans (Araujo et al. 2020). We add to this discussion by offering one plausible mechanism that may drive some of the variation. Specifically, we show that even the extent of human involvement in the decision-making process is subject to cognitive shortcuts and not a phenomenon that is to

be assessed as a natural constant. Neither deference to algorithms nor preference for human involvement is stable; instead, individuals follow or reject algorithmic advice selectively when it aligns with their expectations or relevant contextual cues (Alon-Barkat & Busuioc 2023). More than that, we demonstrated that this very feature was affected *the most* among the institutional dimensions of travel authorization systems. In essence, we documented that the same group of people exhibited a very different degree of aversion to automated decisions when outcomes appeared (un)favorable. We thus add nuance to the discussion on *algorithm aversion* and suggest that it might be less substantive than previous studies suggest. This reading is in line with recent experimental evidence. Schaap et al. (2024) find no default aversion to algorithmic decision makers: when human and machine agents are described in the same terms, acceptance depends on the expected benefits of the decision and on the control people retain over it. Even the label attached to an otherwise identical system changes how much people accept and follow it (Chacon et al. 2026). What looks like aversion to algorithms may therefore be better understood as a preference for discretion – and, as our results show, even this preference shifts when outcomes disappoint.

Our study advances research on motivated reasoning and political psychology (Kunda 1990; Redlawsk 2002; Taber & Lodge 2006; Slothuus & De Vreese 2010; Kahan 2016; Leeper & Slothuus 2014; Baekgaard et al. 2019; Már & Gastil 2020; Boyer et al. 2022). Most empirical work in this tradition examines highly polarized, two-party systems—especially the United States (Leeper & Slothuus 2014). We extend this literature by studying a polarizing topic in a consensus-oriented, multiparty context with low polarization and a strong welfare state. Despite these features, we find clear evidence of motivated reasoning in Finland, comparable in magnitude to previous studies. Our results therefore clarify *when* and *how* motivated reasoning emerges: across populations, MR appears robust, but within populations, its activation is conditional. In particular, strong prior preferences in our case served as the gateway for motivated reasoning. The center group – despite disliking the *Reject* system outcome – was not found to be engaging in motivated reasoning consistently. This does

not, however, mean that this group did not rely on some sort of a cognitive shortcut when forming their legitimacy evaluation. The theory of motivated reasoning suggests *accuracy-oriented* reasoning as a counterpart of goal-oriented reasoning. Under this umbrella, all other cognitive processes involved in forming (legitimacy) judgments are at work. Kunda (1990) suggested that when pursuing accuracy goals, people may rely on heuristics like availability or neglect of the law of large numbers. We thus propose to incorporate the assumption of bounded rationality (Simon 1955) to the study of legitimacy. Future research should therefore examine how motivated reasoning interacts with other cognitive and contextual factors to structure legitimacy evaluations across domains of public administration (Grimmelikhuijsen et al. 2016; Lenz & Viola 2017; Alon-Barkat & Busuioc 2024; Hemesath & Tepe 2025).

The attribute non-attendance we report (Figure 8) can be interpreted as a cognitive shortcut (Gigerenzer & Gaissmaier 2011). Recent eye-tracking studies have shown that conjoint respondents attend to only a subset of attributes rather than weighing all of them, a behavior referred to as survey satisficing (Jenke et al. 2021; Bansak & Jenke 2025). Survey satisficing tracks task burden and the number of choice tasks rather than the content of the profiles, so it should apply evenly across our outcome conditions. The non-attendance we observe does not: it shifts with outcome favorability, following the selective-recruitment and biased-weighting patterns reported above (Figure 8). This outcome-dependence is difficult to reconcile with survey satisficing and points to motivated processing.

Although our study focuses on a specific administrative setting, the results speak more broadly to the challenges that contemporary public authorities face when legitimacy depends not only on institutional design but also on how citizens cognitively approach the task of evaluation. This implies that legitimacy deficits may arise even when the technical design of ADM systems conforms to democratic governance norms of fairness or transparency. Rather than assuming that procedural enhancements will be taken at face value, administrators may need to anticipate that ideologically different audiences will process the same institutional features in systematically different ways, particularly in policy areas where directional

preferences are strong. Recognizing this cognitive layer is therefore essential for understanding where legitimacy vulnerabilities arise and for designing communication and governance strategies that can sustain public confidence in algorithmic decision-making.

5.1 Limitations

Several caveats bound our conclusions. Motivated reasoning is an umbrella term for a set of cognitive mechanisms that a survey experiment design cannot measure directly. We identify it behaviorally, from how procedural attributes are attended to and weighted across randomized outcome conditions. We corroborate these findings with qualitative evidence from the coded open-ended responses, offering an insight into the stated reasons driving the observed choices.

Relatedly, we operationalize outcome favorability via political-attitude tertiles. The mapping from ideology to migration preference is an assumption, and some respondents may be misclassified. We expect such misclassification to attenuate rather than manufacture the patterns we report, but it remains the principal inferential caveat. See Appendix B.1 for further details of inferential and estimation procedures.

Three further limitations concern scope. Finland is a low-polarization, consensus-oriented democracy. This makes the observation of motivated reasoning particularly informative about robustness, but it bounds the magnitude inference. Stronger effects are plausible in higher-polarization contexts. The conjoint task also asks respondents to evaluate decisions made about third-party travelers rather than about themselves. Hence, we are not able to address how skin-in-the-game affects legitimacy judgments. Finally, whether the weighting mechanism we document generalizes beyond migration and ADM in border control to other algorithmically-governed policy domains is an open empirical question we leave for future work.

6 Conclusion

This study set out to explore how citizens judge the legitimacy of automated decision-making in public administration. Our results demonstrate that legitimacy judgments in ADM are not merely reflections of technical and institutional design, but are shaped by preference-consistent information processing activated by outcome favorability. Drawing on the theory of motivated reasoning, we theorized that outcome favorability serves as a motivational trigger that guides how individuals attend to and evaluate procedural features of ADM systems, thereby moderating perceptions of procedural fairness. Across a preregistered survey experiment and open-ended responses, we documented clear evidence for two interrelated mechanisms: selective information recruitment and biased evaluation. Our findings show that well-intentioned procedural safeguards – such as ensuring accountability, embedding human oversight, or offering avenues for appeal – do not insulate systems using ADM from politicized interpretation when the outcomes at stake are substantively contested. At the same time, the absence of motivated reasoning among more moderate respondents suggests that legitimacy need not erode uniformly, and that some segments of the public continue to evaluate ADM systems in relatively accuracy-oriented ways. Taken together, these patterns demonstrate that ADM legitimacy hinges not only on what institutions do, but on how citizens make sense of what they do – an interpretive process shaped by motivated reasoning in polarized contexts.

Our study connects two strands of research that have so far developed largely in parallel. The first concerns the institutional and procedural foundations of legitimacy in automated decision-making. The second emphasizes the cognitive and motivational underpinnings of how citizens interpret institutional performance and outcomes. By bringing these perspectives together, we offer an empirical assessment of how outcome favorability – long recognized as a key element of legitimacy judgments – operates through motivated reasoning in the context of algorithmic governance. Our findings contribute to extant research in public administration in three distinct ways. First, we show how throughput and output legitimacy

interact – specifically, unfavorable outcomes decrease throughput legitimacy perceptions. Second, we provide an attribute-level account of how selective attention and biased weighting operate across key procedural dimensions of ADM systems. Third, by documenting motivated reasoning in a low-polarization context, we refine the boundary conditions under which motivated reasoning is activated in public administration. What we document in the domain of public administration parallels the *winner–loser gap* found in election studies (Anderson 2005), and more generally the moderating effect of outcome favorability on procedural legitimacy perceptions (Esaiasson et al. 2019).

For practitioners and policymakers, our findings imply that legitimacy cannot be secured solely through technical improvements in transparency, human oversight, or accountability. Even well-designed ADM systems may be judged as less legitimate when their outcomes conflict with citizens' prior preferences, meaning that public authorities must anticipate politically asymmetric interpretation of identical procedural safeguards. Recognizing that legitimacy is filtered through motivated reasoning highlights the need for communication strategies, institutional safeguards, and participatory practices that acknowledge these cognitive dynamics and help sustain legitimacy of automated governance.

7 Data availability

The data and replication code underlying this article will be made publicly available on Zenodo upon publication.

8 Bibliography

Alon-Barkat, S. & Busuioc, M. (2023). Human–ai interactions in public sector decision making:"automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, 33(1), 153–169. https://doi.org/10.1093/jopart/muac007

Alon-Barkat, S. & Busuioc, M. (2024). Public administration meets artificial intelligence: Towards a meaningful behavioral research agenda on algorithmic decision-making in government. *Journal of Behavioral Public Administration*, 7. https://doi.org/10.30636/jbpa.71.261

Alon-Barkat, S., Busuioc, M., Schwoerer, K., & Weißmüller, K. S. (2025). Algorithmic dis-

crimination in public service provision: Understanding citizens' attribution of responsibility for human versus algorithmic discriminatory outcomes. *Journal of Public Administration Research and Theory*, 35(4), 469–488. https://doi.org/10.1093/jopart/muaf024

Anderson, C. (2005). *Losers' consent: Elections and democratic legitimacy*. Oxford University Press. https://doi.org/10.1093/0199276382.001.0001

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. https://doi.org/10.1007/s00146-019-00931-w

Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models using marginaleffects for r and python. *Journal of Statistical Software*, 111(9), 1–32. https://doi.org/10.18637/jss.v111.i09

Arnesen, S. (2017). Legitimacy from decision-making influence and outcome favourability: Results from general population survey experiments. *Political Studies*, 65(1_suppl), 146–161. https://doi.org/10.1177/0032321716667956

Aula, V., Hillo, J., & Erkkilä, T. (2026). Ai in taxation—experimental evidence on citizen design preferences and perceptions of legitimacy. *Government Information Quarterly*, 43(2), 102147. https://doi.org/10.1016/j.giq.2026.102147

Baekgaard, M., Christensen, J., Dahlmann, C. M., Mathiasen, A., & Petersen, N. B. G. (2019). The role of evidence in politics: Motivated reasoning and persuasion among politicians. *British Journal of Political Science*, 49(3), 1117–1140. https://doi.org/10.1017/S0007123417000084

Baekgaard, M. & Serritzlew, S. (2015). Interpreting performance information: Motivated reasoning or unbiased comprehension? *Public Administration Review*, 76(1), 73–82. https://doi.org/10.1111/puar.12406

Bansak, K. & Jenke, L. (2025). Odd profiles in conjoint experimental designs: Effects on survey-taking attention and behavior. *Political Analysis*, 33, 315–338. https://doi.org/10.1017/pan.2025.1

Bayes, R. & Druckman, J. N. (2021). Motivated reasoning and climate change. *Current Opinion in Behavioral Sciences*, 42, 27–35. https://doi.org/10.1016/j.cobeha.2021.02.009

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior*, 36(2), 235–262. https://doi.org/10.1007/s11109-013-9238-0

Boyer, M. M., Aaldering, L., & Lecheler, S. (2022). Motivated reasoning in identity politics: Group status as a moderator of political motivations. *Political Studies*, 70(2), 385–401. https://doi.org/10.1177/0032321720964667

Brockner, J., Fishman, A. Y., Reb, J., Goldman, B. M., Spiegel, S., & Garden, C. (2007). Procedural fairness, outcome favorability, and judgments of an authority's responsibility. *Journal of Applied Psychology*, 92(6), 1657–1671. https://doi.org/10.1037/0021-9010.92.6.1657

Brockner, J. & Wiesenfeld, B. M. (1996). An integrative framework for explaining reactions

to decisions: Interactive effects of outcomes and procedures. *Psychological Bulletin*, 120(2), 189–208. https://doi.org/10.1037/0033-2909.120.2.189

Brockner, J. & Wiesenfeld, B. M. (2013). How, when, and why does outcome favorability interact with procedural fairness? *Handbook of Organizational Justice*, 525–553. Psychology Press. https://doi.org/10.4324/9780203774847-30

Brogaard, L., Madsen, J. K., & Petersen, O. H. (2024). Interpreting performance information: Motivated reasoning or unbiased comprehension? a replication and extension. *Public Administration*, 102(1), 318–340. https://doi.org/10.1111/padm.12970

Busch, P. A. (2023). Faced with digital bureaucrats: A scenario-based survey analysis of how clients perceive automation in street-level decision-making. *Government Information Quarterly*, 40(4), 101872. https://doi.org/10.1016/j.giq.2023.101872

Campbell, D., Hensher, D. A., & Scarpa, R. (2011). Non-attendance to attributes in environmental choice analysis: A latent class specification. *Journal of Environmental Planning and Management*, 54(8), 1061–1076. https://doi.org/10.1080/09640568.2010.549367

Chacon, A., Montecino, R., Reyes, T., & Kausel, E. E. (2026). The influence of terminology on use and preference in algorithmic systems: Beyond the hype for chatbots. *Big Data & Society*, 13(1). https://doi.org/10.1177/20539517261431514

Christensen, H. S., Himmelroos, S., & Setälä, M. (2020). A matter of life or death: A survey experiment on the perceived legitimacy of political decision-making on euthanasia. *Parliamentary Affairs*, 73(3), 627–650. https://doi.org/10.1093/pa/gsz014

Dahlberg, S. & Linde, J. (2016). Losing happily? the mitigating effect of democracy and quality of government on the winner–loser gap in political support. *International Journal of Public Administration*, 39(9), 652–664. https://doi.org/10.1080/01900692.2016.1177831

Damgaard, M. & James, O. (2024). Public service users' responses to performance information: Bayesian learning or motivated reasoning? *Journal of Public Administration Research and Theory*, 34(1), –. https://doi.org/10.1093/jopart/muae013

Dias, N. & Lelkes, Y. (2022). The nature of affective polarization: Disentangling policy disagreement from partisan identity. *American Journal of Political Science*, 66(3), 775–790. https://doi.org/10.1111/ajps.12628

Doherty, D. & Wolak, J. (2012). When do the ends justify the means? evaluating procedural fairness. *Political Behavior*, 34(2), 301–323. https://doi.org/10.1007/s11109-011-9166-9

Druckman, J. N., Leeper, T. J., & Slothuus, R. (2018). Motivated responses to political communications: Framing, party cues, and science information. *The feeling, thinking citizen*, 125–150. Routledge. https://doi.org/10.4324/9781351215947-8

Druckman, J. N. & Levendusky, M. S. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly*, 83(1), 114–122. https://doi.org/10.1093/poq/nfz003

Druckman, J. N. & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, 9(2), 111–119. https://doi.org/10.1038/s41558-018-0360-1

Epley, N. & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30(3), 133–140. https://doi.org/10.1257/jep.30.3.133

Esaiasson, P., Arnesen, S., & Werner, H. (2023). How to be gracious about political loss—the

importance of good loser messages in policy controversies. *Comparative Political Studies*, 56(5), 599–624. https://doi.org/10.1177/00104140221109433

Esaiasson, P., Persson, M., Gilljam, M., & Lindholm, T. (2019). Reconsidering the role of procedures for decision acceptance. *British Journal of Political Science*, 49(1), 291–314. https://doi.org/10.1017/S0007123416000508

Folger, R. & Cropanzano, R. (1998). *Organizational Justice and Human Resource Management*. Foundations for Organizational Science. Sage. https://doi.org/10.4135/9781452225777

Ford, S. (2012). A social psychology model of the perceived legitimacy of international criminal courts: Implications for the success of transitional justice mechanisms. *Vanderbilt Journal of Transnational Law*, 45, 405. https://scholarship.law.vanderbilt.edu/vjtl/vol45/iss2/1/

Gaozhao, D., Wright, J. E., & Gainey, M. K. (2024). Bureaucrat or artificial intelligence: people's preferences and perceptions of government service. *Public Management Review*, 26(6), 1498–1525. https://doi.org/10.1080/14719037.2022.2160488

Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62(2011), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, 83(2), 241–262. https://doi.org/10.1111/puar.13483

Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2016). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, 76(3), 404–415. https://doi.org/10.1111/puar.12609

Grimmelikhuijsen, S. & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. *Perspectives on Public Management and Governance*, 5(3), 232–242. https://doi.org/10.1093/ppmgov/gvac008

Haesevoets, T., Verschuere, B., Van Severen, R., & Roets, A. (2024). How do citizens perceive the use of artificial intelligence in public sector decisions? *Government Information Quarterly*, 41(1), 101906. https://doi.org/10.1016/j.giq.2023.101906

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political analysis*, 22(1), 1–30. https://doi.org/10.1093/pan/mpt024

Hainmueller, J., Mummolo, J., & Xu, Y. (2019). How much should we trust estimates from multiplicative interaction models? simple tools to improve empirical practice. *Political Analysis*, 27(2), 163–192. https://doi.org/10.1017/pan.2018.46

Ham, D. W., Imai, K., & Janson, L. (2024). Using machine learning to test causal hypotheses in conjoint analysis. *Political Analysis*, 32(3), 329–344. https://doi.org/10.1017/pan.2023.41

Heidenreich, S., Watson, V., Ryan, M., & Phimister, E. (2018). Decision heuristic or preference? Attribute non-attendance in discrete choice problems. *Health Economics*, 27(1), 157–171. https://doi.org/10.1002/hec.3524

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review*, 39, 105456. https://doi.org/10.1016/j.clsr.2020.105456

Hemesath, S. & 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

Hensher, D. A. & Greene, W. H. (2010). Non-attendance and dual processing of common-metric attributes in choice analysis: A latent class specification. *Empirical Economics*, 39(2), 413–426. https://doi.org/10.1007/s00181-009-0310-x

Hermstrüwer, Y. & Langenbach, P. (2023). Fair governance with humans and machines. *Psychology, Public Policy, and Law*, 29(4), 525–548. https://doi.org/10.1037/law0000381

Hillo, J., Vento, I., & Erkkilä, T. (2025). Algorithmic governance: Experimental evidence on citizens' and public administrators' legitimacy perceptions of automated decision-making. *Public Administration*. https://doi.org/10.1111/padm.70028

Horvath, L., James, O., Banducci, S., & Beduschi, A. (2023). Citizens' acceptance of artificial intelligence in public services: Evidence from a conjoint experiment about processing permit applications. *Government Information Quarterly*, 40(4), 101876. https://doi.org/10.1016/j.giq.2023.101876

Ingrams, A., Kaufmann, W., & Jacobs, D. (2021). In ai we trust? citizen perceptions of ai in government decision making. *Policy & Internet*, 14(2), 390–409. https://doi.org/10.1002/poi3.276

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual review of political science*, 22(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Iyengar, S. & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3), 690–707. https://doi.org/10.1111/ajps.12152

James, O. & Van Ryzin, G. G. (2017). Motivated reasoning about public performance: An experimental study of how citizens judge the affordable care act. *Journal of Public Administration Research and Theory*, 27(1), 197–209. https://doi.org/10.1093/jopart/muw049

Jenke, L., Bansak, K., Hainmueller, J., & Hangartner, D. (2021). Using eye-tracking to understand decision-making in conjoint experiments. *Political Analysis*, 29(1), 75–101. https://doi.org/10.1017/pan.2020.11

Jost, J. T., Baldassarri, D. S., & Druckman, J. N. (2022). Cognitive–motivational mechanisms of political polarization in social-communicative contexts. *Nature reviews psychology*, 1(10), 560–576. https://doi.org/10.1038/s44159-022-00093-5

Kahan, D. M. (2016). *The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It*, (Chapter 1, 1–16). John Wiley& Sons, Ltd. https://doi.org/10.1002/9781118900772.etrds0417

Kane, J. V. (2025). More than meets the itt: A guide for anticipating and investigating nonsignificant results in survey experiments. *Journal of Experimental Political Science*, 12(1), 110–125. https://doi.org/10.1017/XPS.2024.1

Keppeler, F. (2024). No thanks, dear ai! understanding the effects of disclosure and deployment of artificial intelligence in public sector recruitment. *Journal of Public Administration Research and Theory*, 34(1), 39–52. https://doi.org/10.1093/jopart/muad009

Kern, C., Gerdon, F., Bach, R. L., Keusch, F., & Kreuter, F. (2022). Humans versus machines: Who is perceived to decide fairer? experimental evidence on attitudes toward automated decision-making. *Patterns*, 3(10), 100591. https://doi.org/10.1016/j.patter.2022.100591

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3), 480. https://doi.org/10.1037/0033-2909.108.3.480

Leeper, T. J., Hobolt, S. B., & Tilley, J. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis*, 28(2), 207–221. https://doi.org/10.1017/pan.2019.30

Leeper, T. J. & Slothuus, R. (2014). Political parties, motivated reasoning, and public opinion formation. *Political psychology*, 35, 129–156. https://doi.org/10.1111/pops.12164

Lenz, T. & Viola, L. A. (2017). Legitimacy and institutional change in international organisations: a cognitive approach. *Review of International Studies*, 43(5), 939–961. https://doi.org/10.1017/S0260210517000201

Lünich, M. & Kieslich, K. (2024). Exploring the roles of trust and social group preference on the legitimacy of algorithmic decision-making vs. human decision-making for allocating covid-19 vaccinations. *AI & SOCIETY*, 39(1), 309–327. https://doi.org/10.1007/s00146-022-01412-3

Majone, G. (1997). From the positive to the regulatory state: Causes and consequences of changes in the mode of governance. *Journal of public policy*, 17(2), 139–167. https://doi.org/10.1017/S0143814X00003524

Martin, K. & Waldman, A. (2022). Governing algorithmic decisions: The role of decision importance and governance on perceived legitimacy of algorithmic decisions. *Big Data & Society*, 9(1), 20539517221100449. https://doi.org/10.1177/20539517221100449

Martin, K. & Waldman, A. (2023). Are algorithmic decisions legitimate? the effect of process and outcomes on perceptions of legitimacy of ai decisions. *Journal of Business Ethics*, 183(3), 653–670. https://doi.org/10.1007/s10551-021-05032-7

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. https://doi.org/10.1093/jopart/muaa019

Moynihan, D. P., Herd, P., & Harvey, H. (2015). Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory*, 25(1), 43–69. https://doi.org/10.1093/jopart/muu009

Már, K. & Gastil, J. (2020). Tracing the boundaries of motivated reasoning: How deliberative minipublics can improve voter knowledge. *Political Psychology*, 41(1), 107–127. https://doi.org/10.1111/pops.12591

Pierce, L., Rogers, T., & Snyder, J. A. (2016). Losing hurts: The happiness impact of partisan electoral loss. *Journal of Experimental Political Science*, 3(1), 44–59. https://doi.org/10.1017/XPS.2015.8

Redlawsk, D. P. (2002). Hot cognition or cool consideration? testing the effects of motivated reasoning on political decision making. *Journal of Politics*, 64(4), 1021–1044. https://doi.org/10.1111/1468-2508.00161

Schaap, G., Bosse, T., & Hendriks Vettehen, P. (2024). The abc of algorithmic aversion: not agent, but benefits and control determine the acceptance of automated decision-making. *AI & Society*, 39(4), 1947–1960. https://doi.org/10.1007/s00146-023-01649-6

Scharpf, F. (1999). *Governing in Europe: Effective and democratic?* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198295457.001.0001

Schiff, K. J., Schiff, D. S., Adams, I. T., McCrain, J., & Mourtgos, S. M. (2025). Institutional factors driving citizen perceptions of ai in government: Evidence from a survey experiment

on policing. *Public Administration Review*, 85(2), 451–467. https://doi.org/10.1111/puar.13754

Schmidt, V. A. (2013). Democracy and legitimacy in the european union revisited: Input, output and 'throughput'. *Political studies*, 61(1), 2–22. https://doi.org/10.1111/j.1467-9248.2012.00962.x

Schreck, P., Kruse, N., & Conti, G. (2025). The role of organizational motives in the formation of moral legitimacy judgments. *Business Ethics, the Environment & Responsibility*. https://doi.org/10.1111/beer.12801

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99–118. https://doi.org/10.2307/1884852

Skitka, L. J., Bauman, C. W., & Lytle, B. L. (2009). Limits on legitimacy: moral and religious convictions as constraints on deference to authority. *Journal of personality and social psychology*, 97(4), 567. https://doi.org/10.1037/a0015998

Slothuus, R. & De Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of politics*, 72(3), 630–645. https://doi.org/10.1017/s002238161000006x

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. https://doi.org/10.1017/dap.2020.19

Stoetzer, L. F., Munzert, S., Lowe, W., Çalı, B., Gohdes, A. R., Helbling, M., Maxwell, R., & Traunmüller, R. (2023). Affective partisan polarization and moral dilemmas during the covid-19 pandemic. *Political Science Research and Methods*, 11(2), 429–436. https://doi.org/10.1017/psrm.2022.13

Strickland, A. A., Taber, C. S., & Lodge, M. (2011). Motivated reasoning and public opinion. *Journal of health politics, policy and law*, 36(6), 935–944. https://doi.org/10.1215/03616878-1460524

Strömbäck, C., Andersson, D., Västfjäll, D., & Tinghög, G. (2024). Motivated reasoning, fast and slow. *Behavioural Public Policy*, 8(3), 617–632. https://doi.org/10.1017/bpp.2021.34

Su, S. (2022). Updating politicized beliefs: How motivated reasoning contributes to polarization. *Journal of Behavioral and Experimental Economics*, 96, 101799. https://doi.org/10.1016/j.socec.2021.101799

Sunshine, J. & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review*, 37(3), 513–548. https://doi.org/10.1111/1540-5893.3703002

Taber, C. S. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Thibaut, J. & Walker, L. (1978). A theory of procedure. *California Law Review*, 66(3), 541–566. https://doi.org/10.2307/3480099

Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, 57(Volume 57, 2006), 375–400. https://doi.org/10.1146/annurev.psych.57.102904.190038

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

Yalcin, G., Themeli, E., Stamhuis, E., Philipsen, S., & Puntoni, S. (2023). Perceptions of justice by algorithms. *Artificial Intelligence and Law*, 31(2), 269–292. https://doi.org/10.1007/s10506-022-09312-z

Zheng, G., Lee, H., Li, W., Liu, Y., & Tang, T. (2026). Navigating the algorithmic state: Public values and citizens' preferences of ai-driven public service delivery. *Public Administration*. https://doi.org/10.1111/padm.70066

A Appendix

A.1 Tables

Table A1: Automated travel systems outcome treatment description

In the future, travelers coming to Finland from countries that do not need a visa to enter the EU will still need to get an authorisation 72h before their trip. This new system is called ETIAS, which stands for Electronic Travel Identification and Authorisation System. Individuals will apply for ETIAS online. Applications will be quickly processed by computer systems using personal data about travelers. On the next screen, you can see two alternatives of how ETIAS could be set up.		
The default setting of ETIAS will be to accept all asylum seekers to the country.	The default setting of ETIAS will be to prevent asylum seekers from entering the country.	[CONTROL]

Table A2: Conjoint Experiment Design

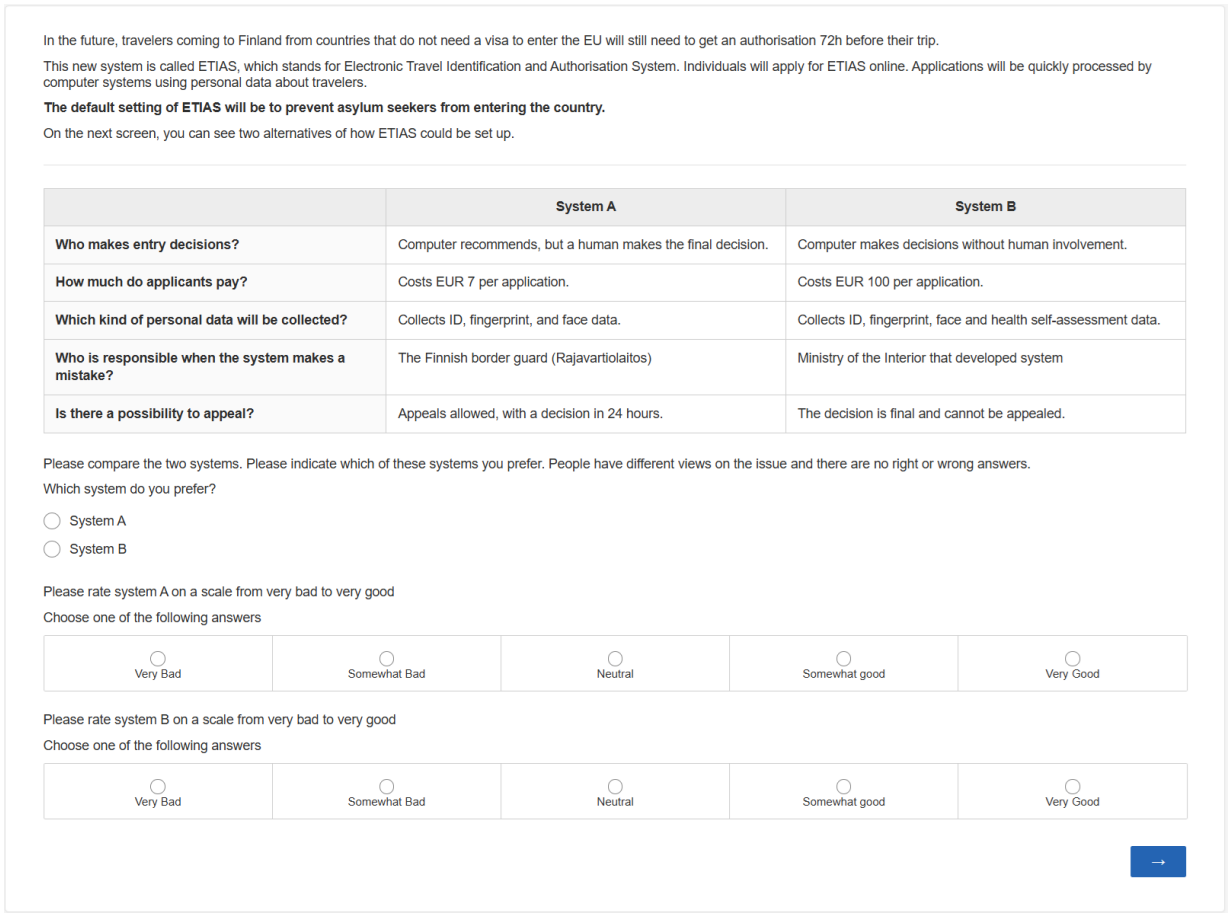
Attribute	**Levels**
1) Who makes entry decisions?	1a) Computer makes decisions without human involvement. 1b) Computer recommends, but a human makes the final decision. 1c) Computer makes decisions, but border guards can change the decision.
2) How much do applicants pay?	2a) Free to use. 2b) Costs EUR 7 per application. 2c) Costs EUR 100 per application.
3) Which kind of personal data will be collected?	3a) Collects ID, fingerprint, and face data. 3b) Collects ID, fingerprint, face and health self-assessment data. 3c) Collects ID.
4) Who is responsible when the system makes a mistake?	4a) The Finnish border guard (Rajavartiolaitos). 4b) Ministry of the Interior that developed the system. 4c) No one.
5) Is there a possibility to appeal?	5a) Appeals allowed, with a decision in 24 hours. 5b) Appeals allowed, with a decision in one week. 5c) The decision is final and cannot be appealed.

Table A3: Balance Table Across Context Conditions

	Accept		Control		Reject		Test
Variable	N	%	N	%	N	%	
Political Attitudes Group	617		606		626		$\chi^2 = 6.723$
center	245	40%	251	41%	259	41%	
left	177	29%	188	31%	208	33%	
right	195	32%	167	28%	159	25%	
Age Groups	609		593		621		$\chi^2 = 6.195$
18–24	23	4%	20	3%	29	5%	
25–34	50	8%	57	10%	67	11%	
35–44	60	10%	58	10%	55	9%	
45–54	101	17%	89	15%	96	15%	
55–64	148	24%	150	25%	164	26%	
65+	227	37%	219	37%	210	34%	
Gender	616		602		625		$\chi^2 = 5.207^*$
Female	329	53%	296	49%	294	47%	
Male	287	47%	306	51%	331	53%	

Statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure A1: Reconstruction of the Qualtrics conjoint task as shown to respondents



In the future, travelers coming to Finland from countries that do not need a visa to enter the EU will still need to get an authorisation 72h before their trip.

This new system is called ETIAS, which stands for Electronic Travel Identification and Authorisation System. Individuals will apply for ETIAS online. Applications will be quickly processed by computer systems using personal data about travelers.

The default setting of ETIAS will be to prevent asylum seekers from entering the country.

On the next screen, you can see two alternatives of how ETIAS could be set up.

	System A	System B
Who makes entry decisions?	Computer recommends, but a human makes the final decision.	Computer makes decisions without human involvement.
How much do applicants pay?	Costs EUR 7 per application.	Costs EUR 100 per application.
Which kind of personal data will be collected?	Collects ID, fingerprint, and face data.	Collects ID, fingerprint, face and health self-assessment data.
Who is responsible when the system makes a mistake?	The Finnish border guard (Rajavartiolaitos)	Ministry of the Interior that developed system
Is there a possibility to appeal?	Appeals allowed, with a decision in 24 hours.	The decision is final and cannot be appealed.

Please compare the two systems. Please indicate which of these systems you prefer. People have different views on the issue and there are no right or wrong answers.

Which system do you prefer?

○ System A

○ System B

Please rate system A on a scale from very bad to very good

Choose one of the following answers

Very Bad	Somewhat Bad	Neutral	Somewhat good	Very Good
○	○	○	○	○

Please rate system B on a scale from very bad to very good

Choose one of the following answers

Very Bad	Somewhat Bad	Neutral	Somewhat good	Very Good
○	○	○	○	○

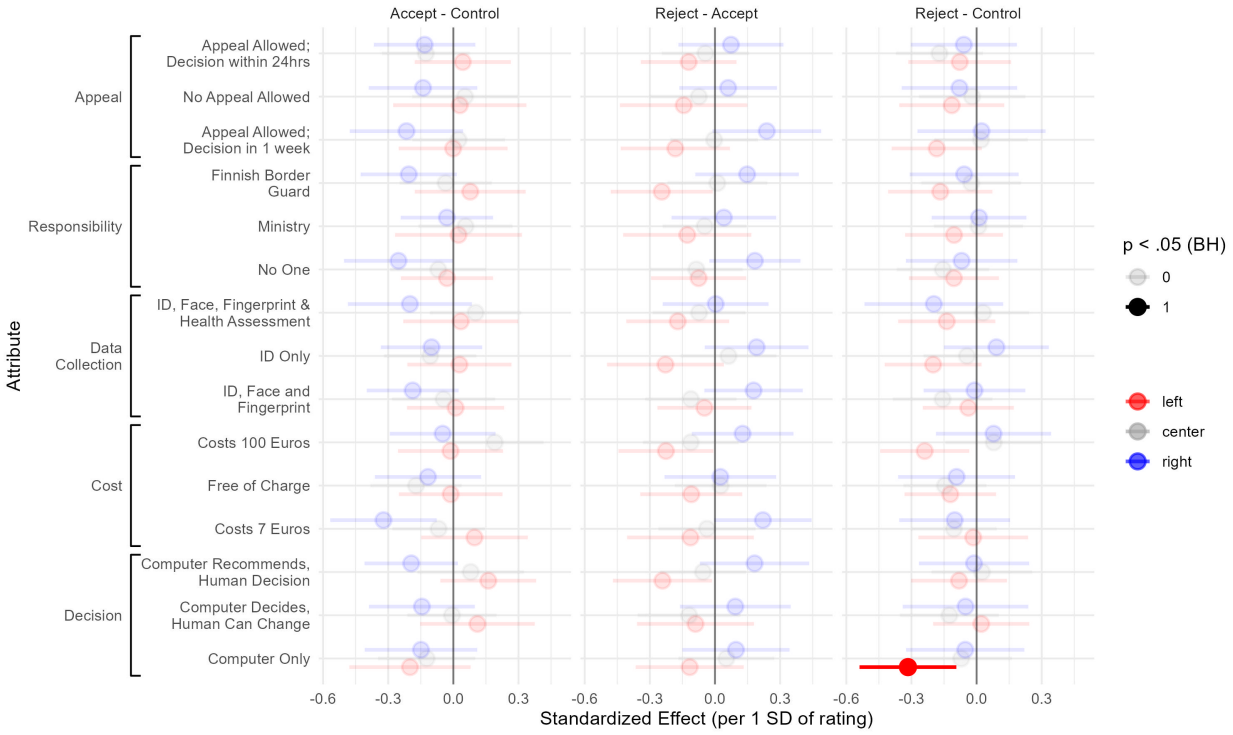
→

Note: Mock reconstruction of a single conjoint task as it appeared to respondents in the Qualtrics survey, shown here for the *reject* treatment. Attribute rows and levels reproduce the conjoint design in Table A2; the vignette text matches Table A1. The specific combination of attribute levels for System A and System B is one randomly drawn pair (of $3^5 \times 3^5$ possible profiles); each respondent saw three such tasks. Rating scale and choice labels reproduce the original instrument.

A.1.1 Robustness

Figure A2 applies the Benjamini–Hochberg procedure to the same estimates shown in the main Figure. The adjustment controls the false discovery rate rather than the family-wise error rate. Transparency encodes BH-adjusted significance within each feature-by-contrast family. Patterns of partisan heterogeneity remain visible after adjustment. The substantive conclusions drawn from the main Figure are robust under the Benjamini–Hochberg correction.

Figure A2: Robustness to multiple testing adjustments



Note: The estimands match Figure 9. Significance is assessed using the Benjamini–Hochberg false discovery rate adjustment. Alpha reflects adjusted significance ($p_{\text{BH}} < .05$) computed within feature-by-contrast families. Color denotes political ideology.

B Online Appendix

B.1 Identification and estimation details

B.1.1 Figure 3

Figure 3 presents a ridge density plot of respondents' political attitudes, constructed by estimating the kernel density of the partisanship measure. Using the empirical cumulative distribution function, the distribution is divided at the 33rd and 66th percentiles, producing three groups – *Left*, *Center*, and *Right* – which are shaded in red, grey, and blue, respectively. The plot illustrates how respondents are concentrated along the political-attitudes continuum and forms the basis for the political-attitudes grouping used in subsequent analyses.

B.1.2 Figure 5

We estimate marginal means for each attribute level in the conjoint design to summarize how likely respondents are to prefer profiles featuring a given procedural feature. These marginal means represent the predicted probability of choosing a profile when each attribute level is set to that value while all other attributes are averaged over their observed distribution. To examine heterogeneity in these baseline procedural evaluations, we compute four separate sets of estimates: one for the full sample and three for respondents on the left, in the center, and on the right of the political spectrum. For each subgroup and for the full sample, marginal means are estimated independently within each of the three outcome contexts (*Control*, *Accept*, and *Reject*), allowing us to compare how different political groups evaluate identical attribute levels across varying system outcomes. Weights for language, region, age, gender, and education are applied. Estimations are carried out using multinomial logit models via the `cregg` package in R (Leeper et al. 2020).

B.1.3 Figure 6

To estimate how the effect of system outcomes on legitimacy varies across levels of political attitudes, we employ the `interflex` package in R using the binning estimator (Hainmueller et al. 2019). This model estimates conditional marginal effects of treatment by comparing respondents' system ratings across outcome conditions while allowing the magnitude of these effects to change flexibly along the moderator, political attitudes. The binning estimator divides the moderator into three empirically balanced groups and computes average treatment effects within each bin, corresponding to the groups shown in Figure 3. We include task-level fixed effects to account for within-respondent correlation, and apply population weights for age, gender, education, language, and region. In the estimation, we control for all procedural attributes in the conjoint design, whether the system was presented on the left or right side of the screen, and for respondents' party affect scores toward all major Finnish parties, as

well as demographic controls for region, age group, gender, and education. Treatment effects are estimated relative to the *Control* outcome, with separate contrasts for the *Accept* and *Reject* conditions.

B.1.4 Figure 7

To detect which attributes respondents actively relied on when making choices, we implement the conditional randomization test (CRT) for conjoint data using the `CRTConjoint` package in R (Ham et al. 2024). For each partisan group (left, center, right) and each outcome context (*Control*, *Accept*, *Reject*), we reshape the data into paired profiles and test one attribute at a time, treating the remaining attributes as conditioning variables. The CRT works by repeatedly simulating counterfactual datasets in which the focal attribute is permuted conditional on all other attributes, and by re-estimating a fixed specification of the choice model to build a reference distribution for the test statistic. In our case, this specification is a logistic-choice model including all left–right attribute indicators, ensuring that any predictive contribution of the focal attribute—whether additive, interactive, or heterogeneous—is captured by the test. P-values are computed from 1,000 randomization draws, and attributes are coded as *attended* when their Benjamini–Hochberg–adjusted p-value falls below 0.05. This procedure is repeated separately for each political-attitude group and each outcome condition, yielding a full matrix of attribute attendance across treatment and partisanship combinations. By construction, this approach does not assume linearity, additivity, or correct model specification, and instead evaluates whether an attribute has any detectable effect on choices once all other attribute information is conditioned upon. The resulting attendance heatmaps in Figure 7 therefore provide a model-agnostic test of which information about procedural features respondents recruited when forming legitimacy judgments. The CRT is estimated without survey weights.

B.1.5 Figure 8

To analyse how respondents justify their legitimacy evaluations, we model the probability that a given procedural attribute is mentioned in the open–ended response. We first construct attribute–specific binary indicators (Appeal, Responsibility, Cost, Data, Decision) using text-coded buckets that identify whether respondents mentioned a particular procedural feature in their justification. We then estimate a series of logistic regression models in which each attribute bucket is treated as the dependent variable and predicted from a full interaction between partisan group and treatment context, while including task fixed effects to absorb within–respondent dependence across choice tasks. Estimation is unweighted, to match the (also unweighted) CRT used for the attendance overlay, and standard errors are clustered at the respondent level to account for repeated observations. From each model,

we generate predicted probabilities for all combinations of partisan group, treatment condition, and attribute, using a balanced prediction grid to ensure comparability across subgroups. These predicted mention probabilities capture the *intensity* with which respondents draw on different procedural considerations when explaining their judgments. To integrate these results with the attribute non-attendance (ANA) analysis, we merge predictions with the CRT–based attendance classifications using harmonized attribute keys for each partisan group and treatment condition. In the resulting heatmap, tile shading represents the predicted probability of mentioning an attribute, while red borders indicate attributes that failed the ANA test and are therefore classified as *not attended*. This combined visualization allows us to distinguish whether an attribute is statistically used in respondents' choices (attendance) from how prominently it features in their justifications (mention probability), thereby linking behavioral and self-reported evidence of motivated information processing.

B.1.6 Figure 9

We estimate a weighted OLS model of respondents' legitimacy ratings in which each of the five randomized attributes (Decision, Cost, Data Collection, Responsibility, Appeal) is interacted with partisan identity (left, center, right) and with treatment condition (Control, Accept, Reject), while including task fixed effects to absorb systematic differences across evaluation tasks. Survey weights align the sample with population benchmarks for age, gender, education, region, and language, and standard errors are clustered by respondent to account for within-person dependence. To interpret the moderated treatment effects, we compute pairwise contrasts of outcome contexts within each attribute level and partisan group (Accept−Control, Reject−Control), using a balanced prediction grid generated from the fitted model using `marginaleffects` in R (Arel-Bundock et al. 2024); for completeness, we additionally report Accept−Reject comparisons. All estimated contrasts are then scaled by the weighted standard deviation of the legitimacy rating, producing standardized effect sizes that allow comparisons across attributes and treatment contrasts; confidence intervals are likewise standardized. We present point estimates with 95% cluster-robust confidence intervals and highlight coefficients significant at the 5%-level, with Appendix Figure A2 showing which findings remain robust under Benjamini-Hochberg correction (notably, the *Computer only* level among left-leaning respondents in the Reject−Control comparison).

We are estimating triple interactions: treatment condition × attribute × partisanship. Confidence intervals reflect this added uncertainty, as they are relatively large. There are two main concerns: false positives and false negatives. First, we run into the danger of identifying non-effects as real because of the high number of tests we run. We account for this by adjusting p-values for false discovery rates (Benjamini & Hochberg 1995). Second,

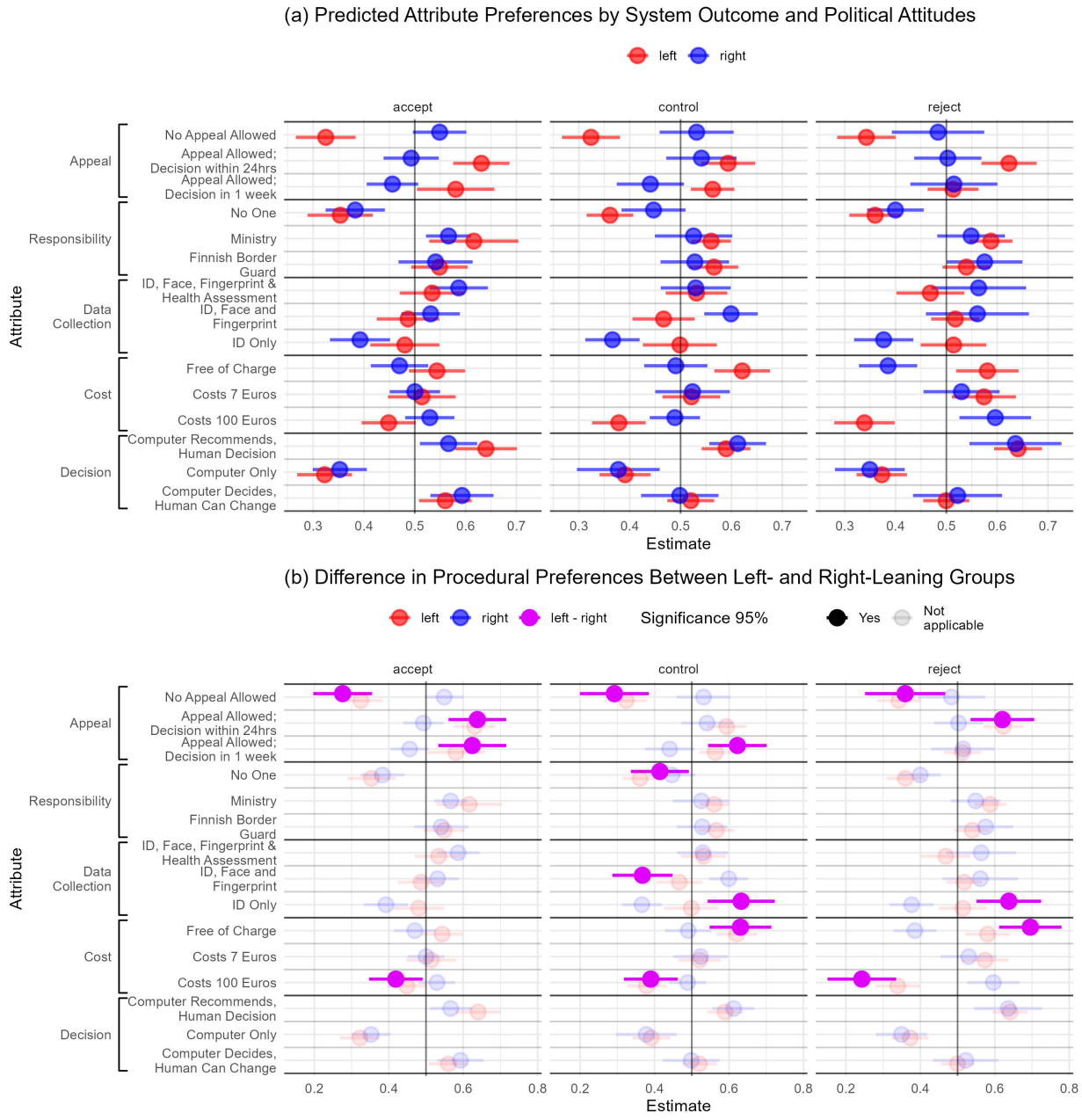
we run into the potential problems of not detecting true effects, because of estimating triple interactions that demand a lot of power (Kane 2025). Since all the effects we report have Cohen's d large enough to be considered relevant (> 0.2), and more importantly, all of the effects in Figure 9 show the same sign for the respective group $\times$ treatment contrasts, we maintain that if anything, more power would probably not lead to falsifying our findings. Because of that, we point out that effect sizes, here, are potentially more important than p-values.

B.2 Figures

B.2.1 Heterogeneity

Panel (a) of Figure B1 shows marginal means across all attribute levels, while Panel (b) reports contrasts between left- and right-leaning respondents. Some attributes are evaluated consistently across groups: both sides favor human-in-the-loop systems and identifiable authorities. Right-leaning respondents become significantly more favorable toward high-cost systems under the *Reject* condition, while left-leaning respondents continue to favor low-cost or free systems. Right-leaning respondents also support more privacy-invasive data regimes, whereas the left show little differentiation across data regimes. Appeals matter more for left-leaning respondents, particularly when they are resolved quickly.

Figure B1: Procedural Preferences by System Outcome and Political Attitudes



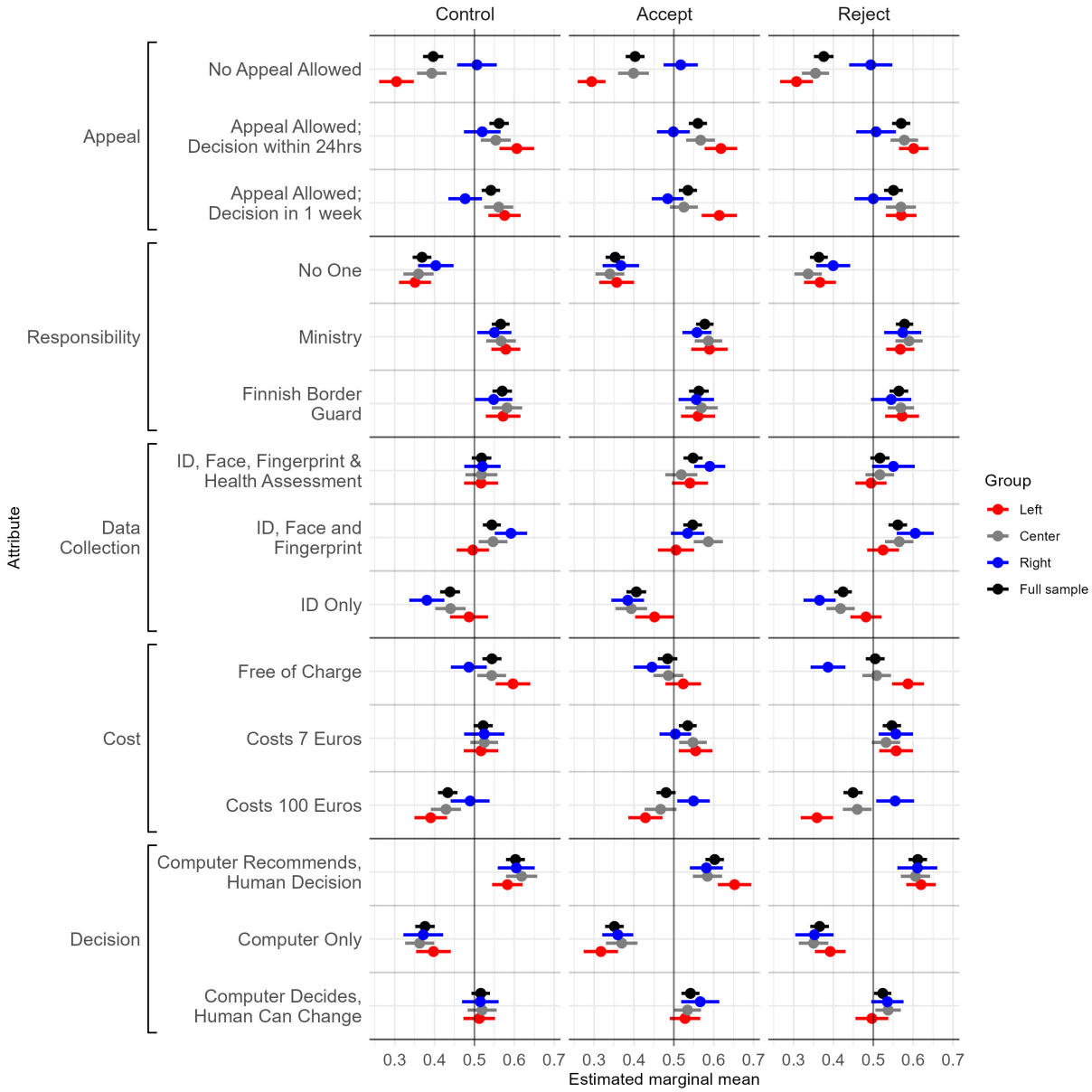
Note: Estimates are based on multinomial logit models stratified by treatment condition (control, accept, reject) and polarization group (left, right). Models are estimated using `cregg::mm()` with weighting applied. Estimates reflect the marginal effects in preferring an average system given a specific attribute level, with responses weighted for population representativeness. Deviations from 0.5 indicate positive or negative preferences, compared to indifference at 0.5. Panel (a) shows Marginal means estimated for all treatments (*Control*, *Reject*, *Accept*) and for the political groups (*left* in red, *right* in blue). Panel (b) shows differences between left- and right-leaning respondents which are computed using `cregg::mm_diffs()` after estimating group-specific multinomial logit models. The Figure shows attribute-level contrasts across treatment conditions, with 95% confidence intervals corrected for multiple comparisons. Only confidence-intervals for contrasts (purple) are shown were the effect is statistically significant at the 95%-level after using BH adjustments. Transparent point estimates in blue and red in the background for right and left leaning groups (corresponding to (a)), respectively.

B.2.2 Sensitivity to survey weighting

Sample weights align the realised sample with population benchmarks on language, region, age, gender, and education. Because the opt-in subpanel underrepresents some demographic strata, weighting changes the relative contribution of respondents to the estimates and can affect both the point estimates and their precision. We re-estimate the headline figures of the main text without applying weights and report the resulting replicates in this section.

Figure B2 reproduces the marginal-mean predictions of Figure 5 without weights. The ordering of attribute levels within each subgroup-by-treatment cell is unchanged, and the same qualitative patterns hold: broad consensus on *Responsibility* and *Decision*, ideological divergence on *Appeal*, *Data Collection*, and *Cost*. Weighting therefore moves the point estimates marginally but does not alter the substantive conclusions of Section 4.1.

Figure B2: Predicted attribute preferences (unweighted)

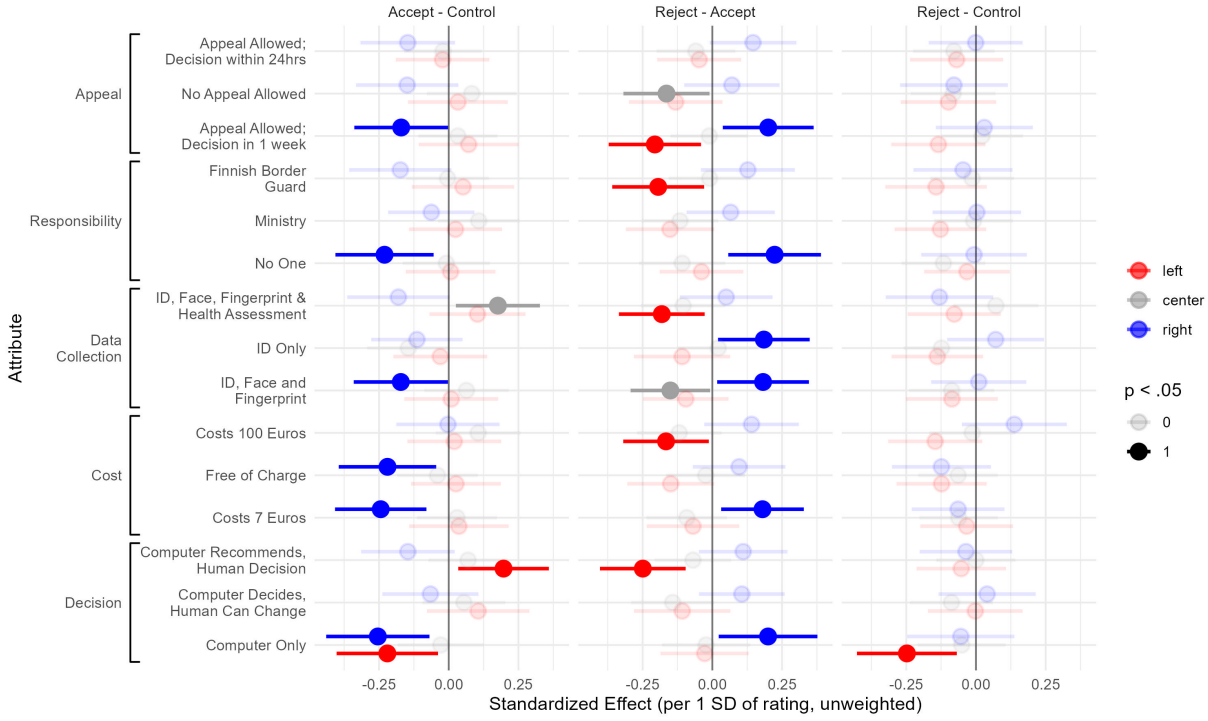


Note: Unweighted replicate of Figure 5. Estimates are marginal means from multinomial logit models via `cregg::mm()`, separately for the full sample and for ideological subgroups (left, center, right), within each outcome condition (*Control*, *Accept*, *Reject*). Survey weights are *not* applied. Values above (below) 0.5 indicate positive (negative) evaluations relative to indifference; 95% confidence intervals shown.

To check the sensitivity of the headline pattern, we re-estimate the standardized contrasts of Figures 9 and A2 without applying weights, using the unweighted standard deviation of the rating for standardization. Figure B3 reports the unweighted estimates with raw significance shading, and Figure B4 reports the same estimates with Benjamini–Hochberg adjustment. The direction and approximate magnitude of the BH-significant left-side contrast (*Computer*

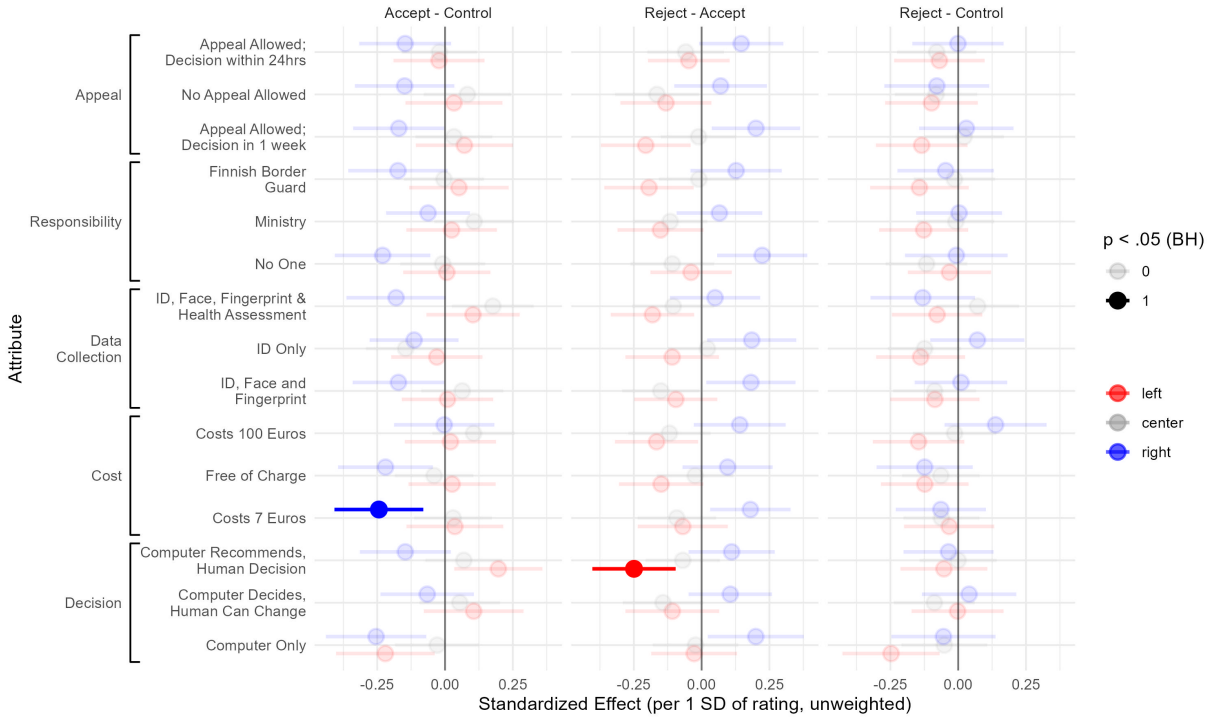
Only, *Reject* vs *Control*) remain in place, and an additional cost-related contrast on the right (*Costs 7 Euros*, *Accept* vs *Control*) reaches BH-adjusted significance only without weights. Weighting therefore masks some contrasts rather than producing them; the substantive conclusions are unchanged, and the broader pattern of ideologically-aligned downweighting of procedural attributes documented in the main analysis is, if anything, more visible without weights. Center-group contrasts that newly reach raw significance in the unweighted estimation do not survive the Benjamini–Hochberg correction and are therefore not interpreted.

Figure B3: Unweighted standardized contrasts



Note: Estimands match Figure 9, but the underlying linear model is estimated without survey weights and standardization uses the unweighted standard deviation of the rating. Significance shading reflects raw $p < .05$. Color denotes political ideology.

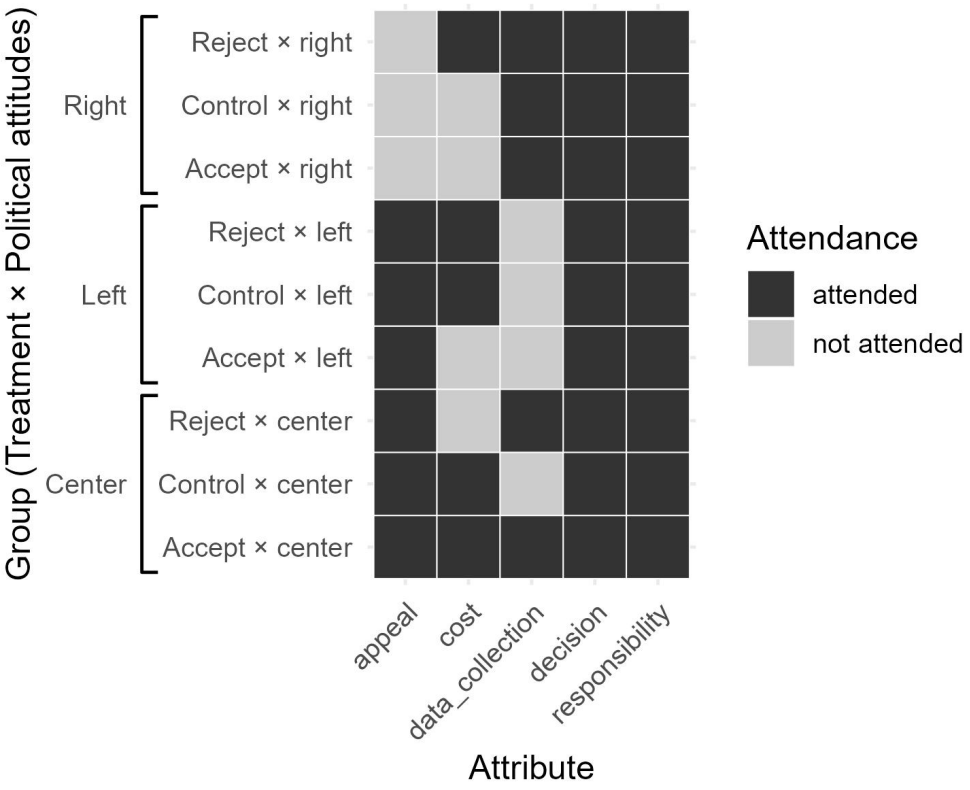
Figure B4: Unweighted standardized contrasts (BH-adjusted)



Note: Same estimates as Figure B3. Significance is BH-adjusted within each feature-by-contrast family ($p_{\text{BH}} < .05$). Color denotes political ideology.

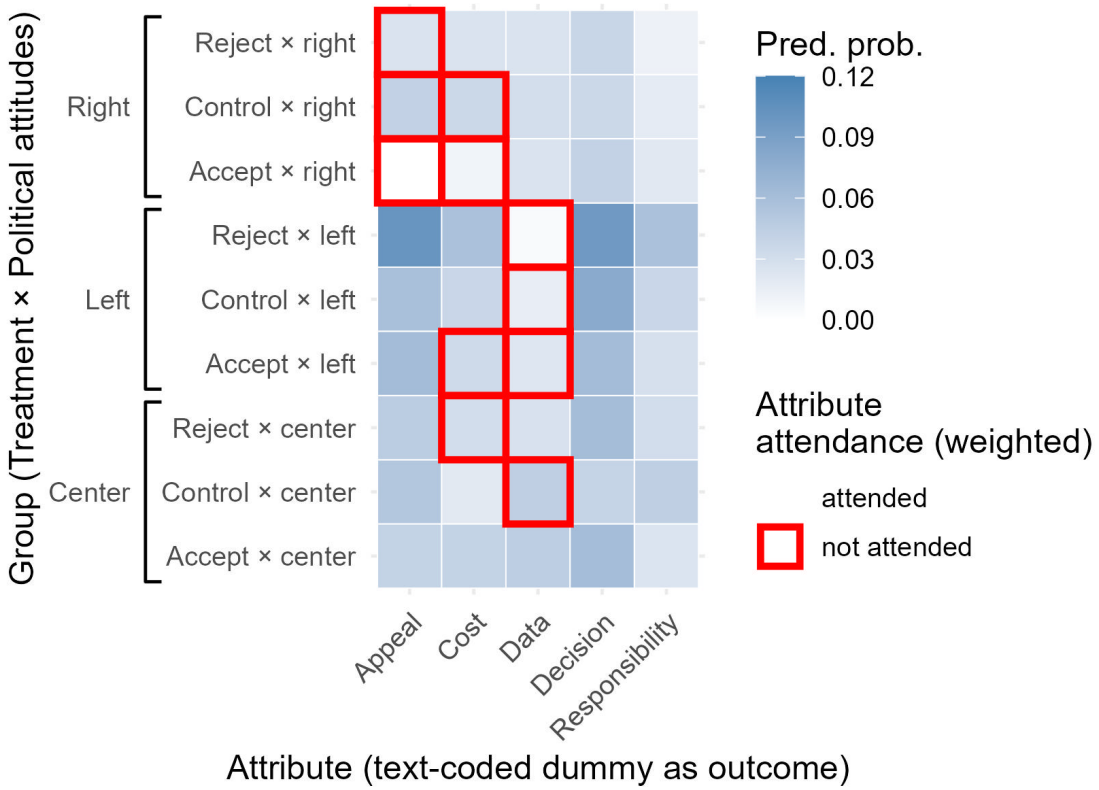
The attribute non-attendance test (Figure 7) is unweighted. As a complementary check, we re-run it under weighting and otherwise hold the procedure fixed (1,000 permutations, BH adjustment within each subgroup). Figure B5 reports the weighted attendance grid and Figure B6 reports the mention-probability heatmap (tiles identical to Figure 8) with the weighted-CRT overlay. Of the 45 subgroup × attribute cells, 42 agree with the unweighted classification; three flip from *attended* to *not attended* under weighting: *Data Collection* for the centre under *Control*; *Cost* for the left under *Accept*; and *Appeal* for the right under *Reject*. The two ideological flips are consistent with motivated reasoning: each group ignores precisely the procedural feature that would otherwise protect the disliked party under its preferred outcome default. Headline conclusions are unchanged, and weighting strengthens rather than weakens the selective-recruitment evidence.

Figure B5: Attribute attendance (weighted)



Note: Weighted-CRT replicate of Figure 7. Test statistic is the deviance reduction of a weighted logistic regression (weights = `combo_weight`); p-values from 1,000 column-permutations of the focal attribute; significance is BH-adjusted within each subgroup. Black cells = attribute attended; grey = not attended.

Figure B6: Mention probability with weighted-CRT attendance overlay



Note: Companion to Figure 8. Tile fill is identical to Figure 8 (unweighted logistic with cluster-robust SEs); the only difference is the attendance overlay: red outlines here mark cells where the *weighted* CRT classifies the attribute as *not attended*, whereas Figure 8 uses the unweighted CRT.