

RESEARCH ARTICLE

Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy

Christopher Starke*  and Marco Lünich 

Department of Social Sciences, University of Düsseldorf, Düsseldorf, Germany

*Corresponding author. E-mail: christopher.starke@uni-duesseldorf.de

Received: 27 March 2020; Revised: 05 August 2020; Accepted: 28 October 2020

Key words: algorithmic decision-making; artificial intelligence; data-driven policy-making; political legitimacy; structured means modeling

Abbreviations: ADM, algorithmic decision-making; AI, artificial intelligence; AI HLEG, Artificial Intelligence High-Level Expert Group; EC, European Commission; EU, European Union; EP, European Parliament; HDM, human decision-making; HyDM, hybrid decision-making; OAP, Open Access Panel; SMM, structured means modeling

Abstract

A lack of political legitimacy undermines the ability of the European Union (EU) to resolve major crises and threatens the stability of the system as a whole. By integrating digital data into political processes, the EU seeks to base decision-making increasingly on sound empirical evidence. In particular, artificial intelligence (AI) systems have the potential to increase political legitimacy by identifying pressing societal issues, forecasting potential policy outcomes, and evaluating policy effectiveness. This paper investigates how citizens' perceptions of EU input, throughput, and output legitimacy are influenced by three distinct decision-making arrangements: (a) independent *human decision-making* by EU politicians; (b) independent *algorithmic decision-making* (ADM) by AI-based systems; and (c) *hybrid decision-making* (HyDM) by EU politicians and AI-based systems together. The results of a preregistered online experiment ($n = 572$) suggest that existing EU decision-making arrangements are still perceived as the most participatory and accessible for citizens (input legitimacy). However, regarding the decision-making process itself (throughput legitimacy) and its policy outcomes (output legitimacy), no difference was observed between the status quo and HyDM. Respondents tend to perceive ADM systems as the sole decision-maker to be illegitimate. The paper discusses the implications of these findings for (a) EU legitimacy and (b) data-driven policy-making and outlines (c) avenues for future research.

Policy Significance Statement

The results of this experimental study suggest that respondents perceive independent algorithmic decision-making (ADM) about the European Union (EU) budget to be illegitimate. EU policy-makers should exercise caution when incorporating ADM systems in the political decision-making process. ADM systems for far-reaching decisions, such as budgeting, may only be used to assist or inform human decision-makers rather than replacing them. An additional takeaway from this study is that the factual and perceived legitimacy of ADM do not necessarily correspond—that is, even ADM systems that produce high-quality outputs, and are implemented transparently and fairly, may still be perceived as illegitimate and might, therefore, be rejected by the electorate. To be socially acceptable, implementation of ADM systems must, therefore, take account of both factual and perceived legitimacy.

1. Introduction

The European Union (EU) currently faces a number of significant crises, most notably the European debt crisis, the distribution of refugees across EU member states, the so-called “Brexit” (withdrawal of the United Kingdom from the EU), and the social and economic consequences of the Covid-19 pandemic. As a result, right-wing populist parties promoting anti-EU messages have gained momentum and threaten the stability of the EU as a whole (Schmidt, 2015). To resolve these crises, the EU must demonstrate responsiveness to citizens’ concerns (*input legitimacy*), effective and transparent procedures (*throughput legitimacy*), and good governance performance (*output legitimacy*) (Weiler, 2012; Schmidt, 2013). However, the EU allegedly lacks legitimacy on all three counts because of a democratic deficit in the institution’s design, the lack of a European identity, the inadequacies of the European public sphere, and the intricacies of producing effective policies for all member states (Follesdal, 2006; Habermas, 2009; Risse, 2014; De Angelis, 2017).

To improve their legitimacy, EU political institutions have increasingly committed to data-driven forms of governance. By integrating digital data into political processes, the EU seeks increasingly to base decision-making on sound empirical evidence (e.g., the Data4Policy project). In particular, algorithmic decision-making (ADM) systems are used to identify pressing societal issues, to forecast potential policy outcomes, to inform the policy process, and to evaluate policy effectiveness (Poel et al., 2018; Algorithm-Watch, 2019). For instance, ADM systems have been shown to successfully support decision-making regarding the socially acceptable distribution of refugees. Trials suggest that this approach increases refugee employment rates by 40–70% as compared to human-led distribution practices (Bansak et al., 2018).

However, little is known about the specific impact of ADM on public perceptions of legitimacy. On the one hand, high public support for digitalization, in general, and autonomous systems, in particular, means that the use of ADM may increase perceived legitimacy (European Commission, 2017). Notably, ADM systems are commonly perceived as true, objective, and accurate and, therefore, capable of reducing human bias in the decision-making process (Lee, 2018). On the other hand, ADM-based policy-making poses a number of novel challenges in terms of perceived legitimacy: (a) Citizens may believe that they have little influence on ADM selection criteria—for instance, which digital data are collected, or on which indicators the algorithm ultimately bases decisions (*input legitimacy*). (b) Citizens may not understand the complex and often opaque technicalities of the ADM process (*throughput legitimacy*). (c) Citizens may doubt that ADM systems can make better decisions than humans, or they may question whether certain decisions produce the desired results (*output legitimacy*).

Few studies have investigated the effects of ADM on perceptions of legitimacy, especially with respect to political decisions. To date, empirical studies have tended to focus on public sector areas, such as education and health, evaluating the effects of ADM as compared to human decision-making (HDM) in terms of variables, such as fairness and trust (Lee, 2018; Araujo et al., 2020; Marcinkowski et al., 2020). While those studies investigate decisions that affect individual citizens (e.g., decisions about loans or university admissions), the political context examined in this study refers to decisions that affect societal groups, or even society as a whole. To bridge this research gap, the present study investigates the extent to which the use of ADM influences the perceived legitimacy of policy-making at EU level. For that, we use EU budgeting decisions as a case in point. Even though fully or semiautomated ADM is unlikely to be implemented in EU decision-making in the near future, it is key to gain empirical insights into their potential consequences. This point refers to the so-called “Collingridge dilemma,” which states that every new technology is accompanied with two competing concerns. “On one hand, regulations are difficult to develop at an early technological stage because their consequences are difficult to predict. On the other hand, if regulations are postponed until the technology is widely used, then the recommendations come too late” (Awad et al., 2020, p. 53). Addressing this dilemma, the study extends the existing literature in three respects. (1) It provides novel insights into the potential of ADM to exacerbate or alleviate the EU’s perceived legitimacy deficit. (2) It clarifies the effects of three distinct decision-making arrangements on perceptions of legitimacy: (a) independent decision-making by EU politicians or HDM; (b) independent decision-making by AI-based systems or ADM; and (c) hybrid decision-making

(HyDM), where politicians select among decisions suggested by ADM systems. (3) Using structural means modeling (SMM) to analyze citizens' perceptions, the study proposes a general measure of input, throughput, and output legitimacy.

2. A Crisis of EU Legitimacy? Input, Throughput, Output

In making effective decisions to resolve major crises, the EU's actions depend on political legitimacy. According to Gurr, "governance can be considered legitimate in so far as its subjects regard it as proper and deserving of support" (Gurr, 1971, p. 185). In his seminal work on legitimacy, Scharpf (1999) distinguished between two dimensions of legitimacy; *input legitimacy* and *output legitimacy*. Input legitimacy is characterized as "responsiveness to citizen concerns as a result of participation by the people" (Schmidt, 2013, p. 2). It, thus, depends on free and fair elections, high voter turnout, and lively political debate in the public sphere (Scharpf, 1999). *Output legitimacy* refers to "the effectiveness of the EU's policy outcomes for the people" (Schmidt, 2013, p. 2)—that is, the EU's problem-solving capacity in pursuing desired goals, such as preserving peace, ensuring security, protecting the environment, and fostering prosperity (Follesdal, 2006). Moving beyond this dichotomy, some scholars (Schmidt, 2013; Schmidt and Wood, 2019) have added throughput as a third dimension of legitimacy, referring to the accountability, efficacy, and transparency of EU policy-makers and their "inclusiveness and openness to consultation with the people" (Schmidt, 2013, p. 2). Also referred to as the "black box" (Steffek, 2019, p. 1), throughput legitimacy encompasses the political practices and processes of EU institutions in turning citizen input into policy output (Steffek, 2019; Schmidt and Wood, 2019).

Ever since the EU was founded, and especially since the failed Constitutional Treaty referenda in France and the Netherlands in 2005, European integration has been dogged by criticisms that the EU lacks legitimacy. Most scholars point to the democratic deficit, the lack of a European identity, and an inadequate public sphere as primary reasons for this alleged crisis of legitimacy (Follesdal and Hix, 2006; Habermas, 2009). The debate centers on four arguments (Follesdal, 2006; Follesdal and Hix, 2006; Holzhaecker, 2007; De Angelis, 2017). First, among key EU political institutions, only the European Parliament (EP) is legitimized by European citizens by means of elections, but scholars argue that the EP is too weak in comparison to the European Commission (EC) (Follesdal and Hix, 2006). While continuous reform of EU treaties has substantially strengthened the EP's role within the institutional design of the EU, it still lacks the power to initiate legislation (Holzhaecker, 2007). Second, the EU's institutional design gives national governments pivotal power over the Council of the EU and the EC. However, as those actors are somewhat exempt from parliamentary scrutiny by the EP and national parliaments, there is a deficit in democratic checks and balances (Follesdal and Hix, 2006). Third, the European elections are not sufficiently "European" (Follesdal, 2006)—that is, "they are not about the personalities and parties at the European level or the direction of the EU policy agenda" (Follesdal and Hix, 2006, p. 536). Instead, national politicians, parties, and issues still dominate campaigns and remain crucial in citizens' voting decisions (Hobolt and Wittrock, 2011). Finally, European citizens are arguably too detached from the EU (Follesdal and Hix, 2006). Public opinion research suggests that although a sense of European identity, trust in European institutions, and satisfaction with EU democracy are on the rise, these pale in comparison to the corresponding scores at national level (Risse, 2014; European Commission, 2019b). Consequently, scholars have argued that the EU lacks a European demos—that is, "a strong sense of community and loyalty among a political group" (Risse, 2014, p. 1207). In addition, the alleged lack of a European public sphere that would enable communication and debate around political issues lends further credence to the claim that the EU suffers from insufficient citizen participation (Habermas, 2009; Kleinen-von Königslöw, 2012).

As all four arguments primarily question the EU's input and throughput legitimacy, many have argued that output is the stronghold for the EU legitimacy. According to Scharpf, "the EU has developed considerable effectiveness as a regulatory authority" (Scharpf, 2009, p. 177). In that regard, the EU enables member states to implement policies that they would otherwise be unable to advance, especially in relation to global policy issues (Menon and Weatherill, 2008). Weiler contended that output legitimacy "is

part of the very ethos of the Commission” (Weiler, 2012, p. 828), but recent crises have also challenged this view; for instance, the austerity measures imposed on debtor states had detrimental effects on the lives of many European citizens (De Angelis, 2017). Debate about the EU’s alleged legitimacy crisis centers primarily on institutional shortcomings in the political system, and public perceptions of legitimacy are neglected. However, Jones (2009) claimed that subjective perceptions are often more important than the normative criteria themselves.

3. ADM for Policy-making in the EU?

In recent years, EU institutions have increasingly sought to address this perceived deficit of legitimacy through evidence-based policy-making: “Against the backdrop of multiple crises, policymakers seem ever more inclined to legitimize specific ways of action by referring to ‘hard’ scientific evidence suggesting that a particular initiative will eventually yield the desired outcomes” (Rieder and Simon, 2016, p. 1). This push for numerical evidence comes at a time when the computerization of society has precipitated the creation and storage of vast amounts of digital data. According to Boyd and Crawford, the so-called big data “offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy” (Boyd and Crawford, 2012, p. 663). It is often presumed that the more data are analyzed, that is, preferably all available data as “N = all” trumps sampling, the greater is the potential to gain insights and receive the best result (Mayer-Schönberger and Cukier, 2013). Digital data are collected, accessed, and analyzed in real time, leading to substantial advances in analytics, modeling, and dynamic visualization (Craglia et al., 2018; Poel et al., 2018; Verhulst et al., 2019). This transformation of real-world phenomena into digital data is expected to provide a timely and undistorted view of societal mechanisms and institutions.

Lately, public discourse around the potential of computerization and big data has included a renewed focus on Artificial Intelligence (AI). According to Katz,

“AI stands for a confused mix of terms—such as ‘big data,’ ‘machine learning,’ or ‘deep learning’—whose common denominator is the use of expensive computing power to analyze massive centralized data. (...) It’s a vision in which truth emerges from big data, where more metrics always need to be imposed upon human endeavors, and where inexorable progress in technology can ‘solve’ humanity’s problems” (Katz, 2017, p. 2).

Indeed, the increasing availability of digital data in combination with significant advances in computing power have underpinned the recent emergence of many successful AI applications, such as self-driving cars and natural language generation, and face recognition. This has, in turn, raised expectations regarding the use of AI for evidence-based or data-driven policy-making (Esty and Rushing, 2007; Giest, 2017; Poel et al., 2018). To exploit technological developments and increasing data availability for policy-making purposes, the EC contracted the Data4Policy project (Rubinstein et al., 2016), arguing that “data technologies are amongst the valuable tools that policymakers have at hand for informing the policy process, from identifying issues, to designing their intervention and monitoring results” (European Commission, 2019a, para. 1).

In that context, van Veenstra and Kotterink (2017, p. 101) noted that “data-driven policy making is not only expected to result in better policies, but also aims to create legitimacy.” Recent reports suggest that ADM systems are already in use throughout the EU to deliver public services, optimize traffic flows, or identify social fraud (Poel et al., 2018; AlgorithmWatch, 2019). Case studies confirm that ADM systems can indeed contribute to better policy (Bansak et al., 2018), using big data to identify emerging issues, to foresee demand for political action, to monitor social problems, and to design policy options (Poel et al., 2018; Verhulst et al., 2019). To that extent, data-driven systems can potentially contribute to the increased legitimacy of *input* (by enabling new forms of citizen participation,) of *throughput* (by making the political process more transparent), and of *output* (by increasing the quality of policies and outcomes).

Yet, despite these promising indications, numerous examples of AI’s pitfalls in political decision-making exist. For instance, a recent report by the research institute AI NOW revealed that ADM systems

may falsely accuse citizens of fraud, arbitrarily exclude them from food support programs, or mistakenly reduce their disability benefits. Incorrect classification by ADM systems has led to a wave of lawsuits against the US government at federal and state levels, undermining both the much vaunted cost efficiency of automated systems and the perceived legitimacy of political decision-making as a whole (Richardson et al., 2019). As a consequence, in tackling the issues that come with the implementation of AI systems in society, the EU has appointed an AI High-Level Expert Group that developed “Ethics Guidelines for Trustworthy AI” (Artificial Intelligence High-Level Expert Group, 2019). All seven key requirements introduced with the guidelines also address issues at the heart of potential legitimacy concerns of the public with regard to proposed ADM systems in governance. Most prominently, when it comes to policy-making, ADM raises important questions concerning the need for *human agency and oversight, transparency, diversity, non-discrimination, and fairness, as well as accountability*. In terms of the three dimensions of legitimacy, ADM systems pose the following challenges: (a) On the input dimension, citizens may lack insight into or influence over the criteria or data that intelligent algorithms use to make decisions. This may undermine fundamental democratic values such as civic participation or representation. (b) On the throughput dimension, citizens may be unable to comprehend the complex and often inscrutable logic that underpins algorithmic predictions, recommendations, or decisions. The corresponding opacity of the decision-making process may violate the due process principle, for example, that citizens receive explanations for political decisions and have the opportunity to file complaints or even go to court. (3) On the output dimension, citizens may fundamentally doubt whether ADM systems actually contribute to better and/or more efficient policy. This may conflict with key democratic principles, such as non-discrimination.

As with all technological innovations, success or failure depends greatly on all stakeholders’ participation and acceptance (Bauer, 1995). In the present context, those stakeholders include EU institutions, bureaucracies, and regulators who may favor the introduction of ADM systems in policy-making, and the electoral body of voters who legitimize proposed policies and implementation. While there are no existing accounts of citizens’ perceptions of ADM systems in the context of political decision-making in the EU, survey data provide some initial insights. Several Eurobarometer surveys have shown that public perception of digital technologies is broadly positive throughout the EU, especially when compared to perceptions of other mega-technologies, such as nuclear power, biotechnology, or gene editing (European Commission, 2015, 2017). According to a recent survey commissioned by the Center for the Governance of Change, “25% of Europeans are somewhat or totally in favor of letting an artificial intelligence make important decisions about the running of their country” (Rubio and Lastra, 2019, p. 10). On that basis, it seems likely that demands to embed AI in the political process will increase, and that political programs will respond to those demands.

4. Hypotheses

The key objective of this study was to investigate whether and to what extent ADM systems in policy-making influence public perceptions of EU input, throughput, and output legitimacy. Previous empirical studies have suggested that different decision-making arrangements (e.g., formal versus descriptive representation, direct voting versus deliberation) can differ significantly in terms of their perceived legitimacy (Esaiasson et al., 2012; Persson et al., 2013; Arnesen and Peters, 2018; Arnesen et al., 2019). However, as those studies did not specifically investigate the potential effects of ADM systems, the present study sought to distinguish between three different decision-making arrangements: (a) independent decision-making by EU politicians (*HDM*); (b) independent decision-making by ADM systems (*ADM*); and (c) *HyDM* by politicians, based on suggestions made by ADM systems. The reasoning for our hypotheses refers to the specific decision-making process tested in this study, namely the distribution of the EU budget to different policy areas.

With regard to perceived input legitimacy, we contend that respondents are likely to perceive the current decision-making process as more legitimate than processes that rely partly or completely on ADM. The primary reason for this assumption is that transferring some (*HyDM*) or all (*ADM*) authority

about EU budgeting decisions to algorithms is likely to diminish the role of democratically elected institutions, which would undermine a fundamental pillar of representative democracies. As a result, algorithmic or hybrid decision systems would probably marginalize the opportunities for citizen participation and thereby decrease input legitimacy. This reasoning, of course, applies primarily to those technologies that directly decide on policy, instead of being involved in less far-reaching stages of the policy cycle, such as agenda setting or policy evaluation (Verhulst et al., 2019). Furthermore, as Barocas and Selbst (2016) point out, minorities and other disadvantaged social groups are often underrepresented in existing digital data, making them vulnerable to be disregarded by ADM. Even though first evidence indicates that data science has some potential to increase citizen participation and representation by assessing policy preferences via opinion mining of social media data (Ceron and Negri, 2016; Sluban and Battiston, 2017), it is unlikely that citizens perceive such indirect and rather unknown forms of political participation to be more legitimate than currently existing democratic procedures. On that basis, we tested the following preregistered hypotheses (see preregistration at Open Science Framework (OSF)):

H1a: HDM leads to higher perceived input legitimacy as compared to ADM.

H1b: HDM leads to higher perceived input legitimacy as compared to HyDM.

H1c: HyDM leads to higher perceived input legitimacy as compared to ADM.

With regard to perceived throughput legitimacy, we argue that implementation of ADM leads to lower levels of perceived legitimacy as compared to the existing political process. Even though EU decision-making processes are often criticized for their lack of transparency, ADM systems suffer from the same deficiency, as they are themselves considered to be a “black box” (Wachter et al., 2018). The extent of transparency of self-learning systems, however, is a major driver of public perceptions of legitimacy (De Fine Licht and De Fine Licht, 2020). A recent EC report, therefore, stressed the urgent need to make ADM more explainable and transparent (Craglia et al., 2018) on the grounds that such systems are typically too complex for the layperson to understand and are largely unable to give proper justifications for decisions. Moreover, their ability to mitigate discrimination in decision-making processes is still subject to contested debates in the literature. While some empirical evidence suggests that ADM may lead to more positive perceptions of procedural fairness (Marcinkowski et al., 2020), ADM is also prone to reproduce and even exacerbate existing societal biases (Barocas and Selbst, 2016). Furthermore, ADM systems lack public accountability because citizens do not know who to turn to regarding policy or administrative failures. Indeed, preliminary empirical evidence suggests that activities that require human skills are perceived as fairer and more trustworthy when executed by humans rather than algorithms (Lee, 2018). On that basis, we formulated the following hypotheses.

H2a: HDM leads to higher perceived throughput legitimacy as compared to ADM.

H2b: HDM leads to higher perceived throughput legitimacy as compared to HyDM.

H2c: HyDM leads to higher perceived throughput legitimacy as compared to ADM.

Several scholars suggest that the EU already legitimizes itself primarily via the output dimension (Scharpf, 2009; Weiler, 2012) due to the aforementioned democratic deficit on the input dimension. Below, we argue why implementing algorithmic or hybrid decision system would mean that the EU is doubling down on output legitimacy. Perceived output legitimacy comprises two key dimensions: citizens’ perceptions of whether political decisions can attain predefined goals (e.g., economic growth, environmental sustainability), and the subjective favorability of such decisions. Assessment of the perceived quality of political output involves both dimensions, and this is where ADM systems are said to have a distinct advantage over human decision-makers, as they can produce novel insights from vast amounts of digital data that would be impossible when relying solely on human intelligence (Boyd and Crawford, 2012). Empirical studies comparing public perceptions of ADM and HDM seem to support this assumption; looking at proxies for legitimacy, ADM systems are evaluated as fairer in distributive terms than HDM (Marcinkowski et al., 2020), especially in high impact situations (Araujo et al., 2020). Building on these empirical findings, we further argue that citizens perceive ADM systems to be most

legitimate when they operate under the scrutiny of democratically elected institutions. Thus, we formulated the following hypotheses.

- H3a: HDM leads to lower perceived goal attainment as compared to ADM.
 H3b: HDM leads to lower perceived goal attainment as compared to HyDM.
 H3c: HyDM leads to higher perceived goal attainment as compared to ADM.
- H4a: HDM leads to lower decision favorability as compared to ADM.
 H4b: HDM leads to lower decision favorability as compared to HyDM.
 H4c: HyDM leads to higher decision favorability as compared to ADM.

5. Method

To test these hypotheses, we conducted an online experiment, applying a between-subjects-design using one factor with three levels: (a) EU politicians making decisions independently (cond^{HDM}); (b) ADM systems making decisions independently (cond^{ADM}); and (c) ADM systems suggesting decisions to be passed by EU politicians (cond^{HyDM}) (see preregistration¹ at OSF). All measurements and stimulus material and the questionnaire's basic functionality were thoroughly tested in multiple pretests involving 321 respondents in total.

5.1. Sample

Respondents were recruited through the noncommercial SoSci Open Access Panel (OAP) during the period April 8–22, 2019. In accordance with German law, SoSci OAP registration involves a double opt-in process, in which panelists first sign up using an email address and must then activate their account and confirm pool membership (Leiner, 2016). Although the SoSci OAP is not representative in terms of sociodemographic variables, its key advantage is participant motivation; as respondents are not compensated for survey participation, their main motivation is topic interest, which is a crucial indicator of data quality (Brüggen et al., 2011). In addition, all questionnaires using the SoSci OAP must first undergo rigorous peer review, so ensuring “major improvements to the instrument before data are collected” (Leiner, 2016, p. 373).

Using Soper's (2019) a priori sample size calculator for structural equation modeling, we determined an optimal sample size of 520, based on the results from a pretest conducted 10 weeks before final data collection. Altogether, 3,000 members of the SoSci OAP were invited by e-mail to participate in the study. In total, 612 respondents completed the questionnaire. A thorough two-step cleaning process was applied for quality control purposes. The first step excluded respondents who failed an attention check regarding the target topic ($n = 14$). In the second step, using the DEG_Time variable (Leiner, 2019), each respondent accumulated minus points for completing single questions or the whole questionnaire too quickly. As the SoSci OAP administrators recommend a threshold score of 50 for rigorous filtering, all respondents with a minus point score of 50, or higher, were excluded from the analysis ($n = 26$). After filtering, the sample comprised 572 respondents—a response rate of 19.1%.

5.2. Treatment conditions (independent variable)

Respondents were randomly assigned to one of the three conditions and received a short text (ca. 250 words per condition) about the decision-making process regarding distribution of the annual EU budget. The stimulus material also included a pie chart, showing budget allocation for different policy areas to inform readers about the range of budgetary items and their distribution in the actual EU budget. For reasons of validity, the text was adapted from the official EU website (European Union, 2019). While the pie chart was identical for all three conditions, the closing paragraph of the text was edited to reflect

¹ Table 4 in the Appendix accounts for all deviations from preregistration.

manipulation of the independent variable: (a) decisions made by politicians of EU institutions only—the status quo (cond^{HDM}; $n = 182$); (b) decisions made by ADM only (cond^{ADM}; $n = 204$); and (c) decisions suggested by ADM and subsequently passed by politicians of EU institutions (cond^{HyDM}; $n = 186$).² In the two latter stimuli that make reference to ADM, we did not mention the specific criteria as to how the AI system would be instructed what an optimal distribution might look like. The wording, which makes reference to “all available data being used to deliver optimal results,” thus merely connects to the abovementioned popular trope that ever larger data sets offer superior results that are beyond human abilities and comprehension. Randomization checks suggest that no differences between the distribution of respondents to the conditions exist in terms of age ($M = 47.26$, $SD = 15.99$; $F(2, 569) = .182$, $p = .834$); gender (female = 45.8 %, male = 53.5 %, diverse = .7 %; $\chi^2(4) = 1.89$, $p = .757$); education (non-tertiary education = 40.6 %, tertiary education = 59.4 %; $\chi^2(2) = .844$, $p = .656$) and political interest ($M = 3.85$, $SD = .82$; $F(2, 569) = .491$, $p = .608$).³ Respondents were not deceived into thinking that cond^{ADM} and cond^{HyDM} are existing decision-making procedures in the EU as it was explicitly stressed that the scenario at hand was only a *potential* decision-making process. At the end of the survey, they were debriefed about the research interest of the study.

5.3. Manipulation check

All respondents answered two items that served as manipulation checks to validate that respondents perceived the differences in the respective conditions. First, perceived technical automation of the decision-making process was assessed by responses (on a five-point Likert scale) to the question *How technically automated was the decision-making process?* The results indicated a significant difference among the three conditions ($F(2, 524) = 389.71$, $p < .001$). Using a Games-Howell post hoc test, cond^{HDM} ($M = 2.11$; $SD = .97$), cond^{ADM} ($M = 4.52$; $SD = .77$), and cond^{HyDM} ($M = 4.16$; $SD = .80$) were found to differ significantly from each other, confirming that respondents recognized the extent to which the described decision-making processes were technically automated.

The perceived involvement of political actors and institutions in the different decision-making arrangements was measured by responses (on a five-point Likert scale) to the question: *What role did politicians or political institutions play in the decision-making process?* Again, there were significant differences among the three conditions ($F(2, 548) = 161.98$, $p < .001$). Using a Games-Howell post hoc

² A translation of the stimulus material can be found in the Appendix. The crucial sentences that addressed the distinct procedures of decision-making were introduced as follows: “The decision on the budget for each year is made in two main steps: ...” The description in the condition *HDM* then reads as follows: “(a) In a first step, the European Commission prepares a draft budget and submits it to the governments of the member states—represented in the Council of the EU—and to the democratically elected European Parliament. (b) The Commission’s budget proposal is then debated, negotiated and, if necessary, adapted in the European Council and the European Parliament. Once the proposal has been accepted by all the institutions involved, the budget for the following year is ready.” In the condition *ADM* the description reads as follows: “(a) As a first step, high-performance computers of the European Court of Auditors bring together all data available at EU level. Examples are available structural and administrative data from the EU and individual member states, economic and social forecasting models, and other data from business and science. On the basis of large data sets, an ‘Artificial Intelligence’ calculates the optimal distribution key of resources for the individual areas of the EU budget within a few hours with the help of the so-called machine learning applications. (b) The resulting model is audited by the Court of Auditors and then presented to the President of the European Commission and the Commissioner for Financial Programming and Budget for signature. Thus, the budget for the following year is ready.” In the last condition *HyDM* it reads as follows: “(a) In a first step, high-performance computers of the European Court of Auditors bring together all data available at EU level. Examples are available structural and administrative data from the EU and individual Member States, economic and social forecasting models, and other data from business and science. On the basis of large data sets, an ‘artificial intelligence’ calculates the optimal distribution key of resources for the individual areas of the EU budget within a few hours with the help of so-called machine learning applications. (b) The budget proposal is then debated, negotiated and, if necessary, adapted in the European Commission, the European Council and the European Parliament. Once the proposal has been accepted by all the institutions involved, the budget for the following year is ready.”

³ The variables *age*, *gender*, and *education* were measured with single-item questions. To measure *political interest*, we adopted three items from Starke et al. (2020): (a) interest in German politics, (b) interest in European politics, (c) interest in non-European politics ($\alpha = .875$).

test, cond^{HDM} ($M = 4.45$; $SD = .88$), cond^{ADM} ($M = 2.63$; $SD = .99$), and $\text{cond}^{\text{HyDM}}$ ($M = 3.41$; $SD = 1.04$), all were found to differ significantly from each other, confirming that respondents recognized the degree to which political actors and institutions were involved in each condition.

5.4. Measures

As Persson et al. (2013, p. 391) rightly noted, “legitimacy is an inherently abstract concept that is hard to measure directly.” To account for this difficulty, measures for *input legitimacy* (dV1), *throughput legitimacy* (dV2), and *output legitimacy* using the two dependent variables, *goal attainment* (dV3), and *decision favorability* (dV4), were thoroughly pretested and validated. All items used in the analysis were measured on a five-point Likert scale, ranging from 1 (*do not agree*) to 5 (*agree*) and including the residual category *don't know*.⁴ The factor validity of all measures was assessed using Cronbach's alpha (α) and average variance extracted (AVE) (Table 1).

Input Legitimacy (dV1). Three items were used to measure perceived input legitimacy, using wording adapted⁵ from previous studies (Lindgren and Persson, 2010; Persson et al., 2013; Colquitt and Rodell, 2015): (a) *All citizens had the opportunity to participate in the decision-making process* (IL1); (b) *People like me could voice their opinions in the decision-making process* (IL2); and (c) *People like me could influence the decision-making process* (IL3). All items were randomized and used as indicators of a latent variable in the analysis.

Throughput Legitimacy (dV2). To measure perceived throughput legitimacy, three items were adapted from Werner and Marien (2018). Respondents were asked to indicate to what extent they perceived the decision-making process described in the stimulus material as (a) *fair* (TL1); (b) *satisfactory* (TL2); and (c) *appropriate* (TL3). All items were randomized and used as indicators of a latent variable in the analysis.

Goal Attainment (dV3). To measure perceived goal attainment, which is considered an important pillar of output legitimacy (Lindgren and Persson, 2010), respondents were asked to indicate to what extent they believed the decision-making process could achieve the goals referred to in the stimulus text (adapted from the official EU website): (a) *Better development of transport routes, energy networks, and communication links between EU countries* (GA1); (b) *Improved protection of the environment throughout Europe* (GA2); (c) *An increase in the global competitiveness of the European economy* (GA3); and (d) *Promoting cross-border associations of European scientists and researchers* (GA4) (European Union, 2019). The order of the items was randomized. As the four goals can be independently attained, the underlying construct is not one-dimensional and reflective. For that reason, we computed a mean index for goal attainment that was used as a manifest variable in the analysis.

Table 1. Descriptives and factorial validity

| | Overall Mean (SD) | cond^{HDM} M (SD) | cond^{ADM} M (SD) | $\text{cond}^{\text{HyDM}}$ M (SD) | Cronbach's alpha | AVE |
|--------------------------|----------------------|--|--|---|---------------------|------|
| Input legitimacy | 1.66 (0.85) | 1.90 (.92) | 1.45 (.75) | 1.66 (.85) | .772 | .542 |
| Throughput legitimacy | 2.76 (1.00) | 2.92 (1.04) | 2.55 (1.00) | 2.84 (.92) | .881 | .715 |
| Goal attainment | 3.28 (.91) | 3.37 (.83) | 3.18 (.92) | 3.31 (.95) | | |
| Decision acceptance | 2.96 (1.05) | 3.15 (.95) | 2.67 (1.11) | 3.12 (.99) | .875 | .747 |

Abbreviation: AVE, average variance extracted.

⁴ The residual category “don't know” was treated as missing values in the analysis. As a consequence, Full Information Maximum Likelihood was used to estimate the model.

⁵ All items were translated from English into German.

Decision Favorability (dV4). In the existing literature, decision acceptance or favorability is commonly used as a measure of legitimacy (Esaiaasson et al., 2012; Werner and Marien, 2018). Conceptualizing decision favorability as the second key pillar of output legitimacy, we used three items to measure dV4. Two of these items were adopted from Werner and Marien's (2018) four-item scale: (a) *I accept the decision* (DF1), and (b) *I agree with the decision* (DF2). As the other two items in their scale refer to the concept of reactance, we opted to formulate one additional item: (c) *The decision satisfies me* (DF3). All items were randomized and used as indicators of a latent variable in the analysis.

5.5. Data analysis

The analysis employed SMM, incorporating all four variables in a single model. As this approach takes account of measurement error due to latent variables, it was adopted in preference to traditional analysis of variance (Breitsohl, 2019). To test the hypotheses, we compared means between groups, using critical ratios for differences between parameters in the specified model. All statistical analyses were performed using AMOS 23. Because of missing data, Full Information Maximum Likelihood estimation was used in conjunction with estimation of means and intercepts (Kline, 2016). Full model fit was assessed using a chi-square test and RMSEA (lower and upper bound of the 90% confidence interval, PClose value), along with the Tucker–Lewis-Index (TLI) measure of goodness of fit (Holbert and Stephenson, 2002; van de Schoot et al., 2012). Differences in means were investigated by obtaining critical ratios (CR); for $CR > 1.96$ or $CR < -1.96$, respectively, the parameter difference indicated two-sided statistical significance at the 5% level.

As the experimental design compared three groups, we tested the measurement models of all latent factors for measurement invariance (van de Schoot et al., 2012; Kline, 2016). This test was necessary to assess whether factor loadings (metric invariance) and item intercepts (scalar invariance) were equal across groups. This “strong invariance” is a necessary precondition to confirm that latent factors are measuring the same construct and can be meaningfully compared across groups (Widaman and Reise, 1997). The chi-square-difference test for strong measurement invariance in Table 2 shows that the assumptions of metric and scalar invariance are violated. Subsequent testing of indicator items identified indicator IL02 as non-invariant. On that basis, a model with only partial invariance was estimated, freeing both the indicator loading and item intercept constraints of IL02. A chi-square-difference test for partial measurement invariance showed better model fit as compared to the configural model ($\Delta\chi^2 = 18.034$, $\Delta df = 16$; $p = .322$). The final model with partial measurement invariance fit the data well ($\chi^2(106) = 173.299$, $p < .001$; RMSEA = .033 (.024; .042); PClose = .999; TLI = .966). The latent means of the specified model with partial invariance were constrained to zero in cond^{HDM}. On that basis, the first condition, in which only EU politicians made decisions about the EU budget, was used as the reference group when reporting the results of group comparisons.

6. Results

Construct means are shown in Table 3. In addition, based on a transformation of Hedge's g , a standardized effect size r as proposed by Steimetz et al. (Steimetz et al., 2009) was manually calculated. This is also reported in Table 3.

Table 2. Descriptives and factorial validity

| | χ^2 | df | p | TLI | RMSEA | PClose |
|-------------------------------------|----------|-----|-------|------|-------------------|--------|
| Configural model | 155.265 | 90 | <.001 | .961 | .036 (.026; .045) | .995 |
| Metric invariance | 198.964 | 102 | <.001 | .949 | .041 (.032; .049) | .963 |
| Scalar invariance | 219.483 | 114 | <.001 | .950 | .040 (.032; .048) | .978 |
| Final model with partial invariance | 173.299 | 106 | <.001 | .966 | .033 (.024; .042) | .999 |

Table 3. Comparisons of structured means of the legitimacy dimensions

| | Means | | | Effect sizes (<i>r</i>) | | |
|-----------------------|---------------------|---------------------|----------------------|---|--|--|
| | cond ^{HDM} | cond ^{ADM} | cond ^{HyDM} | cond ^{HDM} versus cond ^{ADM} | cond ^{ADM} versus cond ^{HyDM} | cond ^{HDM} versus cond ^{HyDM} |
| Input legitimacy | 0 | -.494 | -.262 | .25 | .13 | .13 |
| Throughput legitimacy | 0a | -.346 | .070a | .15 | .19 | .00 |
| Goal attainment | 3.37a | 3.15b | 3.29a,b | .11 | .00 | .00 |
| Decision acceptance | 0a | -.347 | .025a | .15 | .15 | .00 |

Note. Means not sharing any letter are significantly different by the test of critical ratios at the 5% level of significance.

With regard to perceived input legitimacy, we assumed that this would be highest in cond^{HDM} (in which only EU politicians made budget decisions) and lowest in cond^{ADM} (decisions based solely on ADM), with cond^{HyDM} (ADM and EU politicians combined) somewhere between the two. The results indicate that respondents perceived input legitimacy as significantly lower in cond^{ADM} ($\Delta M = -.494$, $p < .001$) and cond^{HyDM} ($\Delta M = -.262$, $p = .011$) compared to cond^{HDM}. As the difference between these conditions was also significant ($\Delta M = -.232$, $p = .009$), hypotheses H1a, H1b, and H1c were supported.

For perceived throughput legitimacy, the results indicate (as expected) that cond^{ADM} was perceived as significantly less legitimate than cond^{HDM} ($\Delta M = -.346$, $p < .001$). No difference was observed between cond^{HDM} and cond^{HyDM} ($\Delta M = -.070$, $p = .481$), but cond^{HyDM} differed significantly from cond^{ADM} ($\Delta M = -.276$, $p = .004$). As a consequence, hypotheses H2a and H2b were supported while H2c was rejected.

In contrast to input and throughput legitimacy, we assumed that cond^{HDM} would score lower than the other two conditions for perceived goal attainment, and that cond^{HyDM} would score higher than the other two conditions. In fact, cond^{HDM} returned the highest mean ($M = 3.37$) and did not differ significantly from cond^{HyDM} ($M = 3.29$; $\Delta M = .083$, $p = .383$). Again, cond^{ADM} scored lowest ($M = 3.15$) and differed significantly from cond^{HDM} ($\Delta M = .223$, $p = .014$) but not from cond^{HyDM} ($\Delta M = -.014$, $p = .151$). These results found no support for hypotheses H3a, H3b, or H3c and even ran counter to the assumptions of H3a.

We anticipated that perceived decision favorability would be highest for cond^{HyDM}, lowest for cond^{HDM}, with cond^{ADM} somewhere between the two. In fact, cond^{ADM} scored significantly lower than cond^{HDM} ($\Delta M = -.347$, $p < .001$) and significantly lower than cond^{HyDM} ($\Delta M = -.372$, $p < .001$). There was no significant difference between cond^{HDM} and cond^{HyDM} ($\Delta M = -.25$, $p = .809$). As a result, H4a and H4b were rejected while H4c was accepted.

7. Discussion

This paper answers the call for more empirical research to understand the nexus of ADM for political decision-making and its perceived legitimacy. How does the integration of AI into policy-making influence people's perceptions of the legitimacy of the decision-making process? In pursuit of preliminary answers to this question, the results of a preregistered online experiment that systematically manipulated levels of autonomy given to an algorithm in EU policy-making yielded three main insights. First, existing EU decision-making arrangements were considered the most participatory—that is, they scored highest on input legitimacy. Second, in terms of process quality (throughput legitimacy) and outcome quality (output legitimacy), no differences were observed between existing decision-making arrangements and HyDM. Finally, decision-making, informed solely by ADM, was perceived as the least legitimate arrangement across all three legitimacy dimensions. In the following sections, we consider the implications of these findings for EU legitimacy, data-driven policy-making, and avenues for future research.

7.1. *Implications for the legitimacy of the EU*

Our findings lend further credence to previous assertions that the EU lacks political legitimacy (Holzhacker, 2007), in that current decision-making arrangements, which solely involve EU politicians, score low on input legitimacy ($M = 1.90$ on a five-point Likert scale). This finding speaks to a previously noted democratic deficit (Follesdal, 2006; Follesdal and Hix, 2006). The present results further reveal that ADM systems do not seem to offer an appropriate remedy; on the contrary, it seems that such systems may even exacerbate the problem, as the existing process is still perceived as having greater input legitimacy than arrangements based wholly (ADM condition) or partly (HyDM condition) on ADM systems. It appears that ADM systems fail to engage citizens in the decision-making process or to make their voices heard. Implementing ADM technologies to assist or replace human political actors is seen as less democratic than the status quo, even though incumbent decision-makers such as the European Commission themselves lack democratic legitimacy. One plausible explanation for this finding is that ADM systems are perceived even more technocratic and detached from voters than EU politicians. For that reason, citizens favor human decision-makers when dealing with human tasks, aligning with earlier findings by Lee (2018).

As the EU depends heavily on public approval, it seems important to explore alternative ways of increasing its legitimacy. Rather than leaving political decisions to ADM systems, less far-reaching forms of data-driven policy-making might help to achieve this goal. Beyond decision making, data-driven applications can help to address input legitimacy deficits by contributing to a much wider range of tasks that include foresight and agenda setting (Ceron and Negri, 2016; Poel et al., 2018). For instance, some existing applications already use public discourse and opinion poll data to predict issues that require political action before these become problematic (Ceron and Negri, 2016; Rubinstein et al., 2016). Further empirical investigation is needed to assess how such applications might affect legitimacy perceptions. Our findings from a German OAP imply citizens' skepticism regarding the potential of ADM systems to increase democratic participation and citizens' representation (input legitimacy).

With regard to the quality of decision-making processes—throughput legitimacy—we found no difference between existing decision-making arrangements and hybrid regimes, involving ADM systems and EU politicians. However, citizens seem to view decision-making based solely on ADM systems as less fair or appropriate than the other two arrangements. Regarding existing EU procedures and practices, critics lament a lack of transparency, efficiency, and accountability (Schmidt and Wood, 2019), yet ADM systems exhibit the same deficiency (Shin and Park, 2019). Inside the “black box,” ADM systems change and adapt decision-making criteria according to new inputs and elusive feedback loops that defy explanation even among AI experts. Under the umbrella term “explainable AI,” a significant strand of computer science literature seeks to enhance ADM's transparency to users and the general public (Miller, 2019; Mittelstadt et al., 2019). For instance, “counterfactual explanations” indicate which ADM criteria would need to be changed to arrive at a different decision (Wachter et al., 2018).

Regarding citizens' perceptions of the effectiveness and favorability of decision-making outcomes (output legitimacy), we found no difference between the existing decision-making process and hybrid regimes incorporating ADM systems and EU politicians. ADM-based systems alone are considered unable to achieve the desired goals, and the respondents in our sample would not approve of the corresponding decisions. It is important to note that decision output was identical for all three experimental conditions, and that only the decision-making process varied. Nevertheless, these result in differing perceptions of output legitimacy, implying that factual legitimacy (as in the actual quality of policies and their outcomes) and perceived legitimacy are not necessarily congruent. In relation to the European debt crisis, Jones (2009) suggested that political institutions must convince the public that they are performing properly, whatever their actual performance. As the interplay between actual and perceived performance also seems important in the case of ADM, we contend that both aspects warrant equal consideration when implementing such systems in policy-making.

Some of the present results run counter to our hypotheses. Given the largely positive attitude to AI in the EU (European Commission, 2015, 2017), and in light of recent empirical evidence (Araujo et al., 2020;

Marcinkowski et al., 2020), we expected ADM to score highly on output legitimacy. However, respondents expressed a more favorable view of HDM and HyDM outcomes, suggesting that they consider it illegitimate to leave important EU political decisions solely to automated systems. ADM systems were considered legitimate as long as humans remained in the loop, indicating that to maintain existing levels of perceived legitimacy, ADM systems should support or inform human policy-makers rather than replacing them. This finding is consistent with the recommendations for “trustworthy AI” by the HLEG calling for human oversight “through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach” (Artificial Intelligence High-Level Expert Group, 2019, p. 16). However, if the decision-making process includes the capability for human intervention or veto, respondents in our study perceived ADM systems as fairly legitimate even in far-reaching decision-making like distributing the EU budget. More research is needed to systematically investigate whether this finding is stable across different political decisions to zoom in on potential differences between collective decisions that affect the whole society (e.g., EU budget) and individual-level decisions that affect single citizens (e.g., loan granting).

Furthermore, the results suggest that implementing hybrid systems for EU policy-making would mean that the EU doubles down on focusing on output legitimacy instead of input legitimacy as citizens do not perceive those systems to solve the democratic deficit of the EU (input legitimacy), yet they consider them to produce equally good policy outcomes. The findings also indicate that increasing factual legitimacy (e.g., by improving the quality of policy outcomes) does not necessarily yield a corresponding increase in perceived legitimacy.

7.2. Implications for data-driven policy-making

Our findings also contribute to the current discussion around data-driven or algorithmic policy-making. To begin, ADM systems as sole decision-makers do not seem to enhance citizens’ assessment of decision-making procedures or outcomes in our sample. However, when such systems operate under the scrutiny of democratically elected institutions (as in the hybrid condition), they are seen to be as legitimate as the existing policy-making process. This suggests that including humans in the loop is a necessary precondition for implementing ADM (Goldenfein, 2019), lending support to the EU’s call for trustworthy AI that highlights the crucial importance of human agency and oversight. With recent reports indicating that this may thus be the more plausible scenario in the immediate future (Poel et al., 2018; AlgorithmWatch, 2019), this finding has important implications for data-driven policy-making, as it shows that citizens view HITL, HOTL, or HIC decision-making as legitimate arguably because politicians can modify or overrule decisions made by ADM systems (Dietvorst et al., 2018).

Of course, our study tests a far-reaching form of algorithmic policy-making, in which algorithms take important budgeting decisions under conditions of limited (HyDM condition) or no (ADM condition) democratic oversight. Yet, as Verhulst, Engin, and Crowcroft point out: “Data have the potential to transform every part of the policy-making life cycle—agenda setting and needs identification; the search for solutions; prototyping and implementation of solutions; enforcement; and evaluation” (Verhulst et al., 2019, p. 1). Public administration has only recently begun to exploit the potential of ADM to produce better outcomes (Wirtz et al., 2019). For instance, the Netherlands now uses an ADM system to detect welfare fraud, and in Poland, the Ministry of Justice has implemented an ADM system that randomly allocates court cases to judges (AlgorithmWatch, 2019). Given the increased data availability and computing power fueling powerful AI innovations, it is reasonable to assume that we have only scratched the surface of algorithmic policy-making and that more far-reaching forms of ADM will be implemented in the future. Moreover, first opinion polls suggest that significant shares of citizens (25% in the EU) agree with AI taking over important political decisions about their country (Rubio and Lastra, 2019). The present findings suggest that implementation processes should be designed to facilitate synergies between ADM and HDM.

7.3. Implications for future empirical research

Four main limitations of this study outline avenues for future empirical research. First, our sample was not representative of the German population. As data were collected using the noncommercial SoSci OAP, the convenience sample was skewed in terms of education. This may have yielded slightly more positive perceptions of current EU legitimacy (HDM condition) as compared to the German population, as previous evidence suggests that higher levels of education are associated with more positive attitudes to the EU (Boomgaarden et al., 2011). Furthermore, the sample may be over-sophisticated in terms of digital literacy and general interest in ADM, which are likely to be associated with perceived legitimacy of as well as trust in algorithmic processes (Cheng et al., 2019). To make stronger claims in terms of the generalizability of the results, future research should use representative national samples.

Second, due to the successful randomization check and better power and fit indices of the model, we did not further control for confounding variables. Future research needs to investigate the effects of individual factors (e.g., EU attitudes, digital literacy) on the perceived legitimacy of HDM, ADM, and HyDM in the EU context.

Third, our study was limited to Germany. While German citizens generally hold more positive views of the EU compared to the European average (European Commission, 2019b), they also favor ADM in politics more than the European average (Rubio and Lastra, 2019). Future studies should investigate the relationship between AI-driven decision-making and perceptions of legitimacy in other national contexts and by means of cross-country comparisons. For instance, preliminary opinion polls suggest that Netherlands' citizens express much higher support for ADM in policy-making than citizens of Portugal (43 versus 19%) (Rubio and Lastra, 2019).

Finally, two of the three decision-making arrangements tested here are hypothetical and are unlikely to be implemented in the immediate future—that is, ADM systems are unlikely to be authorized to allocate the EU's annual budget. On that basis, future research should focus on the effects of less abstract data-driven applications on perceived legitimacy at different stages of the policy cycle and should include varying degrees of transparency of self-learning systems (De Fine Licht and De Fine Licht, 2020). In light of the EU's recent efforts to implement trustworthy AI systems in society, further empirical scrutiny needs to be devoted to assess whether proposed standards and guidelines are also in accordance with public demands and expectations. For instance, citizens may consider it more legitimate to employ AI-based systems to identify existing societal issues requiring political action or to evaluate the success of legislation based on extensive available data, on the condition that such systems will be able to give convincing justifications for their decisions.

8. Conclusion

This study sheds first light on citizens' perceptions of the legitimacy of using ADM in EU policy-making. Based on these empirical findings, we suggest that EU policy-makers should exercise caution when incorporating ADM systems in the decision-making process. To maintain the current levels of perceived legitimacy, ADM systems should only be used to assist or consult human decision-makers rather than replacing them, as excluding humans from the loop seems detrimental to perceived legitimacy. Second, it seems clear that the factual and perceived legitimacy of ADM do not necessarily correspond—that is, even ADM systems that produce high-quality outputs and are implemented transparently and fairly may still be perceived as illegitimate and will, therefore, be rejected. To be socially acceptable, implementation of ADM systems must, therefore, take account of both factual and perceived legitimacy. This study lays the groundwork for further research and hopefully sparks further investigations, addressing the impact of specific nuances of ADM in data-driven policy-making.

Acknowledgments. We would like to thank Nils Köbis for his valuable comments to the manuscript. A preprint of the manuscript is available at [arXiv:2003.11320](https://arxiv.org/abs/2003.11320).

Funding Statement. This research project was self-funded.

Competing Interests. The authors declare no competing interests exist.

Data Availability Statement. Replication data can be found in Zenodo: <https://doi.org/10.5281/zenodo.3728207>

Author Contributions Both authors have contributed to the Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Visualization, Writing—original draft, Writing—review and editing. All authors approved the final submitted draft.

Ethical Standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Supplementary Materials. To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/dap.2020.19>.

References

- AlgorithmWatch (2019) *Automating Society: Taking Stock of Automated Decision-Making in the EU*. Berlin: AW AlgorithmWatch GmbH. Available at https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf
- Araujo T, Helberger N, Kruikemeier S and de Vreese CH (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society* 35, 611–623. <https://doi.org/10.1007/s00146-019-00931-w>
- Arnesen S, Broderstad TS, Johannesson MP and Linde J (2019) Conditional legitimacy: how turnout, majority size, and outcome affect perceptions of legitimacy in European Union membership referendums. *European Union Politics* 20(2), 176–197. <https://doi.org/10.1177/1465116518820163>
- Arnesen S and Peters Y (2018) The legitimacy of representation: how descriptive, formal, and responsiveness representation affect the acceptability of political decisions. *Comparative Political Studies* 51(7), 868–899. <https://doi.org/10.1177/0010414017720702>
- Artificial Intelligence High-Level Expert Group (2019) *Ethics Guidelines for Trustworthy AI*. Available at <https://ec.europa.eu/futurium/en/ai-alliance-consultation>
- Awad E, Dsouza S, Bonnefon JF, Shariff A and Rahwan I (2020) Crowdsourcing moral machines. *Communications of the ACM* 63(3), 48–55. <https://doi.org/10.1145/3339904>
- Bansak K, Ferwerda J, Hainmueller J, Dillon A, Hangartner D, Lawrence D and Weinstein J (2018) Improving refugee integration through data-driven algorithmic assignment. *Science* 359(6373), 325–329. <https://doi.org/10.1126/science.aao4408>
- Barocas S and Selbst A (2016) Big data's disparate impact. *California Law Review* 104(1), 671–729. <https://doi.org/10.15779/Z38BG31>
- Bauer MW (Ed.) (1995) *Resistance to New Technology: Nuclear Power, Information Technology, and Biotechnology*. Cambridge: Cambridge University Press.
- Boomgaarden HG, Schuck ART, Elenbaas M and de Vreese CH (2011) Mapping EU attitudes: conceptual and empirical dimensions of euroscepticism and EU support. *European Union Politics*, 12(2), 241–266. <https://doi.org/10.1177/1465116510395411>
- Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Breitsohl H (2019) Beyond ANOVA. An introduction to structural equation models for experimental designs. *Organizational Research Methods* 22(3), 649–677. <https://doi.org/10.1177/1094428118754988>
- Brüggen E, Wetzels M, De Ruyter K and Schillewaert N (2011) Individual differences in motivation to participate in online panels. *International Journal of Market Research* 53(3), 369–390. <https://doi.org/10.2501/IJMR-53-3-369-390>
- Ceron A and Negri F (2016) The “Social Side” of Public Policy: Monitoring Online Public Opinion and Its Mobilization During the Policy Cycle. *Policy & Internet*, 8(2), 131–147. <https://doi.org/10.1002/poi3.117>
- Cheng X, Guo F, Chen J, Li K, Zhang Y and Gao P (2019) Exploring the trust influencing mechanism of robo-advisor service: a mixed method approach. *Sustainability* 11, 4917. <https://doi.org/10.3390/su11184917>
- Colquitt JA and Rodell JB (2015) Measuring justice and fairness. In Cropanzano RS and Ambrose ML(eds), *The Oxford Handbook of Justice in the Workplace*. Oxford: Oxford University Press.
- Craglia M, Annoni A, Benczur P, Bertoldi P, Delipetrev B, De Prato G, ... Vesnic Alujevic L (2018) *Artificial Intelligence: A European Perspective*. Bruxelles: Publications Office of the European Union. <https://doi.org/10.2760/11251>
- De Angelis G (2017) Political legitimacy and the European crisis: analysis of a faltering project. *European Politics and Society* 18(3), 291–300. <https://doi.org/10.1080/23745118.2016.1229383>
- De Fine Licht K and De Fine Licht J (2020) Artificial intelligence, transparency, and public decision-making. *AI & Society* 35, 917–926. <https://doi.org/10.1007/s00146-020-00960-w>
- Dietvorst BJ, Simmons JP and Massey C (2018) Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. *Management Science* 64(3), 1155–1170. <https://doi.org/10.1287/mnsc.2016.2643>
- Esaiasson P, Gilljam M and Persson M (2012) Which decision-making arrangements generate the strongest legitimacy beliefs? Evidence from a randomised field experiment. *European Journal of Political Research* 51(6), 785–808. <https://doi.org/10.1111/j.1475-6765.2011.02052.x>
- Esty D and Rushing R (2007) The promise of data-driven policymaking. *Issues in Science and Technology* 23(4), 67–72.

- European Commission (2015) *Special Eurobarometer 427: Autonomous Systems*. Available at http://data.europa.eu/euodp/en/data/dataset/S2018_82_4_427_ENG
- European Commission (2017) *Special Eurobarometer 460: Attitudes Towards the Impact of Digitisation and Automation on Daily Life*. Available at http://data.europa.eu/euodp/en/data/dataset/%0AS2160_87_1_460_ENG
- European Commission (2019a) *Data4Policy*. Available at <https://www.data4policy.eu/> (accessed 10 October 2019).
- European Commission (2019b) *Eurobarometer Interactive*. Available at <http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Chart/index> (accessed 14 June 2017).
- European Union (2019) *Wie die EU-Haushaltsmittel ausgegeben werde*. Available at https://europa.eu/european-union/about-eu/eu-budget/expenditure_de%0D (accessed 17 December 2019).
- Follesdal A (2006) The legitimacy deficits of the European Union. *Journal of Political Philosophy* 14(4), 441–468.
- Follesdal A and Hix S (2006) Why there is a democratic deficit in the EU: a response to Majone and Moravcsik. *JCMS: Journal of Common Market Studies* 44(3), 533–562. <https://doi.org/10.1111/j.1468-5965.2006.00650.x>
- Giest S (2017) Big data for policymaking: fad or fasttrack? *Policy Sciences* 50(3), 367–382. <https://doi.org/10.1007/s11077-017-9293-1>
- Goldenfein J (2019) Algorithmic transparency and decision-making accountability. Thoughts for buying machine learning algorithms. In Bertram C, Gibson A and Nugent A (eds), *Closer to the Machine: Technical, Social, and Legal Aspects of AI*. Melbourne: Office of the Victorian Information Commissioner, pp. 41–60.
- Gurr TR (1971) *Why Men Rebel*. Stanford: Princeton University Press.
- Habermas J (2009) *Europe: The Faltering Project*. Oxford: Polity.
- Hobolt SB and Wittrock J (2011) The second-order election model revisited: an experimental test of vote choices in European Parliament elections. *Electoral Studies* 30(1), 29–40. <https://doi.org/10.1016/J.ELECTSTUD.2010.09.020>
- Holbert RL and Stephenson MT (2002) Structural equation modeling in the communication sciences, 1995–2000. *Human Communication Research* 28(4), 531–551. Available at <http://doi.org/10.1111/j.1468-2958.2002.tb00822.x>
- Holzhaecker R (2007) Democratic legitimacy and the European Union. *Journal of European Integration* 29(3), 257–269. <https://doi.org/10.1080/07036330701442232>
- Jones E (2009) Output legitimacy and the global financial crisis: perceptions matter. *JCMS: Journal of Common Market Studies* 47(5), 1085–1105. <https://doi.org/10.1111/j.1468-5965.2009.02036.x>
- Katz Y (2017) Manufacturing an artificial intelligence revolution. *SSRN Electronic Journal*, 1–21. <https://doi.org/10.2139/ssrn.3078224>
- Kleinen-von Königlöw K (2012) Europe in crisis? Testing the stability and explanatory factors of the Europeanization of national public spheres: the European public sphere revisited. *The International Communication Gazette* 74(5), 443–463. <https://doi.org/10.1177/1748048512445153>
- Kline RB (2016) *Principles and Practice of Structural Equation Modeling*, 4th Edn. New York: Guilford Press.
- Lee MK (2018) Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5(1). <https://doi.org/10.1177/2053951718756684>
- Leiner DJ (2016) Our research's breadth lives on convenience samples. A case study of the online respondent pool “SoSci Panel”. *Studies in Communication Media* 5(4), 367–396. <https://doi.org/10.5771/2192-4007-2016-4-36769-134>
- Leiner DJ (2019) Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods*, 13(3), 229–248. <https://doi.org/10.18148/srm/2019.v13i3.7403>
- Lindgren KO and Persson T (2010) Input and output legitimacy: synergy or trade-off? Empirical evidence from an EU survey. *Journal of European Public Policy* 17(4), 449–467. <https://doi.org/10.1080/13501761003673591>
- Marcinkowski F, Kieslich K, Starke C and Lünich M (2020) *Implications of AI (un-)fairness in Higher Education Admissions: The Effects of Perceived AI (un-)fairness on Exit, Voice and Organizational Reputation*. In Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery.
- Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution that will Transform How We Live, Work and Think*. London: Murray.
- Menon A and Weatherill S (2008) Transnational legitimacy in a globalising world: how the European Union rescues its states. *West European Politics* 31(3), 397–416. <https://doi.org/10.1080/01402380801939610>
- Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. *Artificial Intelligence* 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Mittelstadt B, Russell C and Wachter S (2019) *Explaining Explanations in AI*. FAT* 2019 – Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency, pp. 279–288. <https://doi.org/10.1145/3287560.3287574>
- Persson M, Esaiasson P and Gilljam M (2013) The effects of direct voting and deliberation on legitimacy beliefs: an experimental study of small group decision-making. *European Political Science Review* 5(3), 381–399. <https://doi.org/10.1017/S1755773912000173>
- Poel M, Meyer ET and Schroeder R (2018) Big data for policymaking: great expectations, but with limited progress? *Policy & Internet* 10(3), 347–367. <https://doi.org/10.1002/poi3.176>
- Richardson R, Schultz JM and Southerland VM (2019) *Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems*. New York: AI NOW Institute. Available at <https://ainowinstitute.org/litigatingalgorithms-2019-us.html>

- Rieder G and Simon J** (2016) Datatrust: or, the political quest for numerical evidence and the epistemologies of big data. *Big Data & Society* 3(1), 1–6. <https://doi.org/10.1177/2053951716649398>
- Risse T** (2014) No demos? Identities and public spheres in the Euro crisis. *Journal of Common Market Studies* 52(6), 1207–1215. <https://doi.org/10.1111/jcms.12189>
- Rubinstein M, Meyer E, Schroeder R, Poel M, Treperman J, van Barneveld J, Biesma-Pickles A, Mahieu B, Potau X, Svetachova M** (2016) *Ten Use Cases of Innovative Data-Driven Approaches for Policymaking at EU Level*. Available at https://pdfs.semanticscholar.org/ac03/2f2a3e81c5d245cdebfa92c86c6660a29017.pdf?_ga=2.132982414.1684883596.1585744060-853246622.1585590200
- Rubio D and Lastra C** (2019) *European Tech Insights 2019: Mapping European Attitudes to Technological Change and Its Governance. Center for the Governance of Change*. Available at <https://www.ie.edu/cgc/research/tech-opinion-poll-2019/>
- Scharpf FW** (1999) *Governing in Europe*. Oxford: Oxford University Press.
- Scharpf FW** (2009) Legitimacy in the multilevel European polity. *European Political Science Review* 1(2), 173–204. <https://doi.org/10.1017/S1755773909000204>
- Schmidt VA** (2013) Democracy and legitimacy in the European Union revisited: input, output and “throughput.” *Political Studies* 61, 2–22. <https://doi.org/10.1111/j.1467-9248.2012.00962.x>
- Schmidt VA** (2015) *The Eurozone’s Crisis of Democratic Legitimacy: Can the EU Rebuild Public Trust and Support for European Economic Integration?* (FELLOWSHIP INITIATIVE 2014-2015 “Growth, Integration and Structural Convergence Revisited” No. 15). Luxembourg.
- Schmidt VA and Wood M** (2019) Conceptualizing throughput legitimacy: procedural mechanisms of accountability, transparency, inclusiveness and openness in EU governance. *Public Administration*, padm.12615. 97, 727–740. <https://doi.org/10.1111/padm.12615>
- Shin D and Park YJ** (2019) Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior* 98 (April), 277–284. <https://doi.org/10.1016/j.chb.2019.04.019>
- Sluban B and Battiston S** (2017) Policy co-creation in the era of data science. In Presented at the Data for Policy 2017: Government by Algorithm? (Data for Policy). London: Zenodo. <http://doi.org/10.5281/zenodo.892390>, <https://zenodo.org/record/892390#.X7eOiOdCdhE>
- Soper DS** (2019) A-priori sample size calculator for structural equation models. Available at <http://www.danielsoper.com/statcalc>
- Starke C, Marcinkowski F and Wintterlin F** (2020) Social networking sites, personalization and trust in government: empirical evidence for a mediation model. *Social Media + Society*, 1–11.
- Steffek J** (2019) The limits of proceduralism: critical remarks on the rise of “throughput legitimacy.” *Public Administration*, padm.12565. 97(4), 784–796. <https://doi.org/10.1111/padm.12565>
- Steinmetz H, Schmidt P, Tina-Booh A, Wieczorek S and Schwartz SH** (2009) Testing measurement invariance using multigroup CFA: differences between educational groups in human values measurement. *Quality & Quantity* 43(4), 599–616. <https://doi.org/10.1007/s11135-007-9143-x>
- van de Schoot R, Lugtig P and Hox J** (2012) A checklist for testing measurement invariance. *European Journal of Developmental Psychology* 9(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- van Veenstra AF and Kotterink B** (2017) Data-driven policy making: the policy lab approach. In Parycek P, Charalabidis Y, Chugunov AV, Panagiotopoulos P, Pardo TA, Sæbø Ø and Tambouris E (Eds.), *Electronic Participation. ePart 2017. Lecture Notes in Computer Science* (Vol. 10429). Cham: Springer, pp. 100–111. https://doi.org/10.1007/978-3-319-64322-9_9
- Verhulst SG, Engin Z and Crowcroft J** (2019) Data & Policy: a new venue to study and explore policy – data interaction. *Data & Policy* 1, 1–5. <https://doi.org/10.1017/dap.2019.2>
- Wachter S, Mittelstadt B and Russell C** (2018) Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard Journal of Law & Technology*. Available at <http://arxiv.org/abs/1711.00399>
- Weiler JHH** (2012) In the face of crisis: input legitimacy, output legitimacy and the political messianism of European integration. *Journal of European Integration* 34(7), 825–841. <https://doi.org/10.1080/07036337.2012.726017>
- Werner H and Marien S** (2018) The macro-level impact of small-scale involvement processes. Experimental evidence on the effects of involvement on legitimacy perceptions of the wider public. In *ECPR General Conference*. Hamburg.
- Widaman KF and Reise SP** (1997) Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In Bryant KJ, Windle MT and West SG (eds), *The Science of Prevention. Methodological Advances from Alcohol and Substance Abuse Research*. Washington: American Psychological Association, pp. 281–324.
- Wirtz BW, Weyerer JC and Geyer C** (2019) Artificial intelligence and the public sector-applications and challenges. *International Journal of Public Administration*. 42, 596–615. <https://doi.org/10.1080/01900692.2018.1498103>

Cite this article: Starke C and 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. doi:10.1017/dap.2020.19