**Doctoral thesis submitted to**
**the Faculty of Behavioural and Cultural Studies**
**Heidelberg University**
**in partial fulfillment of the requirements of the degree of**
**Doctor of Philosophy (Dr. phil.)**
**in Psychology**

Title of the publication-based thesis

# *Paving New Ways – The Case for Machine Learning in Political Psychology*

presented by
Nils Brandenstein

year of submission
2025

Dean: Prof. Dr. Guido Sprenger
Supervisor: Prof. Dr. Jan Rummel

# Contents

# 1 List of scientific publications of the publication-based dissertation

**Manuscript I**

Brandenstein, N. (2022). Going beyond simplicity: Using machine learning to predict belief in conspiracy theories. *European Journal of Social Psychology*, *52*(5-6), 910–930. https://doi.org/10.1002/ejsp.2859

**Manuscript II**

Brandenstein, N., Ackermann, K., Aeschbach, N., & Rummel, J. (2023). The key determinants of individual greenhouse gas emissions in Germany are mostly domain-specific. *Communications Earth & Environment*, *4*(1), 422. https://doi.org/10.1038/s43247-023-01092-x

**Manuscript III**

Brandenstein, N., Montag, C., & Sindermann, C. (2024). To follow or not to follow: Estimating political opinion from Twitter data using a network-based machine learning approach. *Social Science Computer Review*, *0*(0). https://doi.org/10.1177/08944393241279418

# II Acknowledgements

What a ride! The last five years have proven to be the most challenging, exciting, definitely stressful, but mostly rewarding and joyful time of my life so far. Like most PhD students, I had no idea of what I was getting myself into. But one thing I quickly learned is that the people around you make all the difference and can help you master even the biggest challenges. I am deeply grateful for the continuous support of my colleagues, friends and family.

First, I want to thank my supervisor, Prof. Dr. Jan Rummel. After presenting my master's thesis at his colloquium and without knowing me well, he offered me the opportunity to pursue a PhD in his lab, and most importantly, to pursue *my* PhD. Whenever I tell other researchers that I do political psychology research using machine learning models in an experimental psychology lab, they understandably look confused at first. While our research interests are very different (although they have converged over time), Jan has always supported my ideas, given me the freedom to explore them and - most importantly - believed in my success. He always made time in his busy schedule for discussions about research, teaching, administration, or career planning, and has shaped both me as a researcher and my work in various ways. I know that kind of support is not the norm in psychology labs and I am truly grateful to have him as my supervisor. Next, I want to thank Prof. Dr. Andreas Voß for immediately agreeing to be the second reviewer of my dissertation and for showing genuine interest in my studies.

Great colleagues make for a great work environment and I undoubtedly had the best there are. First, I want to thank the *Sacred Cat Gang*, Wiebke, Steffi, Kathrin, Milena, Lukas, Fabi and Thorben. After starting off as colleagues (I count Fabi and Thorben as MOVs at this point), we now have become very close friends. Thank you for making my PhD (and my life in general) much more fun and exciting. I am so glad I met you! I am also grateful to all my other (former) colleagues for the many insightful discussions and fun times we had. Above all, a big thank you to Lena, my very first colleague and now a good friend, who welcomed me with open arms and helped me a lot in my early days at the PI. Special thanks also go to Stefan who first sparked my interest in machine learning and always made room to talk about the methodological side of my studies. Beyond the PI, I want to thank my collaborators Kathrin, Nicole and especially Conny who not only contributed a lot to our joint project but also made the conferences we attended a lot more fun! I also want to thank all our research assistants who helped me in realizing my projects and contributed greatly to their success.

Naturally, people outside the PI also played a big role in my PhD journey. Above all, I want to thank my girlfriend Alina. I could not have done this without your unconditional support, love and belief in me. Whenever I felt uncertain about a project, frustrated by the results or overwhelmed by stress you had my back and made me believe that things would turn out fine. I cannot think of a better partner and friend by my side. I also want to thank my best friend Orlando who always had a sympathetic ear and never complained when I did not reply for days in a row. Finally, I want to thank my parents and family for always believing in me and supporting my decision to pursue a PhD.

# III Preface

I would not consider myself a very 'political' person. Although I do hold political values such as liberty, equality, or social justice, I have never fully endorsed the politics of any German party or their solutions to domestic or global problems. I believe that many of the proposed policies and political strategies to combat problems in our globalized, complex world will not have the desired impact, or sometimes even unknown adverse consequences down the road. As a result, I often found myself not having extreme opinions, and thus never really involved myself in deep political discussions. That being said, I was always fascinated by politics. Especially today, where our world seems more volatile and confusing than ever with pressing issues like climate change, the rise of authoritarian leaders, widespread misinformation campaigns on social media or deeply divided societies; there is a lot going on that I find important to think about. However, being a psychologist by training, a distinct interest of mine has always been the individual citizen. For instance, I was keen to understand why people cling to to certain political beliefs, even though they sound totally irrational to me or why we know about our impact on the climate but do not behave accordingly. In my pursuit of answers, I came across the field of *Political Psychology*, which seeks to understand political events and humans as "political beings" from a psychological perspective by exploring a wide range of research questions using diverse data sources, designs and methods (Cottam et al., 2004). In the past decades, activities in this area have grown substantially and revealed many insights into how *political cognitions* (e.g., beliefs, attitudes, ideologies) and *political behavior* (e.g., voting, policy support) are shaped. Around the same time when I was reading studies on topics I found interesting, I got in touch with Machine Learning (*ML*) models. Shortly after, I took a few courses to understand how these models work and how they could be used in research settings. This is when I started to see potential limitations in (political) psychology research and how they have likely contributed to different challenges in the past years, like the reproducibility crisis, slow theoretical advancements or ineffective political interventions. Fueled by my changed perspective on the field and the research opportunities it unlocked, this five-year journey began.

In this dissertation, I outline a few shifts in how research in political psychology can be conducted. The first shift argues for a stronger focus on predicting political cognitions and behavior. The second shift argues for considering the complexity of human beings and the usage of complex ML models. Finally, the third shift argues for a more effective utilization of big(ger) datasets and novel data sources. Across the different chapters, I discuss limitations and challenges of current research practices in political psychology and present how the outlined shifts can help addressing them, drawing primarily on recent theoretical and empirical work. Using the three manuscripts which constitute the main contribution of this publication-based dissertation, I specifically show how these shifts can be applied in political psychological research and how they can help to accumulate domain knowledge more rapidly. In these manuscripts, I investigated three distinct topics: 1. Predicting belief in conspiracy theories, 2. Identifying person and situation-related predictors of individual sustainable behavior and 3. Predicting individual political attitudes and behavior using social media data. Besides being suitable to illustrate the potential of the outlined shifts, these topics have been chosen because they represent some of the most pressing issues today, answering the increasing number of calls in recent years encouraging psychological research to address wide-reaching societal challenges (e.g., Gruber et al., 2019; Nielsen et al., 2024). Finally, I reflect on a potential future of political psychology with a focus on how the outlined shifts can be integrated with current research practices to help the field advance.

# 1 The importance of predicting political cognitions and behavior

As an empirical research field, political psychology (and psychology more broadly) seeks to understand human cognitions and behavior through data collection and statistical modeling to develop and test theories. In this context, two key approaches exist to guide the process of knowledge generation: The so-called *explanation-oriented* and the *prediction-oriented* approach (e.g., Shmueli, 2010; Yarkoni & Westfall, 2017). The explanation-oriented approach, which is the most commonly used in psychology, focuses on uncovering the underlying (causal) psychological mechanisms that drive cognitive and behavioral outcomes. In this approach, researchers seek to identify and isolate the effect of a predictor (i.e., independent variable) on an outcome (i.e., dependent variable) of interest. This is typically done by deriving hypotheses from theory, specifying corresponding statistical models and testing them using observational or experimental data (Agrawal et al., 2020; Hofman et al., 2021; Shmueli, 2010). In contrast, the so-called *prediction-oriented* approach primarily focuses on accurately predicting cognitive and behavioral outcomes. In this approach, researchers develop, compare and select statistical models based on their capability to accurately predict an outcome on unseen (*out-of-sample*) data, regardless of how the underlying psychological process (i.e., predictor-outcome relationship) might look like (e.g., Hofman et al., 2021; Shmueli, 2010). A common assumption among (political) psychologists is that this exclusive focus on prediction holds little to no scientific value to inform theory - and thus for understanding the psychological processes underlying cognitions and behavior (Shmueli, 2010). This might be one of the reasons why researchers have almost exclusively followed the explanation-oriented approach in the past, although it also comes with a few limitations which have arguably contributed to a few key challenges in current literature and practice[1] (e.g., De Slegte et al., 2024; Hofman et al., 2021; Shmueli, 2010; Yarkoni & Westfall, 2017). Based on similar arguments made in the past, I argue that a prediction-oriented approach does offer substantial scientific value and can help political psychology to advance as a field.

[1] See also chapter 4 and Manuscript I (Brandenstein, 2022) and Manuscript II (Brandenstein et al., 2023)

The explanation-oriented approach of isolating predictor-outcome relationships has undoubtedly helped political psychology uncover key psychological mechanisms underlying political cognitions and behavior over the years. However, as highlighted in numerous empirical and theoretical works, it also frequently produces statistical models with limited predictive performance (e.g., De Slegte et al., 2024; Hofman et al., 2021; Shmueli, 2010; Yarkoni & Westfall, 2017). This is due to several reasons. First, because the main objective is to uncover underlying (causal) effects, the model's ability to accurately predict the outcome is not an explicit concern in the research process (Hofman et al., 2021). Next, models developed under the explanation-oriented approach tend to focus on simplicity and often do not account for a variety of interacting, small-effect predictors which negatively impacts their predictive capabilities (more on this in the next chapter and in Götz et al. (2022), Hofman et al. (2021), and Shmueli (2010)). The last and more technical reason is illustrated by the *bias-variance-tradeoff* (Yarkoni & Westfall, 2017): Bias refers to systematic errors in the assumptions of a statistical model and it's ability to capture important patterns (i.e., signal) between variables. In contrast, variance describes a model's sensitivity to random noise in the dataset. Importantly, by reducing a model's bias (trying to capture more signal) we run the chance of capturing more random noise, thus increasing it's variance. Since the aim of the explanation-oriented approach is to understand the underlying (causal) effects of a predictor on an outcome, researchers tend to fit their model to a *single* dataset as closely as possible (e.g., by using the ordinary least squares (*OLS*) criterion in regression tasks) through aligning the model parameters with those of the theoretically-assumed data generating process (Yarkoni & Westfall, 2017). As a result, models developed under the explanation-oriented approach try to reduce bias exclusively (Yarkoni & Westfall, 2017). This often produces high variance models that are overly tailored to a single dataset (also known as *overfitting*) and thus perform poorly when predicting outcomes on out-of-sample data.[2].

But why should political psychologists care about prediction to understand psychological processes underlying cognitions and behavior after all? First, focusing on prediction can also help achieving reproducible results. This is because a model's ability to accurately predict outcomes in new, unseen data is a direct indicator of how well its results replicate (Hindman, 2015). Given the recent "reproducibility crisis" which showed that many findings in different subfields of psychology (including political psychology) consistently failed replication (e.g., Camerer et al., 2018; Gerber et al., 2018; Open Science Collaboration, 2015; Walker et al., 2024), evaluating and prioritizing predictive performance may be a key step to enhance the reliability of research results and the drawn conclusions about how cognitions and behavior work. Next, a focus on prediction - and using more complex prediction-focused models - can help to support the generation of new

[2] Besides other factors, the probability of obtaining high variance models is most pronounced in cases where a dataset is small while the number of predictors and/or model complexity is high or when the dataset contains a significant amount of noise (e.g., Hindman, 2015; Yarkoni & Westfall, 2017)

theoretical insights and hypotheses as well as refining and advancing existing explanation-oriented models and theories. Besides other benefits discussed throughout the next chapters, this is primarily because prediction-focused models can automatically learn complex, previously unknown relationships between predictors and outcomes that explanation-oriented models often miss (see also Grimmer et al. (2021), Hindman (2015), and Shmueli (2010)). Next, existing psychological theories may be evaluated by explicitly considering the prediction performance of the derived statistical models (Shmueli, 2010). For instance, if two competing theories exist for why people behave sustainably in their everyday life and model A (derived from theory A) clearly outperforms model B (derived from theory B), theory A arguably provides a more accurate representation of the psychological mechanisms driving sustainable behavior. In this regard, prediction can also serve as a "reality check" for the utility of theoretical explanations: Arguably, even elaborate theories of psychological processes can hold little practical value if the derived statistical models cannot predict the cognitive or behavioral outcome they aim to explain (Shmueli, 2010; Yarkoni & Westfall, 2017). For example, if policymakers design an intervention to promote sustainable behavior based on a psychological theory that explains why some individuals act environmentally friendly but fails to accurately predict the behavior of most other people, the intervention is unlikely to produce substantial large-scale changes.

Based on these considerations, focusing on prediction and obtaining models that predict well should arguably be a key objective in political psychology. However, following the explanation-oriented modeling approach has been - and will be - largely unhelpful in this regard due to the mentioned limitations. To address this, political psychologist may thus shift to a prediction-oriented modeling approach which exhibits two key features: Assessing prediction performance during the modeling phase and using more flexible (complex) models to predict outcomes (Yarkoni & Westfall, 2017). To assess prediction error, researchers use model evaluation techniques such as cross-validation (*CV*) or *train-test-splits* (setting a part of the full dataset aside to evaluate a fitted model on). These techniques allow researchers to reliably estimate a model's reproducibility potential without extra data collection, provide guardrails against overfitting, help to select among competing models and also reduce the potential for problematic analytical procedures often found in the explanation-oriented approach (such as *p-hacking*) (Yarkoni & Westfall, 2017). However, these techniques do not improve prediction performance themselves. Therefore, researchers often deploy complex and flexible ML models in the prediction-oriented approach. This is because - in comparison to explanation-oriented models - ML models are designed and optimized for prediction. One key advantage of ML models is that their bias - and thus the bias-variance ratio - can be freely adjusted, which is also known as *regularization*. Notably, introducing bias into a model often reduces prediction error the most (Yarkoni & Westfall, 2017). This is one of the reasons why ML models have been shown to deliver more accurate predictions and reproducible results across different research areas in comparison to unbiased models such as OLS regression (Hindman, 2015; Yarkoni & Westfall, 2017). As mentioned before, another important hallmark of many ML models is their ability to capture complex relationships between predictors and outcomes which often increases prediction performance even more. In the next chapter, I further elaborate on this feature and demonstrate the unique potential of ML models to gain a deeper understanding of how political cognitions and behavior might work.

# 2 Daring more complexity - Harnessing machine learning models in political psychology

Humans are complex creatures. Researchers across different subfields of psychology have acknowledged the multi-causality of our psychological world and that cognitions and behavior are shaped by an intricate interplay of different psychological, social, cultural, and environmental factors (e.g., Götz et al., 2022; Hofman et al., 2021). Political psychology also reflects this view, with research highlighting that cognitions such as political beliefs and ideology or behaviors such as climate action are likely influenced by multiple, interacting factors (e.g., Feldman & Johnston, 2014; Kollmuss & Agyeman, 2002; Van Valkengoed & Steg, 2019). However, the statistical models and analysis strategies used by political psychologists to study such complex outcomes are still rather simplistic. Notably, this mismatch may lead to undesirable consequences for theory and practice. In accordance with previous considerations, and to address this, I argue that researchers should explicitly account for the multitude of influencing factors and their complex interplay when studying political cognitions and behavior, which may best be achieved by using ML models.

As noted in the previous chapter, following the explanation-oriented approach in political psychology often causes researchers to focus on just one or a few predictors of political cognitions and behavior at a time. Crucially, this can not only lead to models with low predictive accuracy but also potentially misguided conclusions about the relative importance of the studied predictors (e.g., Götz et al., 2022; James et al., 2013, pp. 71-74). To illustrate this, consider the example of a researcher interested in predictors of pro-environmental behavior: They might theoretically assume that individuals who hold pro-environmental attitudes and values travel by car less often. To test this assumption, they collect a dataset measuring these variables, analyze their relationships using a statistical model (e.g., OLS regression) and may initially conclude that pro-environmental attitudes and values are associated with reduced car usage. However, they might later find that these effects diminish considerably when also accounting for infrastructural barriers to public transportation ($PT$) such as increased travel time or perceived stress. Notably, this pattern would have remained undetected if the additional predictors were not included in the model. This isolation-focused analysis strategy in many studies may also lead to undesirable consequences in the real world: If policymakers base their regulation strategies on such results and target predictors that only marginally contribute to a behavior (like environmental attitudes and values in this example), it may come to no surprise that the implemented measures fail to reach the desired impact in the real world[1] (e.g., Abou-Zeid et al., 2012; Götz et al., 2022; Van Valkengoed et al., 2022). Next, even if studies do account for a multitude of potential predictors when studying pressing political issues such as environmental behavior or belief in conspiracy theories, they mostly use explanation-focused linear models such as OLS (e.g., Hunecke et al., 2007; Swami et al., 2011). Besides an increased risk to overfit such models and obtain unreliable results in this scenario (see chapter 1 and Hindman (2015) and Yarkoni and Westfall (2017)), such models can easily miss important patterns underlying political cognitions and behavior. Coming back to the example of car usage: Individuals may be willing to switch to public transport if the time difference is minimal, but become increasingly unlikely to do so as additional travel time grows (a non-linear effect). Similarly, the effect of pro-environmental attitudes on car use may be observable when the time cost of public transport is low, but may disappear when costs are high (a moderating effect). In this example, a researcher using a simple, explanation-focused OLS model would not only miss the non-linear effect but would also need to anticipate the interaction effect in advance in order to capture it. Although it is possible to formulate non-linear models or specify interactions between predictors in explanation-focused models (see for example Imhoff et al. (2022)), anticipating and testing all potential predictor-outcome relationships quickly becomes unattainable with an increasing number of predictors in a model (Agrawal et al., 2020). Thus, researchers may be more likely to miss such effects which can impair the discovery of important psychological processes underlying political cognitions and behavior[1] (e.g., Grimmer et al., 2021).

Based on these considerations and similar previous arguments (e.g., Götz et al., 2022; Grimmer et al., 2021; Hindman, 2015), political psychologists should consider accounting for a wider range of potential predictors and their interactions when studying political cognitions and behavior. An ideal candidate to achieve this are prediction-focused ML models. This is because many ML models, such as Random Forests, not only offer bias-adjustable properties that increase predictive performance and reliability of results (see chapter 1), but also have the ability to automatically capture complex, non-linear relationships between predictors and outcomes (e.g., Breiman, 2001; De Slegte et al., 2024; Grimmer et al., 2021). These features help to identify the most important, robust predictors of cognitive and behavioral outcomes while also uncovering complex underlying relationships with ease (e.g., De Slegte et al., 2024; Grimmer et al., 2021; Hindman, 2015). Similar to the prediction-oriented approach in general, a common assumption is that while ML models are powerful predictive models, they are practically useless to inform theory (e.g., De Slegte et al., 2024). ML models are often considered "black box models" whose internal mechanisms cannot

[1]See also chapter 4 and Manuscript I (Brandenstein, 2022) and Manuscript II (Brandenstein et al., 2023)

be interpreted, rendering them unsuitable to learn about psychological processes underlying cognitions and behavior. Although this assumption has some merit to it, a lot of progress has been made in the past years to enable and improve the interpretability of ML models (for an overview, see Molnar (2022)). An increasingly popular tool are so-called *model-agnostic* interpretation techniques (e.g., SHAP, permutation feature importance, ALE/dependence plots) which can be used to interpret any ML and traditional statistical model. Such techniques allow researchers to analyze different model properties and gain deep insights into the captured predictor-outcome relationships. These include, for instance, the overall importance and interactions of predictors or their relationship type, strength and direction with the outcome (Molnar, 2022). This allows ML models to be used not only for predictive tasks but also in a more explanation-oriented manner, such as evaluating and refining theories (e.g., Agrawal et al., 2020; De Slegte et al., 2024; Fudenberg & Liang, 2020). Another misconception is that ML models only work with large datasets. Although ML models provide specific advantages when dealing with large datasets (more on this in the next chapter), the benefits of many ML models can - ironically - often be the most pronounced in smaller to medium-sized datasets (Hindman, 2015). For instance, regularized linear ML models (e.g., Ridge, Lasso) often perform much better than their traditional counterpart (i.e., OLS) while making effective use of smaller datasets and remaining inherently interpretable (see Manuscript I [2] and II [3] and Hindman (2015), Molnar (2022), and Yarkoni and Westfall (2017)). Even though data sparsity has long been common in the social sciences, the availability of larger datasets from various sources is becoming increasingly prevalent, especially in political psychology. Therefore, the next chapter elaborates on how such large datasets are currently used in the field and discusses several ways to improve their utilization and potential information gain.

[2] Manuscript I: Brandenstein (2022)
[3] Manuscript II: Brandenstein et al. (2023)

# 3 The advent of data abundance - Utilizing large datasets and novel data sources to their full potential

For a long time, researchers in psychology and other social sciences have been afflicted by data sparsity. Conducting experimental and observational studies often comes with high financial and labor costs which limits researchers' ability to collect larger samples (e.g., Dong & Lian, 2021). Also, open access to and availability of existing datasets was restricted for a long time. Although little has changed for conducting tailored, controlled lab experiments, the internet and progress in open science practices (i.e., data sharing) now provides political psychologists with more data than ever: Today, datasets are not only more accessible but have also drastically increased in size (De Slegte et al., 2024). For instance, longitudinal panel studies such as the as the American National Election Studies (n.d.), the European Social Survey (n.d.) or surveys on specific topics collected though international, collaborative efforts (e.g., Imhoff et al., 2022; McBride et al., 2021) now incorporate thousands of samples and a multitude of variables relevant to political psychological research questions. Besides survey datasets, novel data sources like digital trace data (*DTD*, e.g., website activities and browsing behavior) or social media data (*SMD*, e.g., network structures or individual posts) have emerged in recent years, potentially offering new ways to address research questions (e.g., Rafaeli et al., 2019). However, despite some progress made in recent years (e.g., De Slegte et al., 2024; Grimmer et al., 2021), I argue that many political psychologists still fail to fully tap into the potential of today's data abundance. That is, most studies neither account for nor leverage the complexity of large (survey) datasets and under-utilize novel data sources.

Nowadays, an increasing number of studies in political psychology use large survey datasets (like the ones mentioned before) to address their research questions - and for good reason. First, using large sample sizes provides a natural antidote against overfitting, leading to a more realistic picture of singular predictor effect sizes for an outcome (e.g., De Slegte et al., 2024; Grimmer et al., 2021; Yarkoni & Westfall, 2017). On a similar note, many large survey datasets are collected in a way to ensure representativeness of the sample, enabling researchers to derive conclusions on a population level. Arguably, one of the greatest advantages of using large survey datasets is their potential to accelerate scientific discovery (Grimmer et al., 2021). The multitude of variables across diverse topics and large sample sizes enables researchers to perform complex, in-depth analyses of potential predictor-outcome relationships (e.g., Agrawal et al., 2020; Grimmer et al., 2021). In turn, researchers can be much faster in developing and testing psychological theories than it would be possible with individually collected small-scale datasets (e.g., Agrawal et al., 2020; De Slegte et al., 2024; Yarkoni & Westfall, 2017). Arguably, however, many studies in political psychology still miss out on this unique potential. As mentioned in the previous chapters, most researchers continue to focus on just one or a few predictors of political cognitions and behavior at a time, even when working with large survey datasets. While doing so can still be useful to test specific hypotheses, such restrained use of rich data sources significantly limits the ability to explore and uncover intricate, previously unknown variable relationships which could potentially help to gain further insights into political cognitions and behavior (e.g., Agrawal et al., 2020; Grimmer et al., 2021). Also, researchers might still miss the bigger picture of how important different predictors are for cognitive and behavioral outcomes when only analyzing a small subset, as noted in the previous chapter. Importantly, utilizing large datasets to their full potential may also call for the application of complex ML for other reasons. This is best illustrated when considering the potential benefits and analytical challenges associated with datasets from novel sources like DTD and SMD. Arguably, SMD in particular can offer unique possibilities to address research questions and provide a few key advantages over laboratory and survey data, which is why they have gained more scholarly attention in recent years (e.g., Dong & Lian, 2021; Rousidis et al., 2020). SMD offer a plethora of information on individual citizens (*users*) such as posts, likes, followed accounts and other meta-data. In comparison to other data sources, these measurements capture spontaneous, unsolicited expressions and behavior which enables researchers to conduct ecologically-valid, real-time analysis of political cognitions and behavior (e.g., Reveilhac et al., 2022; Schober et al., 2016). Additionally, the pool of potential study subjects is very large with more than 5 billion users across platforms worldwide (Kemp, 2024). These and many other features of SMD have enabled researchers in social science to address research questions of high societal relevance in previously infeasible ways, such as predicting public opinion and electoral outcomes (e.g., Skoric et al., 2020), monitoring and understanding the spread of misinformation (e.g., González-Bailón et al., 2024) or studying mechanisms of political polarization in social networks (e.g., Bail et al., 2018; Nyhan et al., 2023). Notably, the unique advantages of SMD also pose some methodological challenges at the same time. Large corpora of text-based posts, network information or other meta-data from millions of users come in an unstructured format and can be very disorganized, which requires a lot of pre-processing, computational resources and multiple analysis steps to extract the desired information from (e.g., Abbas, 2021; Reveilhac et al., 2022). Although some of these features also apply to survey data (e.g., large sample sizes or unstructured data types like open field responses), the unique combination of challenges associated with handling SMD may be one of the reasons for why SMD have not been widely adopted in political psychology (see

Dong & Lian, 2021, for a recent overview of previous use cases). However, using ML models can once again help to make effective use of such complex data types and facilitate the analysis process of SMD in multiple ways. First, contemporary ML models, such as large language models (e.g., GPT) or other types of deep neural networks, can automatically extract and analyze meaningful and complex features from data without relying on manual coding or traditional rule-based modeling (e.g., Hankar et al., 2025). This enables them to, for instance, identify sentiments and topics in texts or intricate relationships between individuals in a network with a level of contextual understanding that exceeds what most traditional approaches and models can achieve[1] (e.g., Hankar et al., 2025). Moreover, many contemporary ML models are designed to efficiently scale to large data, and can also integrate multiple complex and time-consuming steps of the analysis process like cleaning, dimensionality reduction, feature extraction and pattern recognition into a single pipeline which further reduces the demand for domain knowledge or data processing abilities (e.g., Mu et al., 2024; Mumuni & Mumuni, 2025). Unsurprisingly, an increasing number of studies that do use SMD to address research questions are also using ML models for these reasons (see T.K. et al., 2021, for an overview of recent applications and used models).

[1] See also Manuscript III: Brandenstein et al. (2024)

# 4 Showcasing the potential of machine learning in political psychology: Three original manuscripts

In the previous chapters, I outlined three shifts in how research in political psychology can be conducted: 1. Focusing on prediction, 2. Harnessing complex ML models and 3. Utilizing large datasets and novel data sources to their full potential. Using previous theoretical reflections and practical examples, I illustrated the potential of each shift and discussed how they can address limitations of current approaches. As noted previously, these shifts have not yet been widely adopted in political psychology, although recent progress has been made both within the field and in related disciplines, such as political science (e.g., De Slegte et al., 2024; Douglas et al., 2023; Grimmer et al., 2021; Sindermann et al., 2021). To showcase their value and promote broader use, I demonstrate how they can be applied in political psychology research using the three manuscripts underlying this publication-based dissertation. In these manuscripts, I investigated topics of high societal relevance such as conspiracy theory (*CT*) beliefs, sustainable behavior and social media behavior and aimed to contribute to the understanding of potential psychological mechanisms underlying these cognitions and behavior.

In Manuscript I [1], I analyzed psychological predictors of CT beliefs. Notably, no extensive theory on CT belief susceptibility exists yet and many previously investigated predictors differ in their strength - and even direction - of association with CT belief across studies. In this situation, using a prediction-oriented approach and ML models could help to obtain more reliable results of how the different predictors relate to CT belief (see chapter 1 and 2) which - in turn - may help political psychologists to progress in theory development. To achieve this, I used a medium-sized, representative dataset of more than 2000 UK citizens which was part of a larger, pre-existing panel study. This dataset included a variety of variables, including most of the previously investigated psychological predictors of CT belief as well as an established measurement of CT belief. Analyzing the data, I employed a linear OLS regression, a regularized (linear) Ridge regression (*RR*) and a non-linear Random Forest (*RF*) to predict CT belief from the psychological predictors and evaluated model performance using CV. The results revealed that the RF model showed the best prediction performance, followed by the RR and the OLS performing the worst. Also, the

[1] Manuscript I: Brandenstein (2022)

variance in prediction performance was lower for both regularized ML models than the standard OLS in the CV. These results supported the idea that past reliance on explanation-oriented, unbiased OLS models could have contributed to poor prediction performance and thus varying results across studies. Further, the performance increase of the complex, non-linear RF over the linear models indicated various, interactive effects between the predictors with CT belief. To uncover and further analyze them, I applied model-agnostic interpretation techniques (ALE plots and permutation feature importance). Starting with overall predictor importance in the RF, the results showed predictors like socio-political control, trust in official institutions or paranoid ideation to be the most relevant. Although these key predictors aligned with those most consistently reported across recent studies, some predictors previously identified as important showed less relevance in my study (e.g., analytical thinking style). Analyzing the relationships between these important predictors with CT belief uncovered several non-linear effects. These results join the ranks of very recent studies, showing other CT belief predictors to exhibited non-linear effects as well (e.g., Imhoff et al., 2022).

In Manuscript II [2], I aimed to identify key predictors of individual greenhouse gas (*GHG*) emissions in everyday life. Although different predictors have been identified over the years, past interventions and policies still failed to elicit meaningful changes. I argued that a strong reliance on explanation-oriented modeling has contributed to misguided conclusions about the impact of single predictors and their interaction with each other in the real-world. To address this, I applied a prediction-oriented approach on a large, representative sample of over 10.000 German citizens which I surveyed on different psychological, social and situational predictors potentially affecting sustainable behavior. I compared the prediction performance of a linear OLS and three ML models (Lasso regression, Support Vector Machine and RF) using CV as well as a separate train-test split. The results showed that the RF model outperformed all other models in the CV and the held-out test set. Again, the OLS performed the worst in predicting the outcome and showed the largest prediction variance in the CV among all candidate models. Similar to belief in CTs, this result suggested that previous explanation-oriented linear models may have missed important patterns between various predictors and sustainable behavior - potentially limiting the effectiveness of past climate action strategies. Further analyses of the best performing RF using model-agnostic SHAP values revealed several complex relationships between different predictors with sustainable behavior and helped to identify the most important predictors. Despite some notable differences, the identified key predictors largely aligned with those most consistently reported across previous studies, such as the importance of income or infrastructural barriers for mobility-related behavior. In addition, the RF revealed multiple interaction effects between predictors, such as pro-environmental attitudes and emotions with added travel time, as well as complex relationships

[2] Manuscript II: Brandenstein et al. (2023)

between predictors with sustainable behavior such as job status and mobility related emissions. Importantly, such intricate patterns have not been discovered or reported in previous studies.

Considering the potential of SMD and their limited utilization in political psychology mentioned in the previous chapter, I aimed to use network data of German Twitter (X) users to predict individual following decisions as well as political ideology and behavior in Manuscript III [3]. To achieve this, I first compiled a large dataset of over 270,000 German Twitter users, including information on which of more than 1,400 accounts of German politicians each user followed. I then applied a prediction oriented approach using a neural-network-based Variational Autoencoder (*VAE*) ML model and an unbiased, linear Correspondence Analysis (*CA*) to capture important patterns underlying these following decisions. This captured information was then used to predict following decisions for other users (out-of-sample) as well as self-reported political ideology and voting behavior of a subsample of users. The results showed that the prediction-focused VAE ML model outperformed the linear CA in predicting following decisions as well as self-reported political ideology and behavior. Further, the VAE ML model captured and revealed complex relationships in Twitter users' following decisions which not only helped to improve predicting the outcomes but also supported the "homophily assumption", a theoretical idea for how social networks form. Moreover, the VAE was computationally efficient and integrated multiple analysis steps which had to be performed manually in the CA.

Taken together, the findings from all three manuscripts demonstrate how the outlined shifts can be applied to political psychological research and underline their potential to gain important insights into how political cognitions and behavior can be shaped. Specifically, adopting a prediction-oriented modeling approach not only made it possible to evaluate the reliability of obtained model results and select the best performing model but also underlined the limitations of commonly-used explanation-oriented (linear) models. The use of complex ML models allowed to uncover intricate relationships among predictors of cognitive and behavioral outcomes and helped to identify the most important predictors in each study. These insights can serve as a basis to further develop and refine existing theories — for example, by testing causal relationships — and ultimately contribute to a deeper understanding of the psychological processes underlying political cognitions and behavior. In turn, this may also help inform practice, such as guiding policymakers in designing more effective interventions targeting key determinants of political cognitions and behavior. As described in the manuscripts, these findings were largely derived using large, information-rich datasets, two of which were openly accessible [1,3]. In this context, the results of the third manuscript [3] particularly highlighted the unique potential of social media data as a novel data source for studying political cognitions and behavior.

[3]Manuscript III: Brandenstein et al. (2024)

# 5 A potential future of political psychological research

In this dissertation, I outline several shifts to accelerate the knowledge generation process in political psychology and showcased their potential using the three underlying manuscripts in the previous chapter. While these shifts offer notable benefits, they also come with limitations that political psychologists should take into account when considering their adoption.

In the first chapter, I discussed the potential of focusing on prediction and incorporating a prediction-oriented modeling approach into political psychological research. Although doing so can provide considerable benefits, as shown in the previous theoretical works and my studies, this does not imply that political psychologists should abandon the explanation-oriented approach altogether. After all, explanation-oriented approaches such as controlled experimental studies to isolate causal effects or theory-driven hypotheses testing have undoubtedly helped political psychology to advance significantly and not only remain crucial for a deeper understanding of political cognitions and behavior but can help policymakers to design policies and regulation strategies (Hofman et al., 2021; Yarkoni & Westfall, 2017). However, researchers could certainly benefit from incorporating prediction-oriented practices — such as model evaluation techniques (e.g., cross-validation) or using biased versions of their frequently used models (e.g., regularized regression or multilevel models) — into their current workflow since they can be easily integrated and meaningfully address some of the discussed challenges associated with the explanation-oriented approach (Yarkoni & Westfall, 2017). A recent perspective article has also considered the potential of integrating and combining the prediction- and explanation-oriented approach more directly (Hofman et al., 2021). In detail, a combined approach would aim to predict an outcome in changing situations - such as the spread of (mis-)information that has been experimentally manipulated beforehand - which would enable researchers to identify predictive causal explanations of an outcome (Hofman et al., 2021). Another possibility of integrating both approaches was proposed in a recent study on moral dilemmas (Agrawal et al., 2020): The authors first deployed a complex ML model to reveal how well decisions in moral dilemmas are predictable, and then refined simpler, theory-driven models by comparing them to the ML model. The ultimate goal was to improve the simple model based on theory until it matched the ML model's predictions to maximize both explanatory and predictive power. Although more research is needed in this regard, such first ideas are exciting news and may help political psychology to acquire knowledge much faster by harnessing the unique advantages of both approaches.

In the second chapter, I aimed to highlight the complexity of our psychological world and argued for accounting for the multitude of influencing factors when studying political cognitions and behavior. Concurrently, I showed that ML models offer the most effective means of capturing this complexity and thus argued for their increased utilization in political psychology. Although following these recommendations can provide a lot of benefits, as discussed and shown throughout this dissertation, there are a few considerations to keep in mind. First, analyzing the influence of multiple predictors does not warrant to blindly include any predictor in a model for the sole purpose of improving its predictive capabilities. In this regard, political psychologists should also be aware of other potential pitfalls like the unsuccessful separation of predictors and outcomes (e.g., Brandenstein et al., 2025). To deal with this, researchers could evaluate the suitability of predictors prior to the modeling stage by using theory and (if available) previous study results as a guideline to select potential predictors (e.g., Elhai & Montag, 2020; Hofman et al., 2021); an approach which I also followed in my manuscripts. These recommendations present yet another way how explanation-oriented and prediction-oriented research practices could be integrated more closely. Next, political psychologists aiming to use ML models in their studies should consider a few points to maximize their potential and avoid drawing misguided conclusions. First, researchers should avoid deploying a single ML model only. Instead, they may assess the suitability of both simpler, explanation-oriented as well as various ML models in relation to their research question, while also considering the potential trade-off between model performance and interpretability (e.g., Grimmer et al., 2021). As mentioned in the first chapter, this is best achieved by following a prediction-oriented modeling approach including model evaluation and comparison techniques such as CV. Although most ML models need few assumptions regarding data and variable distributions and are able to learn complex relationships between variables automatically, they still need to be carefully "tuned" in order to effectively utilize their advantages (e.g., De Slegte et al., 2024). Improper tuning or applying models that are too complex for the dataset and research question can backfire and lead to overfitting and unstable predictions (e.g., Hofman et al., 2021; Yarkoni & Westfall, 2017). That being said, different methods to tune ML models algorithmically exist nowadays which significantly reduce these risks and save researchers much time and effort in trying to find optimal settings. Similar to model selection and tuning, political psychologists should also pay close attention when interpreting their ML models. Although many contemporary techniques (such as SHAP) offer deep, previously unthinkable insights into complex ML models, researchers must first make sure to understand how these techniques work, how to accurately interpret and report their results and be mindful of limitations that can arise from specific dataset characteristics. Luckily, well-written resources such as Molnar (2022) deliver detailed explanations of various interpretation methods, further facilitating quick adoption and appropriate use of ML models in political psychology. Arguably, the current fast-paced developments of new ML models tailored to different data types and use cases will only increase their future potential for research in political psychology (see also De Slegte et al., 2024). As discussed in the previous chapter, modern ML models are not only becoming more powerful but also much easier to use. Therefore, researchers may be best advised to familiarize themselves with ML models right now to quickly apply them and utilize their potential in the future.

Finally, in the third chapter, I argued for making more effective use of existing, large survey datasets and the potential offered by novel data sources to study political cognitions and behavior. Despite their unique benefits - discussed throughout the chapter and demonstrated in my studies - an exclusive shift to large survey datasets and data sources like social media is neither feasible nor advisable. Naturally, pre-existing survey data or social media data cannot replace the need for experimental studies to identify causal effects or conducting small-scale observational studies to address a very particular or novel research question. Nevertheless, they pose yet another accessible option for new and established political psychologists alike to catalyze their research. Presumably, the increased acceptance of and participation in open data practices across research disciplines will only increase the availability (and sheer number) of such datasets in the future. In turn, this will provide political psychologists with even more possibilities to study important political cognitions and behavior. To contribute to this movement, the political psychological community should strengthen collaborative efforts across research groups and countries to collect large, information-dense datasets and make them accessible to other researchers. As shown in the last two chapters, data from novel data sources like social media platforms can be particularly valuable for political psychologists and may also help to address major societal challenges such as misinformation, political polarization, and climate change. Unfortunately, using these data sources has become increasingly difficult after major platforms either completely blocked scholarly access or now demand large payments in recent years. Therefore, political psychologist should support and advocate legislation that grants researchers access to user data for scientific research purposes [1].

Wide-reaching adoption of the outlined shifts and recommendations in this dissertation may not be an easy task for the political psychological community. But what would it take to achieve this? Arguably, the biggest obstacle might be long established habits in how research is (or should be) conducted. As mentioned before, the vast majority of (political) psychologists are conducting explanation-oriented research and have probably never used complex predictive ML models or social media data. Importantly, I do not think that this is because researchers have never heard about these options but because they are either not convinced of their potential or lack expertise in how

[1] See also Manuscript III: Brandenstein et al. (2024)

to apply and effectively utilize them. I hope that fellow (political) psychologists who have read my dissertation this far no longer fall into the first category. In any case, however, increasing adoption rates certainly requires communicating the potential of the outlined shifts effectively. In this context, researchers already using prediction-focused ML models or large datasets may try to increase the visibility of their (published) work by, for instance, presenting their work at conferences or use different media outlets to share their findings. Addressing the second reason for why political psychologists may hesitate to adopt the outlined shifts seems more difficult. Although interested researchers may still decide to self-learn these unfamiliar methods, such efforts are unlikely to create wide-reaching changes in the political psychological community. Arguably, a more effective option would be to train aspiring researchers and students in underlying fields, such as psychology or political science, early on. Teaching these methods alongside the long-established ones would enable future political psychologist to decide when and how to use them effectively. Political psychologists may also consider collaborating more frequently with researchers from other domains already familiar with these methods, such as computer scientists. Working in such interdisciplinary teams further allows to bundle resources and share workloads across team members which can lead to more impactful research outcomes.

As illustrated by the various examples throughout this dissertation, political psychology plays a crucial role in understanding the human dimension of societal challenges. In the past, research in this field has revealed valuable insights into the psychological mechanisms underlying individual political cognitions and behavior. Considering the current state of the field and ongoing developments, I am optimistic that political psychology will continue to provide us with findings that can help tackling current societal challenges such as climate change, polarization and misinformation. In this regard, I hope that the shifts outlined in this dissertation, along with the underlying manuscripts, offer a small contribution and support the field's growth.

# References


Abbas, A. M. (2021). Social network analysis using deep learning: Applications and schemes. *Social Network Analysis and Mining*, *11*(1), 106. https://doi.org/10.1007/s13278-021-00799-z

Abou-Zeid, M., Witter, R., Bierlaire, M., Kaufmann, V., & Ben-Akiva, M. (2012). Happiness and travel mode switching: Findings from a Swiss public transportation experiment. *Transport Policy*, *19*(1), 93–104. https://doi.org/10.1016/j.tranpol.2011.09.009

Agrawal, M., Peterson, J. C., & Griffiths, T. L. (2020). Scaling up psychology via Scientific Regret Minimization. *Proceedings of the National Academy of Sciences*, *117*(16), 8825–8835. https://doi.org/10.1073/pnas.1915841117

American National Election Studies. (n.d.). American National Election Studies (ANES). www.electionstudies.org

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Brandenstein, N. (2022). Going beyond simplicity: Using machine learning to predict belief in conspiracy theories. *European Journal of Social Psychology*, *52*(5-6), 910–930. https://doi.org/10.1002/ejsp.2859

Brandenstein, N., Ackermann, K., Aeschbach, N., & Rummel, J. (2023). The key determinants of individual greenhouse gas emissions in Germany are mostly domain-specific. *Communications Earth & Environment*, *4*(1), 422. https://doi.org/10.1038/s43247-023-01092-x

Brandenstein, N., Ackermann, K., & Rummel, J. (2025). The trouble with carbon footprint analysis in behavioral climate research. *Climatic Change*, *178*(3), 57. https://doi.org/10.1007/s10584-025-03900-z

Brandenstein, N., Montag, C., & Sindermann, C. (2024). To follow or not to follow: Estimating political opinion from Twitter data using a network-based machine learning approach. *Social Science Computer Review*, *0*(0). https://doi.org/10.1177/08944393241279418

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Cottam, M. L., Dietz-Uhler, B., Mastors, E., & Preston, T. (2004). What is political psychology? In *Introduction to political psychology* (1st ed., pp. 3–11). Erlbaum.

De Slegte, J., Van Droogenbroeck, F., Spruyt, B., Verboven, S., & Ginis, V. (2024). The Use of Machine Learning Methods in Political Science: An In-Depth Literature Review. *Political Studies Review*, *0*(0). https://doi.org/10.1177/14789299241265084

Dong, X., & Lian, Y. (2021). A review of social media-based public opinion analyses: Challenges and recommendations. *Technology in Society*, *67*, 101724. https://doi.org/10.1016/j.techsoc.2021.101724

Douglas, K. M., Sutton, R. M., Van Lissa, C. J., Stroebe, W., Kreienkamp, J., Agostini, M., Bélanger, J. J., Gützkow, B., Abakoumkin, G., Khaiyom, J. H. A., Ahmedi, V., Akkas, H., Almenara, C. A., Atta, M., Bagci, S. C., Basel, S., Berisha Kida, E., Bernardo, A. B. I., Buttrick, N. R., … Leander, N. P. (2023). Identifying important individual- and country-level predictors of conspiracy theorizing: A machine learning analysis. *European Journal of Social Psychology*, *53*(6), 1191–1203. https://doi.org/10.1002/ejsp.2968

Elhai, J. D., & Montag, C. (2020). The compatibility of theoretical frameworks with machine learning analyses in psychological research. *Current Opinion in Psychology*, *36*, 83–88. https://doi.org/10.1016/j.copsyc.2020.05.002

European Social Survey. (n.d.). European Social Survey. https://www.europeansocialsurvey.org/

Feldman, S., & Johnston, C. (2014). Understanding the Determinants of Political Ideology: Implications of Structural Complexity. *Political Psychology*, *35*(3), 337–358. https://doi.org/10.1111/pops.12055

Fudenberg, D., & Liang, A. (2020). Machine learning for evaluating and improving theories. *ACM SIGecom Exchanges*, *18*(1), 4–11. https://doi.org/10.1145/3440959.3440962

Gerber, A., Huber, G., & Fang, A. (2018). Do Subtle Linguistic Interventions Priming a Social Identity as a Voter Have Outsized Effects on Voter Turnout? Evidence From a New Replication Experiment. *Political Psychology*, *39*(4), 925–938. https://doi.org/10.1111/pops.12446

González-Bailón, S., Lazer, D., Barberá, P., Godel, W., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A., Iyengar, S., Kim, Y., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C., Settle, J., Thorson, E., … Tucker, J. (2024). The Diffusion and Reach of (Mis)Information on Facebook during the U.S. 2020 Election. *Sociological Science*, *11*, 1124–1146. https://doi.org/10.15195/v11.a41

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small Effects: The Indispensable Foundation for a Cumulative Psychological Science. *Perspectives on Psychological Science*, *17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science*, *24*(1), 395–419. https://doi.org/10.1146/annurev-polisci-053119-015921

Gruber, J., Saxbe, D., Bushman, B. J., McNamara, T., & Rhodes, M. (2019). How Can Psychological Science Contribute to a Healthier, Happier, and More Sustainable World? *Perspectives on Psychological Science*, *14*(1), 3–6. https://doi.org/10.1177/1745691618821624

Hankar, M., Kasri, M., & Beni-Hssane, A. (2025). A comprehensive overview of topic modeling: Techniques, applications and challenges. *Neurocomputing*, *628*, 129638. https://doi.org/10.1016/j.neucom.2025.129638

Hindman, M. (2015). Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 48–62. https://doi.org/10.1177/0002716215570279

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*, 181–188. https://doi.org/10.1038/s41586-021-03659-0

Hunecke, M., Haustein, S., Grischkat, S., & Böhler, S. (2007). Psychological, sociodemographic, and infrastructural factors as determinants of ecological impact caused by mobility behavior. *Journal of Environmental Psychology*, *27*(4), 277–292. https://doi.org/10.1016/j.jenvp.2007.08.001

Imhoff, R., Zimmer, F., Klein, O., António, J. H. C., Babinska, M., Bangerter, A., Bilewicz, M., Blanuša, N., Bovan, K., Bužarovska, R., Cichocka, A., Delouvée, S., Douglas, K. M., Dyrendal, A., Etienne, T., Gjoneska, B., Graf, S., Gualda, E., Hirschberger, G., … Van Prooijen, J.-W. (2022). Conspiracy mentality and political orientation across 26 countries. *Nature Human Behaviour*, *6*(3), 392–403. https://doi.org/10.1038/s41562-021-01258-7

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R* (1st ed.). Springer.

Kemp, S. (2024, April). *Digital 2024 April Global Statshot Report* (tech. rep.). Datareportal. Retrieved April 7, 2025, from https://datareportal.com/reports/digital-2024-april-global-statshot

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, *8*(3), 239–260. https://doi.org/10.1080/13504620220145401

McBride, O., Murphy, J., Shevlin, M., Gibson-Miller, J., Hartman, T. K., Hyland, P., Levita, L., Mason, L., Martinez, A. P., McKay, R., Stocks, T. V., Bennett, K. M., Vallières, F., Karatzias, T., Valiente, C., Vazquez, C., & Bentall, R. P. (2021). Monitoring the psychological, social, and economic impact of the COVID-19 pandemic in the population: Context, design and conduct of the longitudinal COVID-19 psychological research consortium (C19PRC) study. *International Journal of Methods in Psychiatric Research*, *30*(1), e1861. https://doi.org/10.1002/mpr.1861

Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). https://christophm.github.io/interpretable-ml-book

Mu, Y., Dong, C., Bontcheva, K., & Song, X. (2024). Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling [Preprint]. https://doi.org/10.48550/ARXIV.2403.16248

Mumuni, A., & Mumuni, F. (2025). Automated data processing and feature engineering for deep learning and big data applications: A survey. *Journal of Information and Intelligence*, *3*(2), 113–153. https://doi.org/10.1016/j.jiixd.2024.01.002

Nielsen, K. S., Cologna, V., Bauer, J. M., Berger, S., Brick, C., Dietz, T., Hahnel, U. J. J., Henn, L., Lange, F., Stern, P. C., & Wolske, K. S. (2024). Realizing the full potential of behavioural science for climate change mitigation. *Nature Climate Change*, *14*(4), 322–330. https://doi.org/10.1038/s41558-024-01951-1

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., … Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, *620*(7972), 137–144. https://doi.org/10.1038/s41586-023-06297-w

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Rafaeli, A., Ashtar, S., & Altman, D. (2019). Digital Traces: New Data, Resources, and Tools for Psychological-Science Research. *Current Directions in Psychological Science*, *28*(6), 560–566. https://doi.org/10.1177/0963721419861410

Reveilhac, M., Steinmetz, S., & Morselli, D. (2022). A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion. *Multimedia Tools and Applications*, *81*(7), 10107–10142. https://doi.org/10.1007/s11042-022-12101-0

Rousidis, D., Koukaras, P., & Tjortjis, C. (2020). Social media prediction: A literature review. *Multimedia Tools and Applications*, *79*(9-10), 6279–6311. https://doi.org/10.1007/s11042-019-08291-9

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social Media Analyses for Social Measurement. *Public Opinion Quarterly*, *80*(1), 180–211. https://doi.org/10.1093/poq/nfv048

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

Sindermann, C., Mõttus, R., Rozgonjuk, D., & Montag, C. (2021). Predicting current voting intentions by Big Five personality domains, facets, and nuances – A random forest analysis approach in a German sample. *Personality Science*, *2*(1). https://doi.org/10.5964/ps.6017

Skoric, M. M., Liu, J., & Jaidka, K. (2020). Electoral and Public Opinion Forecasts with Social Media Data: A Meta-Analysis. *Information*, *11*(4), 187. https://doi.org/10.3390/info11040187

Swami, V., Coles, R., Stieger, S., Pietschnig, J., Furnham, A., Rehim, S., & Voracek, M. (2011). Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories: Conspiracist ideation. *British Journal of Psychology*, *102*(3), 443–463. https://doi.org/10.1111/j.2044-8295.2010.02004.x

T.K., B., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, *40*, 100395. https://doi.org/10.1016/j.cosrev.2021.100395

Van Valkengoed, A. M., Abrahamse, W., & Steg, L. (2022). To select effective interventions for pro-environmental behaviour change, we need to consider determinants of behaviour. *Nature Human Behaviour*, *6*, 1482–1492. https://doi.org/10.1038/s41562-022-01473-w

Van Valkengoed, A. M., & Steg, L. (2019). Meta-analyses of factors motivating climate change adaptation behaviour. *Nature Climate Change*, *9*(2), 158–163. https://doi.org/10.1038/s41558-018-0371-y

Walker, R. M., Zhang, J., & Cheng, E. W. (2024). Puncturing the Knowledge Illusion Does Not Reduce Policy and Political Extremism: Evidence From a Replication and Extension. *Political Psychology*, *45*(3), 581–599. https://doi.org/10.1111/pops.12938

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

# Appendix A1 - Manuscript I

Manuscript I: Going beyond simplicity: Using machine learning to predict belief in conspiracy theories

Note:
This is the final version of the open-access article published in *European Journal of Social Psychology*. The reproduction of this article in this dissertation is permitted under the respective Creative Commons Attribution 4.0 International License.

Brandenstein, N. (2022). Going beyond simplicity: Using machine learning to predict belief in conspiracy theories. *European Journal of Social Psychology*, *52*(5-6), 910–930. https://doi.org/10.1002/ejsp.2859

Received: 17 August 2021 | Accepted: 23 April 2022

DOI: 10.1002/ejsp.2859

EASP WILEY

RESEARCH ARTICLE

# Going beyond simplicity: Using machine learning to predict belief in conspiracy theories

Nils Brandenstein

Department of Psychology, Heidelberg University, Heidelberg, Germany

**Correspondence**
Nils Brandenstein, Heidelberg University, Hauptstr. 47–51, D-69117 Heidelberg, Germany.
Email: nils.brandenstein@psychologie.uni-heidelberg.de

## Abstract

Public and scientific interest in why people believe in conspiracy theories (CT) surged in the past years. To come up with a theoretical explanation, researchers investigated relationships of CT belief with psychological factors such as political attitudes, emotions, or personality. However, recent studies have put the robustness of these relationships into question. In the present study, a prediction-based analysis approach and machine learning models are deployed to detect and remedy poor replicability of CT belief associations. The analysis of a representative dataset with 2025 UK citizens supports the assumption that the current simplicity of the field's analysis routine, exhibiting high sample-specificity and neglecting complex associations of psychological factors with CT belief, may obscure important relationships. The results further point towards key components of conspiratorial mindsets like general distrust and low socio-political control. Important implications for building a coherent theory of CT belief are derived.

KEYWORDS
belief, conspiracy, machine learning, political attitudes, prediction, social psychology

## 1 | INTRODUCTION

In light of the recent COVID-19 pandemic, conspiracy theories (CTs) and their proponents have gained a lot of public and scientific attention. CTs are usually defined as a set of narratives that contradicts an official explanation for a given event, causally blaming a small special interest group with a malevolent plot for its occurrence (e.g., Douglas et al., 2016; Swami et al., 2011). Since CTs permeate countries and social classes (e.g., Oliver & Wood, 2014) and can crucially influence our way of thinking (e.g., Douglas & Sutton, 2008; Lewandowsky, Oberauer et al., 2013) as well as our behaviour (e.g., Imhoff & Lamberty, 2020), their investigation is also of applied socio-political relevance. Although many CTs contain radical views regarding the causes for certain events, believing in CTs seems to be a general human phenomenon (Miller et al., 2016). As a result, scholars have tried to understand what makes people susceptible to CT belief. Douglas et al. (2017) argued that CTs appeal to people when they promise to satisfy psychological needs, which can be classified into epistemic motives (e.g., the need for understanding, sense-making), existential motives (e.g., the need for security and having control over one's environment), and social motives (e.g., the need of belonginess and maintaining a positive self and group image). Especially in comparison with non-conspiratorial explanations for given events, CTs can serve as a simple and consistent set of causal explanations, which arguably satisfy these psychological needs. Since not all people adhere to CT belief to the same extent, researchers investigated a variety of psychological factors like socio-political attitudes and personality factors that may account for individual differences in susceptibility and motive satisfaction. Consequently, robust associations between CT belief and psychological factors should be found that are directly linked to epistemic, existential, and social motives. Upon closer inspection of previous studies, however, the result pattern is not as clear-cut as one may expect.

### 1.1 | Epistemic motives in relation to CT belief

First, concerning studies on epistemic motives, a host of previous research has shown that belief in CTs is higher among people experiencing feelings of uncertainty and/or showing little dispositional tolerance for uncertainty (e.g., Maftei & Holman et al., 2022; Van Prooijen & Jostmann, 2013). Feelings of uncertainty are directly linked to the epistemic motive and sense-making process and CT belief adaptation may serve as a means to reduce them. However, some studies did not find an association between dispositional uncertainty and CT belief (e.g., Moulding et al., 2016).

Furthermore, scholars have argued that low levels of education and impoverished analytical thinking styles should increase belief in CTs (e.g., Douglas et al., 2016; Van Prooijen, 2017). Since CTs deliver overly simplified explanations of complex events, they often have an intuitive appeal. Thus, people with lower cognitive abilities and less analytical thinking may be more likely to accept CTs at face value. On the other hand, higher levels of education and analytical thinking style should lead to more scepticism and scrutiny of CTs, which should reduce belief in them. Although empirical studies generally support this theoretical notion (Douglas et al., 2016; Swami et al., 2014; Van Prooijen, 2017), Ståhl and Van Prooijen (2018) showed that high levels of analytical thinking only reduce belief in CTs, if people also value epistemic rationality, that is the motivation to form beliefs on rational grounds.

High levels of religious denomination have also been shown to be positively associated with belief in CTs. Scholars expected this relationship because coincidental events can also be seen as deliberately caused by a powerful agent (i.e., God, influential groups), a feature that both CT belief and religious belief share (e.g., Newheiser et al., 2011). Upon further investigation, however, Jasinskaja-Lahti and Jetten (2019) were able to show that this positive association of religiosity and CT belief is alleviated when accounting for trust in official institutions and that religious and non-religious people adhere to CTs to the same extent.

More general personality traits like the Big Five (Extraversion, Openness, Agreeableness, Neuroticism, Conscientiousness) have also been linked to epistemic motives and increased CT belief in the past. In this regard, CTs may, for instance, satisfy curiosity for people high in Openness (Swami et al., 2013), help individuals with high levels of Neuroticism to reduce uncertainty (Hollander, 2018), or bolster suspicion and antagonism towards others for people low in Agreeableness (Swami & Furnham, 2012). In a recent meta-analysis, Goreis and Voracek (2019) further investigated these proposed relationships. They were, however, unable to find evidence for a substantial linkage between any of the Big Five personality traits and CT belief. To summarize previous study results on epistemic motives, psychological factors like analytical thinking style and (to some extent) uncertainty seem to be the more robust correlates of CT belief, while general personality traits or religious denomination may be far less suited to explain individual belief differences.

### 1.2 | Existential motives in relation to CT belief

Turning to existential motives, a host of studies have shown that people are more likely to adapt CT belief when they experience a lack of control over their life and what is happening in the world (e.g., Abalakina-Paap et al., 1999; Imhoff, 2015; Van Prooijen & Acker, 2015), which directly triggers the existential motive to feel safe and secure. In a recent study, Stojanov et al. (2020) conducted six experiments on the relationship of perceived lack of control and CT belief and showed the constructs to be correlated in general. However, they did not find a substantial effect of experimentally manipulated lack of control on CT belief adaptation, which indicates trait rather than state lack of control to be associated with CT belief. Experiencing a stable, continuous lack of control in everyday life might thus be more relevant for CT belief adaptation than a temporary, situationally triggered lack of control. Furthermore, this linkage seems to depend on the type of control beliefs. While socio-political control (i.e., perceived influence in the political decision-making process) seems to be robustly associated with CT belief, results regarding the relationship between personal control (i.e., achieving goals) or interpersonal control (i.e., control over social situations) and CT belief are mixed (Imhoff & Lamberty, 2018). Researchers also investigated the disposition to perceive structure in random and unrelated events as a predictor of CT belief. While some studies found evidence for these dispositional “nothing happens by accident” beliefs to be positively related to CT belief (e.g., Moulding et al., 2016; Van Prooijen et al., 2018), others found no difference between individuals with high and low priors for randomness (Dieguez et al., 2015).

Further, low self-esteem has been linked to higher CT endorsement in the past (e.g., Swami et al., 2011). Arguably, individuals with low self-esteem may be more likely to believe CTs since holding these beliefs enables them to blame others for their personal problems (Abalakina-Paap et al., 1999). This external attribution may then help individuals to deal with feelings of insecurity arising from personal problems. However, this linkage does not seem to be very robust either (cf. Cichocka et al., 2016).

The same holds true for the construct of death anxiety (i.e., dispositional fear of death and its consequences), which arguably triggers the existential motive of feeling safe and secure the most. People high in death anxiety are assumed to be more susceptible to CTs, since holding these beliefs may alleviate feelings of existential threat by making the future more predictable and help individuals to understand the root cause of their feelings (Van Prooijen, 2019). Although a study by Newheiser et al. (2011) found a positive relationship between death anxiety and endorsement of a specific CT belief, others did not (e.g., Bruder et al., 2013).

Recent preliminary findings also suggest that low dispositional resilience (i.e., the ability to bounce back from stress) increases CT belief adaptation (Miller, 2020). Individual differences in dispositional resilience are assumed to be related to the coping strategies used to deal with stressful life events (e.g., Wood & Bhatnagar, 2015). People low in resilience often use passive, dysfunctional coping strategies like avoidance and blaming others (Yi et al., 2005). As a result, vulnerable individuals may be more inclined to believe CTs as a coping strategy to reduce feelings of existential threat and to make sense of the stressful situation (see Van Prooijen, 2019 for a discussion on existential motives and sense-making processes). However, more research is needed to support this idea and clarify the functional basis of this relationship.

In summary, research on existential motives indicates that dispositional control beliefs (especially socio-political control) might be a key component of CT belief, while other psychological factors like self-esteem and death-anxiety seem to be less relevant.

### 1.3 | Social motives in relation to CT belief

Finally, research on psychological factors fuelling social motives for CT belief adaptation involves group memberships, socio-political attitudes, and trust. A fair number of previous studies focused on the relationship of Right-Wing-Authoritarianism (*RWA*) and Social Dominance Orientation (*SDO*) with CT belief. Although being markedly different constructs, both RWA and SDO correspond to generalized attitudes and prejudice towards distinct groups (Imhoff & Bruder, 2014), which in turn may increase susceptibility to CT beliefs. More specifically, people high in RWA exhibit negative attitudes towards nonconforming, potentially dangerous individuals, while people high in SDO show negative attitudes towards derogated, low-status groups (e.g., Asbrock et al., 2010). In this regard, CTs may be used to further bolster these generalized attitudes and satisfy the social motive of belonginess and positive in-group evaluation. Consistent with this idea, most researchers found RWA and SDO to be positively correlated with CT belief (e.g., Grzesiak-Feldman & Irzycka, 2009; Kofta et al., 2020; Swami, 2012). However, the association of RWA and/or SDO with greater CT endorsement does not always seem to arise (e.g., Imhoff & Bruder, 2014; Oliver & Wood, 2014). Similar to these group perceptions, narcissism and collective narcissism have been proposed to be associated with CT belief adaptation fuelled by defensive motives (defending the view of oneself or one's in-group). In general, studies seem to support this assumption (e.g., Cichocka et al., 2016; Golec de Zavala & Federico, 2018; Kay, 2021).

Moreover, different forms of institutional trust (e.g., trust in government, parties, science) as well as interpersonal trust have been the focus of investigation in many previous studies. The overwhelming majority supports the idea that lower trust in these entities is associated with stronger CT belief (e.g., Bruder & Kunert, 2022; Einstein & Glick, 2015; Lewandowsky, Gignac et al., 2013). In a nationwide study conducted by Miller et al. (2016) in the US, trust in government and interpersonal trust functioned as a moderator of the relationship between political ideology, knowledge, and CT beliefs. Therefore, trust may also be a more complex psychological factor, influencing the relationship of other psychological factors with CT belief (cf. Mari et al., 2022; Miller et al., 2016).

Somewhat similar to interpersonal trust, the concept of subclinical paranoid ideation, that is, being suspicious of other people's behaviours and intentions in general (Van Prooijen & Van Lange, 2021, p. 240), has recently been gaining much attention in the CT belief literature. Studies have shown that paranoid ideation seems to be closely related to CT belief (e.g., Darwin et al., 2011; Grzesiak-Feldman & Ejsmont, 2008). However, since paranoid ideation and conspiracy belief share the same notion of assuming people to pursue malicious plans, their construct independence has been questioned. In a recent meta-analysis, Imhoff and Lamberty (2018) found further empirical support for their association but also showed that they indeed seem to be different constructs. However, the functional basis of their relationship may still be open to debate.

Consistent with the social motive of the desire to belong, research has shown that people generally exhibit higher levels of CT belief if they feel ostracized and isolated from society (e.g., Leman & Cinnirella, 2013; Moulding et al., ; Poon et al., 2020). In this regard, CT belief allegedly helps individuals to deal with their loneliness and feel connected again (Douglas et al., 2017). In conclusion, social motives seem to be highly interwoven with CT beliefs. More specifically, distrust in other people and official institutions, generalized social attitudes as well as feelings of isolation seem to go hand in hand with heightened CT belief susceptibility.

### 1.4 | Structural factors in relation to CT belief

On top of psychological motives, structural factors like age, gender, living area, and socio-economic status (SES) have also been investigated in CT belief research. The general line of reasoning to expect differences between these factors revolves around members of marginalized groups (e.g., minorities, low SES, women and youth), who, for instance, exhibit less perceived control or adhere more to paranoid thinking styles due to their social standing (cf. Darwin et al., 2011; Stempel et al., 2007). However, studies have yielded mixed results. Differences between genders remain highly obscure with no real observable trend (cf. Dyrendal et al., 2021; Mari et al., 2022; Swami et al., 2011). Similarly, studies reported CT belief to be positively related to age (Van Prooijen, 2017), negatively related (Jensen et al., 2021; Mari et al., 2022), or not related at all (Douglas et al., 2016). Differences in living areas which cause individuals to be more or less physically isolated from others have also been proposed to be predictive of CT belief. Again, some studies found individuals living in remote or rural areas to adhere more to CT beliefs (Mari et al., 2022; Siddiqui, 2020), while others did not (Sternisko et al., 2021). Socio-economic factors like income and—more generally—SES may also serve as potential predictors of CT belief. Indeed, stigmatization and feelings of permanent insecurity through material strain (i.e., precarity) associated with low income and SES has been shown to increase CT belief susceptibility (Adam-Troian et al., 2021). Nevertheless, a clear-cut relationship between income and SES with CT belief does not seem to arise (cf. Douglas et al., 2016, 2019; Freeman et al., 2020).

In general, the mixed pattern of relationships between structural factors and CT belief seems to suggest that there may be little to no difference between groups in general. Positive findings in previous literature might also be attributable to study sample characteristics or psychological factors that may mediate the relationship (e.g., existential threats, perceived control, or trust). However, there is still more research needed on structural factors to draw a conclusion.

## 1.5 | *Quo vadis*?—The current state of conspiracy theory research

Although this overview is non-exhaustive (see Table 1 for a more detailed report), it yet shows that several psychological factors are associated with individual CT belief susceptibility to different extents. Still, as indicated by the numerous contradicting findings, many of these associations seem to be inconsistent or obscure. However, obtaining robust relationships of psychological factors with CT belief tendencies is crucial in the current state of research, as a coherent theoretical framework is still lacking (Van Prooijen & Douglas, 2018). Addressing this issue, Stojanov and Halberstadt (2020) argued that these inconsistencies result from the field's discord over how to measure belief in CTs in the first place. Over the years, different CT belief scales and singular items with varying degrees of specificity have been used (e.g., Brotherton et al., 2013; Bruder et al., 2013). Consequently, study results may differ as a function of CT belief measurement. Although measurement plurality can certainly explain some of the inconsistencies in the literature, it does not explain why associations between psychological factors and CT belief differ between studies using identical assessment instruments or experimental and correlational designs. This is true for several of the previously mentioned psychological factors, like intolerance for uncertainty (cf. Maftei & Holman, 2022; Moulding et al., 2016), the Big Five (see Goreis & Voracek, 2019), self-esteem (see Cichocka et al., 2016), or internal control (see Stojanov et al., 2020), to mention just a few.

One way to deal with inconsistencies is certainly to explicitly consider measurement and design plurality as boundary conditions in current theorizing. However, I argue that inconsistencies in CT belief associations with psychological factors can also be tackled from another angle—namely by overcoming the simplicity of current analysis routines.

First, researchers regularly do not conduct cross-validation of their model results. This is problematic due to a phenomenon called *bias-variance-tradeoff*. In most cases when fitting an explanatory model, we are trying to reduce *bias*, so that the parameter estimates of the model fit the true parameter values most closely. Bias is contrasted by *variance*, describing a model's sensitivity to small changes in the dataset. Now, if researchers are only concerned about finding the best fit for one particular dataset (reducing bias) they will produce models that cannot be generalized to new data, as their variance increases drastically (Yarkoni & Westfall, 2017). Additionally, the regular use of frequentist statistics like Null hypothesis significance testing (*NHST*) and *p* (e.g., Bakker et al., 2012) in combination with small sample sizes can cause further generalization problems, as Stojanov et al. (2020) and OpenScienceCollaboration (2015) illustrated.

Second, most researchers assume linear relationships between psychological factors and CT belief, while many studies in psychology—even in CT research itself—show much more complex associations. Van Prooijen et al. (2015) and Imhoff et al. (2022), for instance, showed that the relationship between political ideology and CT belief follows a quadratic U-shape with the political extremes being more prone to CT belief. Finally, most studies only test single candidate factors for CT belief at a time, ignoring interactions and covariates. This can be problematic since complex psychological phenomena are likely to have many interplaying small-effect causes rather than few big ones (Götz et al., 2021). Goreis and Voracek (2019) argued that belief in CTs falls into this area, since individual beliefs can be influenced by a variety of proximal and distal factors and their interactions. Some of the previously stated exemplary findings on psychological factors support this notion empirically, such as the findings that analytical thinking is only negatively associated with CT belief if people also value rationality (Ståhl & Van Prooijen, 2018), the moderating role of knowledge and trust on the relation between political attitude and CT endorsement (Miller et al., 2016), or vanishing relationships between CT belief and psychological factors, when accounting for covariates like paranoid ideation (Imhoff & Lamberty, 2018). Thus, current practice in CT research can lead to an overestimation of the impact of single predictors. What is apparent from these problems is that contradictory results may be elicited by a mismatch between the current theoretical standing of the field and the study design typically used by researchers.

## 1.6 | The present research

In the present study, I suggest using a predictive analysis approach and machine learning models (ML) to address the discord of CT belief correlates in previous studies. A predictive approach aims to find the ratio between bias and variance that minimizes the prediction error on unseen data. In turn, models are obtained that replicate well and whose results are more trustworthy. This is achieved by using cross-validation and model regularization. The former allows us to estimate how well results replicate and the latter renders them more replicable by tuning the model fit (often by introducing bias, see above). Therefore, if our goal is to find robust associations between potential psychological factors and CT belief, a predictive approach will be arguably better suited for this purpose than just evaluating the model fit on one dataset, as has been done in most previous studies. On top of this, using ML models can help identify robust correlates of CT belief. Since many ML models are flexible and non-linear approximators, they can find the optimal bias-variance-ratio more easily and generally obtain better replicability than traditional methods (Hindman, 2015). Plus, ML models can account for all possible predictor relationships and interactions in the data. Recently, researchers in the social sciences have acknowledged these advantages and started using ML models in research areas in

**TABLE 1** Overview of variable relationships with belief in conspiracy theories in previous studies

| Variable | Author | Direction and strength of relationship with CT belief |
|---|---|---|
| Age | 1. Imhoff et al. (2022) | [O\|+], $b = .00$, $b = 0.34^{*}$ |
| | 2. Douglas et al. (2016) | [−\|O], $b = -0.16^{*}$, $b = 0.01$ |
| | 3. Van Prooijen (2017) | [+], $b = 0.15^{***}$ |
| Gender | 1. Dyrendal et al. (2021) | [O\|+], $t = -1.74$, $t = 2.88^{*}$ [male more CT belief] |
| | 2. Imhoff et al. (2022) | [O\|+], $b = 0.05$, $b = 0.34^{*}$ [male more CT belief] |
| | 3. Swami et al. (2011) | [+], $t = 2.62^{*}$ [women more CT belief] |
| Living area | 1. Siddiqui (2020) | [−\|O], $r = -.23^{*}$, $r = -.30^{*}$, $r = -.28^{*}$, $r = -.005$ |
| | 2. Marques et al. (2021) | [O\|−], $b = 0.04$, $b = 0.02$, $b = 0.06^{**}$, $b = 0.04^{*}$ |
| | 3. Stoica and Umbres (2021) | [O], $r = .003$, $r = .03$ |
| Education | 1. Douglas et al. (2016) | [O\|−], $r = -.13$, $r = -.26^{**}$ |
| | 2. Van Prooijen (2017) | [−], $r = -.15^{***}$; $r = -.26^{***}$ |
| | 3. Van Prooijen and Acker (2015) | [O\|−], $r = -.21^{***}$, $r = -.11^{***}$, $r = -.13^{***}$, $r = -.16^{***}$, $r = -.02$ |
| Socio-economic status (SES) / Income | 1. Adam-Troian et al. (2021) | [−], $r = -.25^{***}$ [precarity] |
| | 2. Freeman et al. (2020) | [O], $r = .02$, $r = -.02$ [income] |
| | 3. Douglas et al. (2016) | [O\|−], $r = -.13$, $r = -.12^{*}$ [income], $r = -.06$, $r = -.16$ [SES] |
| Religion | 1. Oliver and Wood (2014) | [O\|+], $r = .36^{***}$, $r = .10^{***}$, $r = .19^{***}$, $r = .00$, $r = .11^{***}$ [importance of religion] |
| | 2. Jasinskaja—Lahti and Jetten (2019) | [+\|−], $r = .25^{*}$, $r = -.04^{*}$ [importance of religion; when accounting for trust: negative] |
| | 3. Newheiser, Farias and Tausch (2011) | [−], $r = -.49^{***}$ [religiosity belief inventory] |
| Ideology | 1. Imhoff et al. (2022) | [+\|O], $b = 0.10^{***}$, $b = 0.19^{***}$ [quadratic effect], $b = 0.14^{**}$, $b = 0.12$ [linear effect] |
| | 2. Imhoff and Bruder (2014) | [O], $r = .07$ |
| | 3. Van Prooijen et al. (2015) | [+\|O], $b = -0.32^{*}$, quadratic: $b = 0.35^{*}$ [Study2a] |
| Right-wing authoritarianism (RWA) | 1. Imhoff and Bruder (2014) | [O\|+], $r = .05, .16, .29^{***}$ [Study 4] |
| | 2. Bruder et al. (2013) | [+], $r = .28^{***}$ |
| | 3. Imhoff and Lamberty (2018) | [+], $r = .21^{**}$ |
| Social dominance orientation (SDO) | 1. Imhoff and Bruder (2014) | [O], $r = .03, -.05, .14$ |
| | 2. Bruder et al. (2013) | [+], $r = .16^{*}$ |
| | 3. Dyrendal et al. (2021) | [+], $r = .24^{***}$, $r = .12^{**}$ [SEM paths] |
| Self-esteem | 1. Swami (2012) | [O], $r = -.03$ |
| | 2. Cichocka et al. (2016) | [O], $r = -.08$, $r = .05$ |
| | 3. Swami and Furnham (2012) | [−], $r = -.16^{**}$ |
| Paranoid ideation | 1. Bruder et al. (2013) | [+], $r = .45^{***}$ |
| | 2. Darwin et al. (2011) | [+], $r = .47^{**}$ |
| | 3. Imhoff and Lamberty (2018) | [+], $z = .38^{***}$ [meta-analysis] |
| Resilience | 1. Miller (2020) | [−], $b = -0.19^{*}$ [supplementary section C] |
| Death anxiety | 1. Imhoff and Bruder (2014) | [O], $r = .08$ |
| | 2. Newheiser et al. (2011) | [+], $r = .32^{***}$ |
| | 3. Bruder et al. (2013) | [O], $r = .10$ |

(Continues)

**TABLE 1** (Continued)

| Variable | Author | Direction and strength of relationship with CT belief |
|---|---|---|
| Control: Internal | 1. Bruder et al. (2013) | [O], $r = .03$ |
| | 2. Imhoff and Lamberty (2018) | [O], $r = -.13$, $r = .06$ |
| | 3. Kofta et al. (2020) | [−], $r = .14^{**}$ [study 4] |
| Control: Perception of randomness | 1. Moulding et al. (2016) | [+], $r = .26^{**}$ [World assumption scale: Randomness] |
| | 2. Van Prooijen et al. (2018) | [+], $r = .37^{***}$, $r = .23^{**}$ [illusory pattern perception] |
| | 3. Dieguez et al. (2015) | [O], $r = .13$, $r = -.04$, $r = -.10$ [prior subjective randomness] |
| Control: Socio-political | 1. Bruder et al. (2013) | [−], $r = -.22^{***}$ |
| | 2. Imhoff and Bruder (2014) | [−], $r = -.23^{***}$ |
| | 3. Imhoff and Lamberty (2018) | [−], $r = -.19^{**}$, $r = -.22^{**}$ |
| Uncertainty | 1. Moulding et al. (2016) | [O\|−], $r = .09$, $r = .16$, $r = .09$, $r = .13$, $r = .30^{**}$ |
| | 2. Kofta et al. (2020) | [−], $r = .13^{**}$ [study 4] |
| | 3. Miller (2020) | [−], $b = .22^{***}$ [supplementary section C] |
| Interpersonal trust | 1. Green and Douglas (2018) | [−], $b = −0.24^{**}$ |
| | 2. Imhoff and Lamberty (2018) | [−], $r = -.16^{**}$ |
| | 3. Goertzel (1994) | [−], $r = -.37^{*}$ |
| Trust in government | 1. Bruder and Kunert (2022) | [−], $b = −0.48^{***}$ |
| | 2. Mari et al. (2022) | [−], $r = -.22^{***}$ |
| | 3. Imhoff and Lamberty (2018) | [−], $r = -.53^{**}$, $r = -.28^{**}$ |
| Trust in political parties | 1. Krouwel et al. (2017) | [−], $r = -.27^{**}$ |
| | 2. Leiser et al. (2017) | [−], $r = .46^{*}$ |
| Trust in scientists | 1. Bruder and Kunert (2022) | [−], $b = −0.42^{*}$ |
| | 2. Rothmund et al. (2022) | [−], $r = -.45^{**}$ |
| Extraversion | 1. Goreis and Voracek (2019) | [O], $r = .01$ [meta-analysis] |
| | 2. Swami and Frunham (2012) | [O], $r = .03$ |
| | 3. Imhoff and Lamberty (2018) | [O], r = -.012, r = -.006 |
| Openness | 1. Goreis and Voracek (2019) | [O], $r = .02$ [meta-analysis] |
| | 2. Imhoff and Lamberty (2018) | [O], $r = .01$, $r = -.04$ |
| | 3. Swami et al. (2010) | [+], $r = .23^{***}$ |
| Agreeableness | 1. Goreis and Voracek (2019) | [O], $r = -.02$ [meta-analysis] |
| | 2. Brotherton et al. (2013) | [O], $r = .11$ |
| | 3. Swami et al. (2011) | [−], $r = -.07^{*}$ |
| Neuroticism | 1. Goreis and Voracek (2019) | [O], $r = .03$ [meta-analysis] |
| | 2. Imhoff and Lamberty (2018) | [+\|O], $r = .13^{*}$, $r = .14$ [accounting for paranoia] |
| | 3. Hollander (2018) | [+], $r = .03^{*}$, $r = .08^{***}$, $r = .09^{***}$, $r = .08^{***}$ |
| Conscientiousness | 1. Goreis and Voracek (2019) | [O], $r = .01$ [meta-analysis] |
| | 2. Imhoff and Lamberty (2018) | [O], $r = -.05$, $r = -.02$ |
| | 3. Swami and Furnham (2012) | [+], $r = .12^{***}$ |
| Analytical thinking | 1. Swami et al. (2014) | [−], $r = -.25^{***}$ |
| | 2. Swami and Furnham (2012) | [−], $r = -.17^{***}$ |
| | 3. Stahl and Van Prooijen (2018) | [O], $b = −0.07$ [accounting for rationality] |

(Continues)

**TABLE 1** (Continued)

| Variable | Author | Direction and strength of relationship with CT belief |
|---|---|---|
| Isolation/Ostracism | 1. Poon et al. (2020) | [+], $b = .35^{***}$ [ostracism] |
| | 2. Graeupner and Coman (2017) | [+], $r = .19^{*}$ [social exclusion] |
| | 3. Moulding et al. (2016) | [+], $r = .34^{**}$ [sense of isolation] |
| Need for uniqueness | 1. Imhoff and Lamberty (2017) | [+], $r = .128^{*}$, $r = .203^{**}$, $r = .139^{*}$ |
| | 2. Kay (2021) | [+], $r = .11^{*}$ |
| | 3. Lantian et al. (2017) | [+], $r = .17^{*}$ [Study 2] |
| Narcissism | 1. Cichocka et al. (2016) | [+], $r = .24^{***}$ [collective narcissism] |
| | 2. Sternisko et al. (2021) | [+], $r = .30^{*}$, $.48^{*}$ [collective narcissism] |
| | 3. Kay (2021) | [+], $r = .23^{*}$ [Narcissistic Personality Inventory] |

*Note*. Exemplary studies limited to three for each variable (if available). Indication of direction: [+] significant positive relationship found, [–] significant negative relationship found, [O] no significant relationship found. Coefficients displayed as found in original study, rounded to two decimal places.
All listed variables are also available in the present study's dataset (except from Need for Uniqueness and Narcissism). More information on measurements used in the present study can be found in Table 2.
$^{*}p < .05$.
$^{**}p < 0.01$.
$^{***}p < 0.001$.

which clear theoretical assumptions do not yet exist, prior findings fail to generalize, or when predictor variables and their relationships with a criterion are manifold (e.g., Walsh et al., 2017). Since this is arguably the case for CT research, ML models represent a promising option to alleviate inconsistent results and extract how psychological factors are associated with CT belief. In this study, I will use the popular Random Forest (RF) Machine Learning algorithm (Breiman, 2001) to predict belief in CTs based on a variety of previously investigated psychological factors, which all have been linked to CT belief in the past. Further, I will render support to the interpretability of the RF results and compare the performance to linear models on the same dataset by using a predictive modelling approach.

## 2 | METHODS

### 2.1 | Dataset overview

The dataset used in the present work is part of a panel study conducted by the COVID-19 Psychological Research Consortium (McBride et al., 2021), in which the authors collected data on several topics over multiple representative waves in the UK, including CT belief. Representativeness of the study sample was assured by defining strata for age, sex, and household income for the UK prior to sampling. Target quotas for each stratum were obtained from Eurostat population estimates (Eurostat, 2020) and the Office for National Statistics (2017). The final sample quotas for age, sex and household incomes were met within 1% deviation of the representative UK target quotas. The dataset contains $N = 2025$ adults (1047 female, $M_{age} = 45.4$ years, $SD = 15.9$), all of whom were residents of the UK. The complete dataset, information about the panel study, quotas, as well as design and conduct can be found in McBride et al. (2021) and the corresponding OSF repository.

### 2.2 | Measures and data preprocessing

Belief in CTs was measured using the Conspiracy Mentality Questionnaire (CMQ) by Bruder et al. (2013), a validated scale to assess a person's general tendency (i.e., mindset) to believe in CTs. This measurement was chosen due to its beneficial features of being independent of events and societal contexts while still exhibiting high correlations with belief in more specific CTs (Bruder et al., 2013; Dyrendal et al., 2021). The dataset further contains all previously investigated psychological factors listed in Table 1 (except from Narcissism and Need for Uniqueness), which were used as *features* (i.e., predictors) to predict belief in CTs. The inclusion of predictors was based on the literature review described in the Introduction but limited to data availability. Caveats to the generalizability of the findings due to missing factors in the current dataset will be addressed in the Discussion.

Due to the big number of predictors investigated in the present study, all information regarding their measurement (including operationalization and scale reliability scores) is mainly presented in a tabular form (see Table 2). In what follows, the characteristics of the used measurements will be presented in an abbreviated form. Structural factors (e.g., age, gender) as well as political ideology, self-esteem, and trust in different entities (i.e., people, government, parties, scientists) were assessed with single item questions. The Big Five personality factors, control beliefs, paranoid ideation and uncertainty were assessed using subscales of existing instruments. More specifically, facets of the Big Five personality factors (i.e., Extraversion, Openness, Agreeableness, Neuroticism, Conscientiousness) were assessed using the respective two item subscales of the Short Big Five Inventory by Rammstedt and John (2007). Similarly, the different types of control beliefs (i.e., internal, randomness and socio-political, see Table 1) were measured using the three-item control belief subscales *internal*, *chance*

**TABLE 2** Overview of variables, operationalization and scale reliability coefficients in the present study

| Variable | Description, source and example item | Number of items, response format | Reliability (Cronbach's $\alpha$) |
|---|---|---|---|
| Outcome | | | |
| Conspiracy mentality | Mean CMQ Score (Conspiracy Mentality Scale, Bruder et al., 2013) "I think that politicians usually do not tell us the true motives for their decisions" | 5 items, 11-point scale (Certainly not - Certainly) | .85 |
| Features (Predictors) | | | |
| Age | Age of participant in years | 1 item, open text | – |
| Gender | Gender of participant | 1 item, categorical (male; female; other: open field) | – |
| Living area | Area of living | 1 item, categorical (city; suburb; town; rural area) | – |
| Education | Highest qualification | 1 item, categorical (no qualifications; O-level/GCSE or similar; A-level or similar; diploma; undergraduate degree; postgraduate degree; technical qualification; other) | – |
| Income | Approximate annual household income | 1 item, categorical (1: £0–15,490; 2: £15,491–£25,340; 3: £25,341–£38,740; 4: £38,741–£57,930; 5: > = £57,931) | – |
| Religion | Religion of participant | 1 item, categorical (Christian; Muslim; Jewish; Buddhist; Sikh; Atheist; Agnostic; or other) | – |
| Ideology | Self-reported rating on 10-point political ideology scale | 1 item, 10-point scale (left - right) | – |
| Right-wing authoritarianism (RWA) | Total scale mean score: Right Wing Authoritarianism (Short Authoritarianism Scale, Bizumic & Duckitt, 2018) "What our country needs most is discipline, with everyone following our leaders in unity" | 6 items, 5-point scale (Strongly disagree – Strongly agree) | .73 |
| Social dominance orientation (SDO) | Total scale mean score: Social Dominance Orientation (Social Dominance Scale; Ho et al., 2015) "An ideal society requires some groups to be on top and others to be on the bottom" | 8 items, 5-points scale (Strongly oppose – Strongly favour) | .84 |
| Self-esteem | Single-Item Self-esteem Scale (SISES) (Robins, Hendin, & Trzesniewski, 2001) "I have high self-esteem" | 1 item, 7-point scale (Not very true of me – Very true of me) | – |
| Paranoid ideation | Mean score of subscale "Persecution", Persecution and Deservedness Scale (Persecution and Deservedness Scale; Melo, Corcoran, Shryane & Bentall, 2009) "I'm often suspicious of other people's intentions towards me" | 5 items, 5-point scale (Strongly disagree – Strongly agree) | .86 |
| Resilience | Total scale mean score: Resilience (Brief Resilience Scale; Smith et al., 2008) "I tend to take a long time to get over setbacks in my life" | 6 items, 5-point scale (Strongly disagree – Strongly agree) | .88 |
| Death anxiety | Total scale mean score: Death Anxiety (Death Anxiety Inventory; Tomás-Sábado, Gómez-Benito, & Limonero, 2005) "I find it difficult to accept the idea that it all finishes with death" | 17 items, 5-point scale (Totally disagree – Totally agree) | .94 |

(Continues)

**TABLE 2** (Continued)

| Variable | Description, source and example item | Number of items, response format | Reliability (Cronbach's α) |
|---|---|---|---|
| Control: Internal | Mean score of subscale "Internal", Locus of control (Locus of Control Scale; Sapp & Harrod, 1993) "My life is determined by my own actions" | 3 items, 7-point scale (Strongly disagree – Strongly agree) | .71 |
| Control: Chance [Randomness] | Mean score of subscale "Chance", Locus of control (Locus of Control Scale; Sapp & Harrod, 1993) "To a great extent, my life is controlled by accidental happenings" | 3 items, 7-point scale (Strongly disagree – Strongly agree) | .70 |
| Control: Powerful others [Socio-political] | Mean score of subscale "Powerful others", Locus of control (Locus of Control Scale; Sapp & Harrod, 1993) "Getting what I want requires pleasing those people above me" | 3 items, 7-point scale (Strongly disagree – Strongly agree) | .85 |
| Uncertainty-intolerance | Mean score of subscale "Uncertainty is stressful and upsetting", Intolerance of uncertainty (Intolerance of Uncertainty Scale, Buhr & Dugas, 2002) "I always want to know what the future has in store for me" | 12 items, 5-point scale (Not characteristic of me at all – Entirely characteristic of me) | .90 |
| Interpersonal trust | Trust in people ("Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people? ") | 1 item, 5-point scale (Mostly people can be trusted – Need to be very careful) | – |
| Trust in government | Trust in government ("Could you indicate how much trust you have in the government") | 1 item, 5-point scale (Completely trust - Do not trust at all) | – |
| Trust in political parties | Trust in political parties ("Could you indicate how much trust you have in political parties") | 1 item, 5-point scale (Completely trust – Do not trust at all) | – |
| Trust in scientists | Trust in scientists ("Could you indicate how much trust you have in scientists") | 1 item, 5-point scale (Completely trust – Do not trust at all) | – |
| Extraversion | Mean score of subscale "Extraversion", Big Five personality factors (Short Big Five Inventory; Rammstedt & John, 2007) "I see myself as someone who is reserved" | 2 items, 5-point scale (Disagree strongly – Agree strongly) | .57 (Spearman- Brown) |
| Openness | Mean score of subscale "Openness", Big Five personality factors (Short Big Five Inventory; Rammstedt & John, 2007) "I see myself as someone who has an active imagination" | 2 items, 5-point scale (Disagree strongly – Agree strongly) | .13 (Spearman- Brown) |
| Agreeableness | Mean score of subscale "Agreeableness", Big Five personality factors (Short Big Five Inventory; Rammstedt & John, 2007) "I see myself as someone who is generally trusting" | 2 items, 5-point scale (Disagree strongly – Agree strongly) | .27 (Spearman- Brown) |
| Neuroticism | Mean score of subscale "Neuroticism", Big Five personality factors (Short Big Five Inventory; Rammstedt & John, 2007) "I see myself as someone who get nervous easily" | 2 items, 5-point scale (Disagree strongly – Agree strongly) | .65 (Spearman- Brown) |
| Conscientiousness | Mean score of subscale "Conscientiousness", Big Five personality factors (Short Big Five Inventory; Rammstedt & John, 2007) "I see myself as someone who does a thorough job" | 2 items, 5-point scale (Disagree strongly – Agree strongly) | .51 (Spearman-Brown) |

(Continues)

**TABLE 2** (Continued)

| Variable | Description, source and example item | Number of items, response format | Reliability (Cronbach's α) |
|---|---|---|---|
| Analytical thinking | Total sum of right answers: Cognitive Reflection Task (Frederick, 2005) "If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?" | 5 items, multiple choice with 4 answer options | – |
| Loneliness [Isolation/Ostracism] | Total scale mean score: Loneliness (Three item Loneliness Scale; Hughes, Waite, Hawkley, & Cacioppo, 2004) "I lack companionship" | 3 items, 3-point scale (Hardly ever, Some of the time, Often) | .88 |

*Note*. Variable names describe investigated construct, brackets indicate construct description in previous studies (cf. Table 1). Reliability scores only calculated for (sub)scales with at least two continuous items. Further information regarding variable measurement in the present study can also be retrieved from the codebook of the original dataset (McBride et al., 2021).

and *powerful others* of the Locus of Control scale by Sapp and Harrod (1993). Paranoid ideation was assessed using the five-item *Persecution* subscale of the Persecution and Deservedness Scale by Melo et al. (2009). The 12-item subscale *Uncertainty is stressful and upsetting* of the Intolerance of Uncertainty scale by Buhr and Dugas (2002) was deployed to measure uncertainty. Analytical thinking was measured using the total sum of right answers to the Cognitive Reflection Task by Frederick (2005) with five questions. All other psychological factors in the present study (RWA, SDO, Resilience, Death-Anxiety, and Loneliness) were measured using the entirety of existing, validated scales for each respective construct. In other words, all item responses in these scales were averaged (including subscales), resulting in total scale means.

Prior to model building, the dataset was preprocessed. At first, missing values on the features were imputed using the mean across participants for each feature (< 10% of total values). The responses of six participants were excluded from the analysis (each representing a unique level in the gender variable) since model prediction on unseen data can fail, if a predictor level only occurs in one data subset. All categorical features were *one-hot-encoded* (i.e., dummy-coded). The dependent variable and numerical features were standardized.

## 2.3 | Model selection

Representing the current approach, I modelled a standard ordinary least-squares regression model (OLS). Following the predictive approach, I also deployed a regularized linear Ridge Regression (RR), which penalizes the size of the parameter estimates (contingent on the number of features in the model) and shrinks them towards zero, leading to a more biased model compared to OLS (Hoerl & Kennard, 1970). This regularization technique is also known as "L2 regularization" and was chosen over stepwise procedures, as either retaining or dropping variables exhibits high variance and is often ineffective (e.g., Hastie et al., 2009, pp. 93ff.). The last deployed ML model, a Random Forest Regressor (RF), was chosen, since it can cope with smaller datasets and many predictors, and can account for complex interactions and non-linear relationships. RFs can be best described as types of non-linear models, based on the idea of *decision trees*. These trees learn input–output relationships by iteratively splitting the so-called "predictor space" in distinct, non-overlapping regions. Due to its non-linear approximation, a decision tree can obtain a perfect prediction on the training data if grown deep enough, which can lead to overfitting. Tackling this problem, RFs comprise an ensemble of single decision trees. For every decision tree in the RF, a random subsample of predictors as well as a random bootstrap subsample of participants is drawn. Consequently, the single trees are decorrelated and the overall RF is regularized. In the end, the outputs of the single trees are aggregated to retrieve one final prediction for each sample in the RF. For more information, see Breiman (2001).

## 2.4 | Analytical procedure

All analysis steps were performed in a *Jupyter Notebook* (Kluyver et al., 2016) using the Python programming language version 3.8, the *numpy* (Van Der Walt et al., 2011) and *pandas* (McKinney, 2011) library for data management as well as the *scikit-learn* library (Pedregosa et al., 2011) for model building and data preprocessing. The full analysis code can be found in the corresponding OSF project (see Appendix). For all models, the available features indicated in Table 1 were entered as independent variables and the conspiracy mentality (CMQ) as the dependent variable.

To assess and compare model performance (i.e., capability to generalize to unseen data) of the RF, RR and OLS, cross-validation (CV) was used. Note that although a designated held-out test set is often more desirable to test model generalizability, CV yields a robust estimate of prediction performance and is even preferable when working with small- to medium-sized datasets (cf. Raschka, 2018). Therefore, CV represents a great option for CT belief research, since researchers do not need to extensively collect data and can work with smaller datasets at their disposal while still being able to estimate generalization performance of modelled associations between psychological factors and CT belief, as well as compare competing models. CV works by splitting the entire dataset into *k* groups (folds), each unique group once functioning as the test fold (out-of-sample) and the remaining *k*-1 groups functioning as the training data (in-sample) in *k* rounds. The final generalization performance is obtained by averaging the prediction error over the *k* test folds.

However, this procedure is sensitive to the composition of the *k* training folds and the partitioning scheme (Rodriguez et al., 2009). In other words, the number of folds chosen and specific train fold characteristics (e.g., extreme cases or outliers) can lead to a noisy estimate of the model's capability to generalize to new data, especially when dealing with small samples. Consequently, it has been suggested to repeat the CV procedure to account for small perturbations in the single folds, thereby stabilizing the estimation of a given model's capability to generalize to new data (e.g., Krstajic et al., 2014; Rodriguez et al., 2009). Thus, the present study conducted a 10 × 5 repeated CV. Five folds have been selected, as this number generally results in models with neither excessively high bias nor high variance (other things being equal), leading to realistic replicability estimates (James et al., 2013, p. 184). Consequently, each of the training folds contained $N = 1616$ observations and each of the test folds contained $N = 403$ observations.

In "flexible" models like the RR and RF, the researcher must fine tune the so-called *hyperparameters*, which control the amount of regularization in the model. Fine tuning these hyperparameters was achieved by automatically finding the best settings using a *GridSearch* algorithm (Pedregosa et al., 2011). Hereby, the researcher must only define a potential grid of potential hyperparameters, and the algorithm automatically selects the best combination. This procedure was built into the CV. Before training and validating the model on the five test sets (outer-loop), each training set was split into five additional folds (inner-loop) and the best hyperparameters were selected in this inner loop, before completing the outer loop. In other words, the tuning procedure was *nested* into the cross-validation procedure, carried out by a search algorithm. In doing so, the risk of obtaining overfitted, high variance models is further reduced, yielding a more realistic picture of out-of-sample prediction performance, since finding the best hyperparameters and evaluating model performance in the outer loop are not mixed.

After retrieving the optimal hyperparameters in the inner-loop, the RF as well as both linear models were iteratively trained on the five training folds and tested on the corresponding five test folds (outer-loop) to get an average prediction performance. Note that the inner loop was not applied to the OLS model, as there are no hyperparameters to be tuned. The whole nested CV procedure was repeated ten times to get the final generalization performance, as previously described. Studies have shown that this (repeated) nested cross-validation is an essential method for reliable model assessment (He & Chalise, 2020; Krstajic et al., 2014; Varma & Simon, 2006). Figure S4 in the Appendix presents an illustrative overview of the repeated nested cross-validation.

## 2.5 | Performance validation

To compare the linear models and the RF regarding their generalization performance to unseen data, the root mean squared error (RMSE) as well as the explained variance $R^2$ were calculated. Further, the permutation feature importance (PFI) as well as the Accumulated Local Effects plots (ALE) (Altmann et al., 2010; Apley & Zhu, 2020) were calculated for the RF. These metrics enable deep insights into variable relationships in a ML model (i.e., RF) and will be briefly explained.

The permutation feature importance measures how much a particular predictor contributes to the overall performance in the model. It is obtained by iteratively shuffling the rows of each feature, thereby removing its information (i.e., signal). In each step, the model is run again with one permuted feature, the new prediction performance is obtained and compared to the un-permuted, original model. This procedure is repeatedly applied (here: ten times) for each feature at a time to get an average estimate of the importance. In short, the values we obtain for each predictor express the average decrease explained variance $R^2$ when the information of this feature is removed from the model. Note that this importance measure is relative and that the values should not be interpreted in an absolute manner, similar to standard parameter estimates of linear models. Also, note that the relative importance of a feature also takes interaction effects with other features in the model into account, meaning that the interaction effect between two features is present in both importance measures (i.e., the features main and interaction effects). Therefore, the PFI values do not add up to the total amount of explained variance, but a larger sum (Molnar, 2022). For more information on permutation-based feature importance, see Altmann et al. (2010) or Molnar (2022).

ALE plots go a step further than permutation importance. First, we choose a feature for which we want to know how it affects CT belief prediction on average. Now, for each possible value *v* of our feature, we calculate the change in prediction by looking at a small "window" of values around *v* and calculating the difference in prediction between the upper and lower window which embeds our value of interest. To illustrate this method, let us say we want to predict the price of a house based on several features, like living area in m$^2$, neighbourhood, year of building, etc. We now want to know, how each value *v* of the feature "living area" influences our price prediction.

> For the effect of living area at 30 m$^2$, the ALE method uses all houses with about 30 m$^2$, gets the model predictions pretending these houses were 31 m$^2$ minus the prediction pretending they were 29 m$^2$. This gives us the pure effect of the living area and is not mixing the effect with the effects of correlated features. (Molnar, 2022)

These calculated differences are then accumulated and centred to obtain the ALE curves. Consequently, ALE values can be interpreted as a simple main effect of a feature at different values *v* in comparison to the average effect of the predictor values (centred at zero). Put simply, the ALE curves show us how our features are associated with the outcome prediction. Note, however, that ALE values are not the same as, for instance, linear regression coefficients and cannot be compared one to one, since they are calculated differently. For more information, see Molnar (2022).

## 3 | RESULTS

Prior to model building and evaluation, Pearson correlations between all continuous variables were calculated (Table 3). The strongest associations with CT belief were obtained with interpersonal trust ($r = .25$, $p < .001$), trust in government ($r = .21$, $p < .001$), locus of control: powerful others ($r = .24$, $p < .001$) and paranoid ideation ($r = .19$, $p < .001$). On the other hand, psychological factors like self-esteem ($r = -.05$, $p > .99$), SDO ($r = -.07$, $p > .99$) or the Big Five personality traits ($ps > .99$), except Neuroticism ($r = .10$, $p = .01$), were not substantially correlated with CT belief. In general, these observed correlations of psychological factors with CT belief are mostly in line with the results pattern of previous research (cf. Table 1). Some of the predictors were intercorrelated to a mentionable degree, for example resilience and neuroticism ($r = -.69$, $p < .001$), the subjective control dimensions "chance" and "powerful others" ($r = .69$, $p < .001$), and trust in government and trust in parties ($r = .62$, $p < .001$). Expectedly, most of the medium to medium-high correlations were found in the context of different subscales of the same or theoretically linked constructs. Note that these potential multicollinearities only pose a problem to the interpretability of linear regression model coefficients but not to their overall prediction performance on unseen data per se (Harrell Jr, 2015, p. 105). Following Jacobucci and Grimm (2020) suggestions when comparing linear and ML models in psychological research, reliability of predictors and outcome were inspected using Cronbach's alpha. All scales obtained high to sufficient scores ($\alpha > .70$) with one exception: the two-item subscales of the short Big Five scale yielded reliability scores below .70 (tested with the Spearman-Brown Coefficient), although the authors of the short Big Five scale Rammstedt and John (2007) reported sufficient reliability in their validation study. An overview over the reliability scores can be found in Table 2.

### 3.1 | Cross-validation results

Figure 1 shows the prediction performance of the linear regression (OLS), ridge regression (RR) and Random Forest (RF) obtained in the cross-validation procedure. Notably, this analysis estimates how well and consistently (on average) our models predict CT belief on unseen data (test folds), which gives us an estimate for their generalization performance, if we were to apply them on further, unseen data. Along with the mean prediction performance, the standard error across the ten repeated CVs is reported. This metric can be used as an estimate of the uncertainty around the generalization estimate (i.e., prediction variance) to compare model stability (Krstajic et al., 2014). Therefore, both the average prediction performance estimate and the standard error can be used to make an informed decision about which model (i.e., OLS, RR, RF) is best suited to robustly capture the associations between personal factors and CT belief.

Looking at the prediction error, the OLS performed the worst out of all three models, $RMSE_{OLS} = 0.921$ ($SE = 0.58^{-3}$). In contrast, regularizing a linear model helped to improve the average prediction error as well as model stability, as shown by the results of the RR, $RMSE_{RR} = 0.917$ ($SE = 0.45^{-3}$). In comparison to both linear models, the RF delivered the lowest prediction error overall, $RMSE_{RF} = 0.911$ ($SE = 0.50^{-3}$), only exhibiting a slightly higher standard error than the RR across the CV runs. Looking at the explained variance $R^2$, the results paint a similar picture. The ranking of all three models remained the same with the OLS showing the lowest average $R^2$ value and highest standard error, $R^2{}_{OLS} = 14.87\%$ ($SE = 0.11\%$). Again, the RF performed the best $R^2{}_{RF} = 16.80\%$ ($SE = 0.08$) and the RR explained more variance than the OLS, $R^2{}_{RR} = 15.54\%$ ($SE = 0.08$). Figure 1 graphically illustrates these differences in model performance, averaged over all ten CV repetitions. Overall, the results indicate that the OLS performs the worst when predicting CT belief on out-of-sample data. In comparison, the RR showed higher performance overall, indicating that regularization of a linear model helped to predict CT belief more robustly and accurately. This pattern illustrates that even when dealing with a medium-sized dataset like the present study, low-capacity linear models can capture a non-negligible amount of noise in the data, yielding models with higher variance. Furthermore, associations of personal factors and CT belief might not be limited to singular, linear relationships, since a non-linear, flexible RF showed the best generalization performance to unseen data with comparatively stable performance across different test sets. Further, the mean prediction error and explained variance in the RF did not fall within the respective 1 SE range of the RR or OLS (see Figure 1),

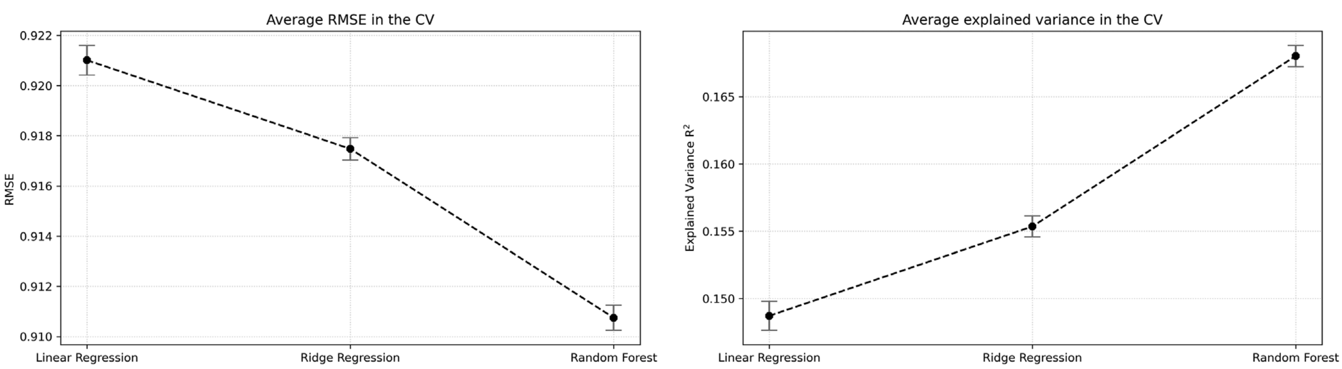



**FIGURE 1** Graphical illustration of model cross-validation performance (out-of-sample). *Note*. Black dots represent average RSME and explained variance $R^2$, respectively. Error bars indicate $\pm$ 1 SE from the mean over all ten test folds in the repeated cross-validation

**TABLE 3** Descriptive statistics and correlations of scales and numerical variables used in the study

| Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Self-esteem | 4.09 | 1.58 | – | | | | | | | | | | | | | | | | | | | | | | |
| 2. Ideology | 5.32 | 1.86 | 0.22*** | – | | | | | | | | | | | | | | | | | | | | | |
| 3. Paranoid ideation | 2.49 | 1.00 | −0.22*** | 0.01 | – | | | | | | | | | | | | | | | | | | | | |
| 4. SDO | 2.40 | 0.69 | 0.15*** | 0.44*** | 0.19*** | – | | | | | | | | | | | | | | | | | | | |
| 5. RWA | 3.03 | 0.67 | 0.06 | 0.38*** | 0.04 | 0.33*** | – | | | | | | | | | | | | | | | | | | |
| 6. Loneliness | 1.59 | 0.62 | −0.36*** | −0.10** | 0.45*** | −0.02 | −0.07 | – | | | | | | | | | | | | | | | | | |
| 7. Control: Internal | 4.89 | 1.07 | 0.33*** | 0.13*** | −0.18*** | −0.04 | 0.05 | −0.18*** | – | | | | | | | | | | | | | | | | |
| 8. Control: Powerful others | 3.37 | 1.41 | −0.16*** | −0.03 | 0.58*** | 0.10** | 0.01 | 0.35*** | −0.20*** | – | | | | | | | | | | | | | | | |
| 9. Control: Chance | 3.82 | 1.23 | −0.22*** | −0.05 | 0.49*** | 0.06 | −0.02 | 0.36*** | −0.12*** | 0.69*** | – | | | | | | | | | | | | | | |
| 10. Extraversion | 2.91 | 0.96 | 0.44*** | 0.07 | −0.20*** | 0.03 | 0.04 | −0.24*** | 0.18*** | −0.13*** | −0.16*** | – | | | | | | | | | | | | | |
| 11. Agreeableness | 3.37 | 0.80 | 0.15*** | −0.04 | −0.38*** | −0.19*** | 0.01 | −0.23*** | 0.15*** | −0.21*** | −0.19*** | 0.18*** | – | | | | | | | | | | | | |
| 12. Conscien-tiousness | 3.72 | 0.87 | 0.20*** | 0.09* | −0.26*** | −0.07 | 0.18*** | −0.19*** | 0.22*** | −0.24*** | −0.24*** | 0.17*** | 0.23*** | – | | | | | | | | | | | |
| 13. Neuroticism | 2.85 | 1.05 | −0.51*** | −0.11*** | 0.40*** | −0.04 | −0.03 | 0.42*** | −0.30*** | 0.34*** | 0.34*** | −0.37*** | −0.25*** | −0.23*** | – | | | | | | | | | | |
| 14. Openness | 3.25 | 0.83 | 0.03 | −0.07 | −0.02 | −0.15*** | −0.11*** | 0.07 | 0.02 | −0.02 | 0.02 | 0.07 | 0.05 | 0.10*** | 0.07 | – | | | | | | | | | |
| 15. Death anxiety | 2.58 | 0.88 | −0.11*** | 0.07 | 0.46*** | 0.19*** | 0.07 | 0.28*** | −0.12*** | 0.42*** | 0.37*** | −0.07 | −0.17*** | −0.25*** | 0.39*** | .00 | – | | | | | | | | |
| 16. Resilience | 3.27 | 0.84 | 0.50*** | 0.13*** | −0.43*** | 0.03 | 0.07 | −0.44*** | 0.36*** | −0.38*** | −0.39*** | 0.35*** | 0.26*** | 0.30*** | −0.69*** | .00 | −0.39*** | – | | | | | | | |
| 17. Uncertainty | 2.93 | 0.76 | −0.25*** | −0.03 | 0.50*** | 0.03 | 0.01 | 0.38*** | −0.14*** | 0.45*** | 0.42*** | −0.26*** | −0.26*** | −0.2*** | 0.52*** | 0.04 | 0.54*** | −0.51*** | – | | | | | | |
| 18. Interpersonal trust | 3.29 | 1.02 | −0.10* | 0.15*** | 0.38*** | 0.15*** | 0.22*** | 0.16*** | −0.10** | 0.21*** | 0.20*** | −0.09* | −0.28*** | −0.05 | 0.15*** | −0.06 | 0.22*** | −0.16*** | 0.21*** | – | | | | | |
| 19. Trust: government | 3.41 | 1.10 | −0.20*** | −0.32*** | 0.09* | −0.24*** | −0.25*** | 0.09* | −0.12*** | 0.09** | 0.10*** | −0.09* | −0.09* | −0.06 | 0.13*** | 0.07 | .00 | −0.14*** | 0.06 | 0.12*** | – | | | | |
| 20. Trust: parties | 3.73 | 0.96 | −0.19*** | −0.13*** | 0.01 | −0.13*** | −0.10** | 0.04 | −0.08 | .00 | .00 | −0.08 | −0.07 | 0.02 | 0.07 | 0.04 | −0.08 | −0.06 | −0.02 | 0.12*** | 0.62*** | – | | | |
| 21. Trust: scientists | 2.35 | 0.99 | −0.05 | 0.06 | 0.24*** | 0.20*** | 0.08 | 0.07 | −0.22*** | 0.20*** | 0.11*** | −0.02 | −0.17*** | −0.14*** | 0.08 | −0.04 | 0.16*** | −0.11*** | 0.07 | 0.21*** | 0.28*** | 0.23*** | – | | |
| 22. Analytical thinking | 1.90 | 1.61 | .00 | −0.08 | −0.19*** | −0.10** | −0.19*** | −0.07 | 0.08 | −0.15*** | −0.10** | −0.07 | −0.05 | 0.01 | −0.11*** | 0.01 | −0.21*** | 0.09* | −0.07 | −0.17*** | 0.06 | 0.11*** | −0.15*** | – | |
| 23. Conspiracy belief (CMQ) | 7.03 | 1.83 | −0.05 | .00 | 0.19*** | −0.07 | 0.10** | 0.11*** | .00 | 0.24*** | 0.16*** | 0.03 | −0.03 | 0.05 | 0.09* | 0.05 | 0.10** | −0.06 | 0.15*** | 0.25*** | 0.21*** | 0.16*** | 0.09* | −0.13 | – |

*Note.* $N = 2019$, correlation coefficients rounded to two decimal places.
*$p < .05$.
**$p < .01$.
***$p < .001$ (Bonferroni-adjusted alpha for 253 bivariate correlations).

providing a further argument for its superiority to linear models. In the next step, the importance of personal factors for CT belief prediction in the RF, as well as their associations, will be analysed.

### 3.2 | Feature relationship and importance

Before calculating the permutation feature importance, the RF was fitted on the complete dataset, again using a Grid Search algorithm (Pedregosa et al., 2011) to find the best hyperparameters. Looking at the importance plot (Figure 2), the most relevant feature for prediction performance in the RF turned out to be the belief in powerful others controlling people's lives (socio-political control). When the information of this feature was removed, the average explained variance decreased by more than 10%. Decrease in explained variance of the second most important feature, trust in the government, is almost on a par with belief in powerful others. The third highest scoring feature, interpersonal trust, turned out to be less important for the overall model performance, although it still exhibits an 8% average decrease in variance, when information on this feature is removed. All following features are trailing in comparison to the three mentioned ones with less than 6% average decrease in explained variance. Nevertheless, the RF attributed incremental importance to them, showing that most of the features still contain relevant information for CT belief prediction. Furthermore, the estimation error of feature importance throughout permutations seems to be relatively low for most features, as seen in the lower and upper boxplot bounds.

To investigate the association of the top three features with belief in CTs, ALE plots were calculated (Figure 3), which reveal linear as well as non-linear trends. Starting with the most relevant feature, belief in powerful others, a value of + 1 standard deviation from the sample mean increased predicted CT belief by 0.10 compared to the prediction when using the average of the feature. In other words, the function of belief in powerful others and belief in CTs shows a positive slope (i.e., coefficient) of ∼ .10. Note however, that the function is not perfectly linear, as we can see the slope to ramp up significantly for people reporting values between ½ SD to 1 SD from the mean. For all other features, the change in slopes seems to be even more pronounced. We can observe a steeper, almost exponential slope of the functions for values above the feature means than for values below the feature means. More precisely, a 1 SD increase in distrust in information from government or trust in people leads to a higher amount of CT belief than a 1 SD decrease lowers the amount of CT belief. In summary, the plots illustrate non-linear associations between psychological factors and belief in CTs.

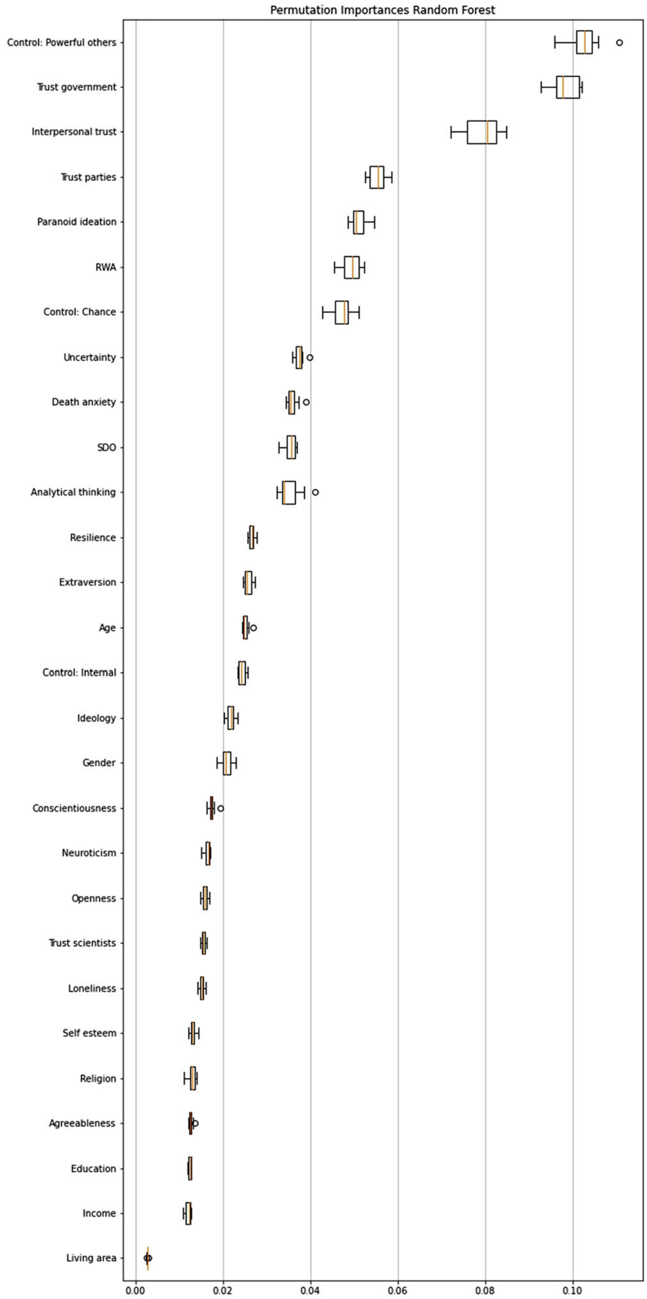



**FIGURE 2** Permutation feature importance (PFI) of RF model features. *Note*. Permutation feature importance calculated using ten permutations per feature on the full dataset. Upper and lower box-plot bounds indicate ± 1.5 IQR, outliers represented as circles

## 4 | DISCUSSION

Belief in CTs and its determinants has gained much scientific interest in the past decade. Although numerous studies have investigated the relationship of psychological factors with belief in CTs, recent studies have put many of the proposed associations into question (e.g., Goreis & Voracek, 2019). The present study argues that methodological problems like high sample-specificity, a focus on binary linear relationships and the neglection of predictor interactions may have contributed to these inconsistencies in the literature. As a partial solution to these problems, a predictive analysis approach (cross-validation and regularization) and machine learning algorithms were proposed and applied to a recently collected, openly available large dataset (McBride et al., 2021).

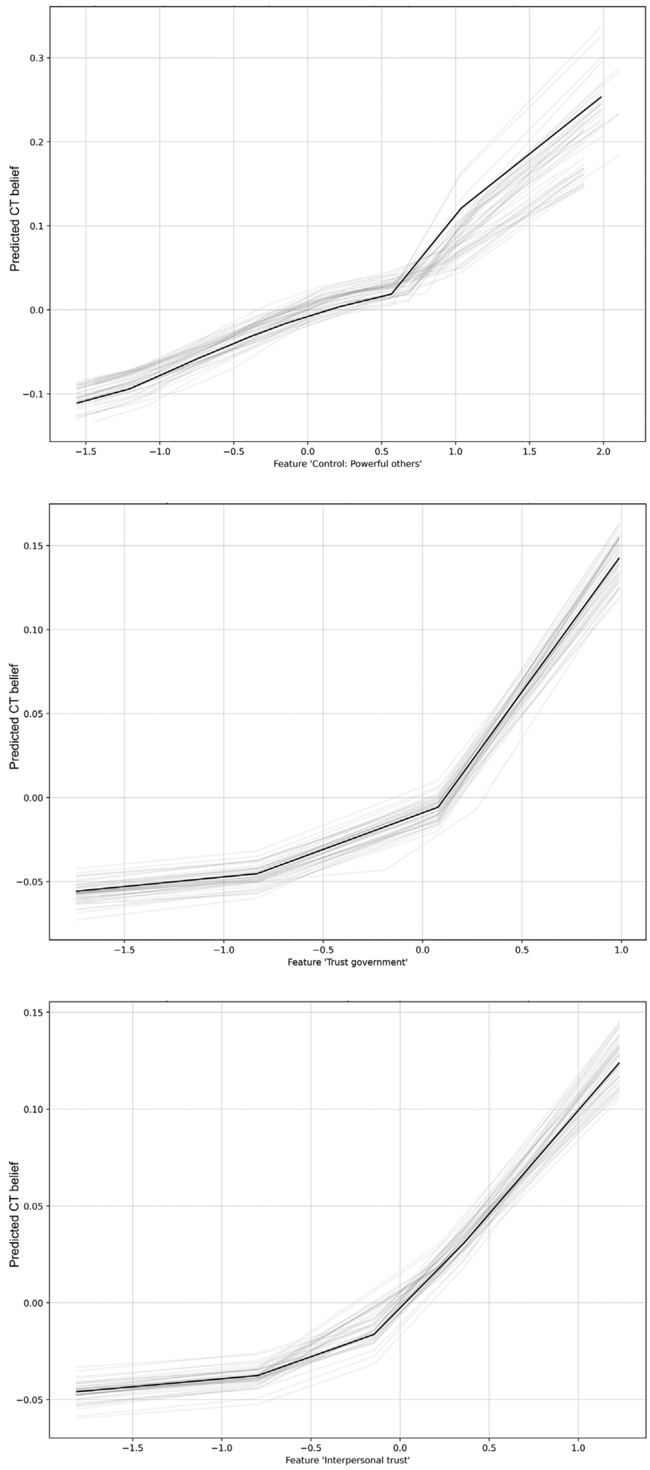



**FIGURE 3** Accumulated-local-effects plots (ALE) of the three most important features in the RF. *Note*. Features are negatively coded, higher values indicate more distrust/less control. Values for predicted CT belief and features are standardized. Grey lines indicate single Monte Carlo simulations for the dataset, solid black line represents the average accumulated local effect (see Molnar, 2022)

## 4.1 | Methodological insights for studies on conspiracy theory belief

As expected, the present results show that CT belief prediction performance of a standard linear model (OLS) is low and model variance seems to be an issue, as seen in direct performance comparison to a regularized linear regression. Already from a mere methodological view, it is thus not surprising that associations between psychological factors and CT belief are inconsistent across different studies, since researchers who follow the current analysis approach may end up with vastly divergent coefficients on small datasets. Reducing model variance by using a regularized linear regression (ridge regression, RR) helped to alleviate these issues and improve performance. Therefore, future research should incorporate cross-validation and model comparison techniques when defaulting to linear models before interpreting potential associations of personal factors and CT belief, since they might be inflated. Furthermore, the chosen non-linear Random Forest outperformed both linear models in the cross-validation, yielding more accurate CT belief predictions. Most importantly, this surplus in performance and generalization capability supports the existence of more complex interactions and non-linear relationships between predictors of CT belief than many previous study designs accounted for. This notion is further supported by the accumulated local effects (ALE) plots indicating non-linear relationships, which draw on similar findings in this research area (e.g., Van Prooijen et al., 2015). Adding to this point, the permutation feature importance revealed multiple psychological factors to be relevant for CT belief prediction, supporting the existence of interplaying effects as noted in Götz et al. (2021).

Comparing the prediction performance in the present study ($R^2_{RF}$ CV ~17%) with previous research is not an easy task, since previous studies neither reported generalization performance on (big) unseen datasets nor conducted any kind of cross-validation techniques (to the author's knowledge). Furthermore, several previous studies did not employ linear prediction models of CT belief or report effect sizes, rendering comparisons between previous findings and the present ones even more difficult. Studies that did, report effect sizes ranging from $R^2 = 10\%$ (Cichocka et al., 2016), up to $R^2 = 40\%$ (Galliford & Furnham, 2017), putting the results of the present study somewhere in the middle of previously reported effect sizes. Nevertheless, belief in CTs could be predicted to a mentionable degree, which arguably represents a more reasonable estimate, considering the limitations of other studies mentioned before. In sum, the present results advocate the use of extensive analysis strategies in CT research, such as cross-validation, model comparison, inclusion of multiple predictors and their interactions, as well as accounting for non-linear relationships.

## 4.2 | Theoretical and practical implications of the current research

Apart from conceptual findings, the present study also delivers concrete theoretical insights. Looking at the relative importance of psychological factors for predicting CT belief, striking similarities to previous research can be observed. The most important predictor in the Random Forest turned out to be socio-political control (locus of control—powerful others). This result is in line with previous research on socio-political control, since the belief in powerful others controlling one's environment seems to be one of the most robust and strong correlates of CT belief (see Table 1). Furthermore, socio-political control was a strong predictor of CT belief, regardless of other covariates in the model, which largely rules out other potential explanations for their association raised in the past, like spurious relationships (cf. Stojanov et al., 2020). Theoretical considerations also speak for the strong association of socio-political control and CT belief, since a perceived lack of control is directly tied to social motives and (for instance) the tendency to detect agency of powerful actors in random events (Van Prooijen, 2019). Although the causal direction of this relationship is still debated (cf. Jolley & Douglas, 2014; Stojanov et al., 2020), belief in socio-political control seems to be deeply interwoven with a conspiratorial mindset.

The same applies to dimensions of distrust in official institutions and people in general. Especially for levels of distrust above average, the probability of CT belief seems to increase almost exponentially (see Figure 3). Again, this finding is supported by previous literature, showing comparatively robust relationships between trust and CT belief. Since official institutions like the government or parties are generally seen as high-power groups (e.g., Imhoff & Bruder, 2014) potentially being orchestrators of societal events, this relationship between distrust and conspiracy mentality becomes apparent. Similarly, interpersonal trust was often associated with heightened CT endorsement (e.g., Imhoff & Lamberty, 2018). Van Prooijen et al. (2021) state multiple potential reasons for this linkage: if those in power are unwilling to protect the ordinary citizen against malevolent intentions of others, people may be less likely to risk being exploited by others leading to more distrust. People high in conspiracy mentality also showed reduced investments in a behavioural trust board game, independent of interpersonal trust cues. Thus, high levels of distrust in official institutions and society seem to go hand in hand with conspiracy mentality.

Although still being important for CT belief prediction, all other predictors are trailing in comparison to the aforementioned ones. This might be due to several reasons. First, most of the remaining predictors showed much less clear and strong associations in previous literature, such as self-esteem, the Big Five and internal locus of control (see Table 1). Thus, their low importance might arise from their inherent low association with CT belief in the first place. However, other predictors showed comparatively high associations with CT belief in the past. This finding might be explained by shared variance with other constructs in the model or the predictors solely being moderators of (and/or being moderated by) other constructs, as already shown for ideology (Miller et al., 2016), analytical thinking style (Ståhl & Van Prooijen, 2018) or neuroticism (Imhoff & Lamberty, 2018). Consequently, these constructs might still be associated with CT belief but, when accounting for other predictors, their share in explained variance shrinks. Recent studies on feature importance in ML models indicate that the importance of highly correlated predictors (in comparison to uncorrelated predictors) might be slightly overestimated in the permutation, which could have influenced their relative position in the estimation (Wei et al., 2015).[1] Nevertheless, the relative order of predictors falls in line with the general result patterns of previous research. For instance, CT belief associations with paranoid ideation, authoritarianism and uncertainty seem to be much stronger and robust than with, for example, self-esteem or the Big Five, which is mostly in line with their relative rank in the feature importance plot (see Figure 2). In sum, researchers may especially focus on psychological factors like control beliefs, trust and generalized attitudes to understand why some people are more susceptible to CT beliefs than others.

On top of these theoretical insights, practical implications for dealing with CT beliefs arise. Building bridges between citizens and (local) governments may not only be helpful for a stable society but also for potential CT belief reduction through increased trust. Although others did not find manipulated individual control to be associated with reduced CT belief (Stojanov et al., 2020), empowering individuals' perception of socio-political control may still be a promising approach. As suggested by previous research in this area, both increasing trust and perceived socio-political control may best be tackled from regulatory angles, that is by increasing involvement of individuals in political decision-making processes (Parent et al., 2005). Studies in this area have shown that increasing government responsiveness (Kahne & Westheimer, 2006), in-person and online interaction with citizens (Arshad & Khurram, 2020) or e-government initiatives (Parent et al., 2005) can be fruitful strategies to deal with political distrust and efficacy, which may then lead to lower CT belief susceptibility.

### 4.3 | Limitations of the current research

Naturally, the present study has its limitations. First, some potentially important variables which have been shown to be related to CT beliefs in previous research, like for example need for uniqueness (Imhoff & Lamberty, 2017), ostracism (Poon et al., 2020) or narcissism (Cichocka et al., 2016), were not available in the dataset analysed here (see Table 1). Although exhibiting a fair amount of explained variance, the prediction performance of the models in the present study could potentially be improved by including these missing variables. However, even though they might account for some unexplained variance or share variance with other constructs (e.g., loneliness, self-esteem) the present results and previous studies still point towards key components of a conspiratorial mindset like trust, socio-political control or generalized attitudes.

[1] Note that this only affects predictors that are at least to some extent related to the outcome. The feature importance in the RF was also calculated using Shapeley Values (Lundberg et al., 2018), which are not influenced by the presence of correlated features. The results revealed a similar picture to the permutation importance, with trust in government, interpersonal control, and socio-political control being the most important predictors. The summary plot can be found in Figure S5 in the supporting information or in the linked OSF project.

Second, the measurement of constructs must be addressed. Although the used locus of control scale by Sapp and Harrod (1993) shows good internal consistency and construct validity, it has not been as widely used in CT research as other measures. This leads to comparability limitations for the association of CT belief, especially for the control dimension *chance*. Adding to that, the control subscale *powerful others* (socio-political control) and *chance* were highly correlated ($r$ = .69, Table 3), which could also explain the relatively high feature importance of control subscale chance in the Random Forest, as mentioned earlier. Further, there seems to be an overlap between item formulations of socio-political control and the conspiracy mentality questionnaire CMQ (e.g., "My life is chiefly controlled by powerful others" vs. "I think that there are secret organizations that greatly influence political decisions"). Although this semantic overlap could cast doubt in their construct independency, both constructs were only correlated to a small degree ($r$ = .24) and have multiple distinct correlates (Table 3). In addition, robust associations between similar measurements of socio-political control and CT belief across studies further support the significance of their relationship (see Table 1).

Even though the used cross-validation procedure has been shown to deliver robust generalization estimation performance, the final models fitted on the whole dataset may not perform as precisely on a newly collected, designated test set (cf. Bates et al., 2021). Since the Random Forest also exhibited a non-negligible amount of variance in the cross-validation, it is unclear if it would consistently outperform linear models in other CT belief conceptualizations. Note, however, that Krstajic et al. (2014) argued that the use of standard errors to assess model performance and variance needs to be redefined in the repeated cross-validation context and thus may not be as informative in the present study. Speaking of CT measurement, the current study used the CMQ (Bruder et al., 2013) to assess people's general conspiracy mentality and not belief in specific CTs. This measurement of CT belief was chosen, since specific CTs limit the generalizability of findings to specific socio-political events that are heavily influenced by cultural context and situational variables and do not reflect the underlying mentality or belief readiness (Stojanov et al., 2020). Therefore, numerous studies have used the CMQ to measure CT belief in the past and have also shown general conspiracy mentality to be a direct predictor of belief in more specific CTs (e.g., Bruder et al., 2013; Dyrendal et al., 2021; Swami et al., 2010). These features combined made the CMQ a well-suited option for the present research question. Nevertheless, the results of the present study may not be directly translated to belief in other, more specific CTs. While being comparably large and representative, the dataset exclusively consisted of UK residents and the transferability of results to other cultures and countries is pending. Finally, since analysing concrete predictor interactions was beyond the scope of the present study, future research on CT belief should investigate these relationships more thoroughly. In summary, the proposed approach should be extended to other datasets, operationalizing CT belief in different ways, and should include a variety of potential predictors and analyse their interactions.

## 5 | GENERAL CONCLUSION

Investigating CTs and why people are drawn to them is more than just an end in itself. However, we must be aware of potential pitfalls in our research approach. The present study showed that cross-validation and replication efforts are warranted in CT belief research. Arguably, this emerging research area can accumulate knowledge at a higher rate if correlational results are robust and generalizable, especially when inductively building theories on these findings. Otherwise, we end up in the current situation of contradictory and inconsistent results, ultimately leading to confusion and hindering progress. Moreover, ML models can be a fruitful alternative to standard linear models. Using them, we gain insights about complex associations and interactions of psychological factors with CT belief that go beyond the simplicity of current findings. Adopting all these practices, we can focus our research efforts and built on a solid foundation to ultimately progress in establishing a comprehensive CT belief framework.

### AUTHOR CONTRIBUTIONS

The author Nils Brandenstein conducted all preparation steps of the study as well as the present manuscript on his own. There are no further authorships to be mentioned.

### ACKNOWLEDGEMENTS

I would like to express my greatest appreciation to Prof. Dr Jan Rummel for his kind and enduring advice on the manuscript.

### CONFLICT OF INTEREST

This manuscript has not been published (except from a PrePrint on PsychArchive) and is not under consideration for publication elsewhere. There are no conflicts of interest and funding to disclose.

### ETHICS STATEMENT

I declare that the present research is conducted ethically, results are reported honestly, the submitted work is original and not (self-)plagiarized, and authorship reflects individuals' contributions.

### DATA AVAILABILITY STATEMENT

The data and analysis code used in the manuscript can be found under the following OSF repository: https://osf.io/ckjyp/?view_only=bcbf25c0f2c241ef808bf84f48a8186f. The data reported in this manuscript were obtained from publicly available data as part of a panel study conducted by the COVID-19 Psychological Research Consortium (McBride et al., 2021) and is openly accessible on the OSF.

### ORCID

*Nils Brandenstein* https://orcid.org/0000-0003-2663-8793

### REFERENCES

Abalakina-Paap, M., Stephan, W. G., Craig, T., & Gregory, W. L. (1999). Beliefs in conspiracies. *Political Psychology*, *20*(3), 637–647. https://doi.org/10.1111/0162-895x.00160

Adam-Troian, J., Chayinska, M., Paladino, M. P., Uluğ, Ö. M., Vaes, J., & Wagner-Egger, P. (2021). Of precarity and conspiracy: Introducing a socio-functional model of conspiracy beliefs. https://doi.org/10.31234/osf.io/mnfrd

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(4), 1059–1086. https://doi.org/10.1111/rssb.12377

Arshad, S., & Khurram, S. (2020). Can government's presence on social media stimulate citizens' online political participation? Investigating the influence of transparency, trust, and responsiveness. *Government Information Quarterly*, *37*(3), 101486. https://doi.org/10.1016/j.giq.2020.101486

Asbrock, F., Sibley, C. G., & Duckitt, J. (2010). Right-wing authoritarianism and social dominance orientation and the dimensions of generalized prejudice: A longitudinal test. *European Journal of Personality*, *24*(4), 324–340. https://doi.org/10.1002/per.746

Bakker, M., Van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543–554. https://doi.org/10.1177/1745691612459060

Bates, S., Hastie, T., & Tibshirani, R. (2021). Cross-validation: What does it estimate and how well does it do it? *arXiv*, preprint arXiv:2104.00673. https://doi.org/10.48550/arXiv.2104.00673

Bizumic, B., & Duckitt, J. (2018). Investigating right wing authoritarianism with a very short authoritarianism scale. *Journal of Social and Political Psychology*, *6*(1), 129–150. http://doi.org/10.5964/jspp.v6i1.835

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology*, *4*, 279–279. https://doi.org/10.3389/fpsyg.2013.00279

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, *4*, 225–225. https://doi.org/10.3389/fpsyg.2013.00225

Bruder, M., & Kunert, L. (2022). The conspiracy hoax? Testing key hypotheses about the correlates of generic beliefs in conspiracy theories during the COVID-19 pandemic. *International Journal of Psychology*, *57*(1), 43–48. http://doi.org/10.1002/ijop.12769

Buhr, K., & Dugas, M. J. (2002). The intolerance of uncertainty scale: Psychometric properties of the English version. *Behaviour Research and Therapy*, *40*(8), 931–945. https://doi.org/10.1016/s0005-7967(01)00092-4

Cichocka, A., Marchlewska, M., & De Zavala, A. G. (2016). Does self-love or self-hate predict conspiracy beliefs? Narcissism, self-esteem, and the endorsement of conspiracy theories. *Social Psychological and Personality Science*, *7*(2), 157–166. https://doi.org/10.1177/1948550615616170

Darwin, H., Neave, N., & Holmes, J. (2011). Belief in conspiracy theories. The role of paranormal belief, paranoid ideation and schizotypy. *Personality and Individual Differences*, *50*(8), 1289–1293. https://doi.org/10.1016/j.paid.2011.02.027

Dieguez, S., Wagner-Egger, P., & Gauvrit, N. (2015). Nothing happens by accident, or does it? A low prior for randomness does not explain belief in conspiracy theories. *Psychological Science*, *26*(11), 1762–1770. https://doi.org/10.1177/0956797615598740

Douglas, K. M., & Sutton, R. M. (2008). The hidden impact of conspiracy theories: Perceived and actual influence of theories surrounding the death of Princess Diana. *The Journal of Social Psychology*, *148*(2), 210–222. https://doi.org/10.3200/socp.148.2.210-222

Douglas, K. M., Sutton, R. M., Callan, M. J., Dawtry, R. J., & Harvey, A. J. (2016). Someone is pulling the strings: Hypersensitive agency detection and belief in conspiracy theories. *Thinking and Reasoning*, *22*(1), 57–77. https://doi.org/10.1080/13546783.2015.1051586

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current Directions in Psychological Science*, *26*(6), 538–542. https://doi.org/10.1177/0963721417718261

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology*, *40*, 3–35. https://doi.org/10.1111/pops.12568

Dyrendal, A., Kennair, L. E. O., & Bendixen, M. (2021). Predictors of belief in conspiracy theory: The role of individual differences in schizotypal traits, paranormal beliefs, social dominance orientation, right wing authoritarianism and conspiracy mentality. *Personality and Individual Differences*, *173*, 110645. https://doi.org/10.1016/j.paid.2021.110645

Einstein, K. L., & Glick, D. M. (2015). Do I think BLS data are BS? The consequences of conspiracy theories. *Political Behavior*, *37*(3), 679–701. https://doi.org/10.1007/s11109-014-9287-z

Eurostat (2020). *Population on 1 January by age and sex*. Eurostat. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1037/e640112011-028

Freeman, D., Waite, F., Rosebrock, L., Petit, A., Causier, C., East, A., Jenner, L., Teale, A.-L., Carr, L., & Mulhall, S. (2020). Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychological Medicine*, *52*(2), 1–13. https://doi.org/10.1017/s0033291720001890

Galliford, N., & Furnham, A. (2017). Individual difference factors and beliefs in medical and political conspiracy theories. *Scandinavian Journal of Psychology*, *58*(5), 422–428. https://doi.org/10.1111/sjop.12382

Golec de Zavala, A., & Federico, C. M. (2018). Collective narcissism and the growth of conspiracy thinking over the course of the 2016 United States presidential election: A longitudinal analysis. *European Journal of Social Psychology*, *48*(7), 1011–1018. https://doi.org/10.1002/ejsp.2496

Goertzel, T. (1994). Belief in Conspiracy Theories. *Political Psychology*, *15*(4), 731–742. http://doi.org/10.2307/3791630

Goreis, A., & Voracek, M. (2019). A systematic review and meta-analysis of psychological research on conspiracy beliefs: Field characteristics, measurement instruments, and associations with personality traits. *Frontiers in Psychology*, *10*, 205–205. https://doi.org/10.3389/fpsyg.2019.00205

Götz, F., Gosling, S., & Rentfrow, J. (2021). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, *17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Graeupner, D., & Coman, A. (2017). The dark side of meaning-making: How social exclusion leads to superstitious thinking. *Journal of Experimental Social Psychology*, *69*, 218–222. https://doi.org/10.1016/j.jesp.2016.10.003

Green, R., & Douglas, K. M. (2018). Anxious attachment and belief in conspiracy theories. *Personality and Individual Differences*, *125*, 30–37. http://doi.org/10.1016/j.paid.2017.12.023

Grzesiak-Feldman, M., & Ejsmont, A. (2008). Paranoia and conspiracy thinking of Jews, Arabs, Germans, and Russians in a Polish sample. *Psychological Reports*, *102*(3), 884–886. https://doi.org/10.2466/pr0.102.3.884-886

Grzesiak-Feldman, M., & Irzycka, M. (2009). Right-wing authoritarianism and conspiracy thinking in a Polish sample. *Psychological Reports*, *105*(2), 389–393. https://doi.org/10.2466/pr0.105.2.389-393

Harrell, E. Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-19425-7

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (Vol. 1, 2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

He, J., & Chalise, P. (2020). Nested and repeated cross validation for classification model with high-dimensional data. *Revista Colombiana de Estadística*, *43*(1), 103–125. https://doi.org/10.15446/rce.v43n1.80000

Hindman, M. (2015). Building better models: Prediction, replication, and machine learning in the social sciences. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 48–62. https://doi.org/10.1177/0002716215570279

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new $SDO_7$ scale. *Journal of Personality and Social Psychology*, *109*(6), 1003–1028. http://doi.org/10.1037/pspi0000033

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.2000.10485983

Hollander, B. A. (2018). Partisanship, individual differences, and news media exposure as predictors of conspiracy beliefs. *Journalism & Mass Communication Quarterly*, *95*(3), 691–713. https://doi.org/10.1177/1077699017728919

Hughes, M. -E., Waite, L. J., Hawkley, L. C., & Cacioppo, J. T. (2004). A Short Scale for Measuring Loneliness in Large Surveys. *Research on Aging*, *26*(6), 655–672. http://doi.org/10.1177/0164027504268574

Imhoff, R. (2015). Beyond (right-wing) authoritarianism: Conspiracy mentality as an incremental predictor of prejudice. In M. Bilewicz, A. Cichocka & W. Soral, *The psychology of conspiracy* (pp. 1, 140–160). London: Routledge. https://doi.org/10.4324/9781315746838

Imhoff, R., & Bruder, M. (2014). Speaking (un–) truth to power: Conspiracy mentality as a generalised political attitude. *European Journal of Personality*, *28*(1), 25–43. https://doi.org/10.1002/per.1930

Imhoff, R., & Lamberty, P. (2018). How paranoid are conspiracy believers? Toward a more fine-grained understanding of the connect and disconnect between paranoia and belief in conspiracy theories. *European Journal of Social Psychology*, *48*(7), 909–926. https://doi.org/10.1002/ejsp.2494

Imhoff, R., & Lamberty, P. (2020). A bioweapon or a hoax? The link between distinct conspiracy beliefs about the Coronavirus disease (COVID-19) outbreak and pandemic behavior. *Social Psychological and Personality Science*, *11*(8), 1110–1118. https://doi.org/10.31234/osf.io/ye3ma

Imhoff, R., & Lamberty, P. K. (2017). Too special to be duped: Need for uniqueness motivates conspiracy beliefs. *European Journal of Social Psychology*, *47*(6), 724–734. https://doi.org/10.1002/ejsp.2265

Imhoff, R., Zimmer, F., Klein, O., António, J. H. C., Babinska, M., Bangerter, A., Bilewicz, M., Blanuša, N., Bovan, K., Bužarovska, R., Cichocka, A., Delouvée, S., Douglas, K. M., Dyrendal, A., Etienne, T., Gjoneska, B., Graf, S., Gualda, E., Hirschberger, G., … van Prooijen, J.-W. (2022). Conspiracy mentality and political orientation across 26 countries. *Nature Human Behaviour*, *6*(3), 392–403. http://doi.org/10.1038/s41562-021-01258-7

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. https://doi.org/10.1177/1745691620902467

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, 1st ed.). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Jasinskaja-Lahti, I., & Jetten, J. (2019). Unpacking the relationship between religiosity and conspiracy beliefs in Australia. *British Journal of Social Psychology*, *58*(4), 938–954. https://doi.org/10.1111/bjso.12314

Jensen, E. A., Pfleger, A., Herbig, L., Wagoner, B., Lorenz, L., & Watzlawik, M. (2021). What drives belief in vaccination conspiracy theories in Germany? *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.678335

Jolley, D., & Douglas, K. M. (2014). The effects of anti-vaccine conspiracy theories on vaccination intentions. *Plos One*, *9*(2), e89177. https://doi.org/10.1371/journal.pone.0089177

Kahne, J., & Westheimer, J. (2006). The limits of political efficacy: Educating citizens for a democratic society. *PS: Political Science & Politics*, *39*(2), 289–296. https://doi.org/10.1017/s1049096506060471

Kay, C. S. (2021). The targets of all treachery: Delusional ideation, paranoia, and the need for uniqueness as mediators between two forms of narcissism and conspiracy beliefs. *Journal of Research in Personality*, *93*, 104128. https://doi.org/10.1016/j.jrp.2021.104128

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C.F. Loizides & B. Schmidt(2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. *Proceedings of the 20th International Conference on Electronic Publishing* (87–90). Amsterdam: IOS Press Ebooks. https://www.doi.org/10.3233/978-1-61499-649-1-87

Kofta, M., Soral, W., & Bilewicz, M. (2020). What breeds conspiracy antisemitism? The role of political uncontrollability and uncertainty in the belief in Jewish conspiracy. *Journal of Personality and Social Psychology*, *118*(5), 900–918. https://doi.org/10.1037/pspa0000183

Krouwel, A., Kutiyski, Y., van Prooijen, J.-W., Martinsson, J., & Markstedt, E. (2017). Does extreme political ideology predict conspiracy beliefs, economic evaluations and political trust? Evidence from Sweden. *Journal of Social and Political Psychology*, *5*(2), 435–462. http://doi.org/10.5964/jspp.v5i2.745

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, *6*(1), 10–10. https://doi.org/10.1186/1758-2946-6-10

Lantian, A., Muller, D., Nurra, C., & Douglas, K. M. (2017). I know things they don't know!. *Social Psychology*, *48*(3), 160–173. https://doi.org/10.1027/1864-9335/a000306

Leman, P. J., & Cinnirella, M. (2013). Beliefs in conspiracy theories and the need for cognitive closure. *Frontiers in Psychology*, *4*, http://doi.org/10.3389/fpsyg.2013.00378

Leiser, D., Duani, N., & Wagner-Egger, P. (2017). The conspiratorial style in lay economic thinking. *PLOS ONE*, *12*(3), e0171238. http://doi.org/10.1371/journal.pone.0171238

Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The role of conspiracist ideation and worldviews in predicting rejection of science. *Plos One*, *8*(10), e75637. https://doi.org/10.1371/journal.pone.0075637

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science*, *24*(5), 622–633. https://doi.org/10.1177/0956797612457686

Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv*, preprint arXiv:1802.03888.

Maftei, A., & Holman, A.-C. (2022). Beliefs in conspiracy theories, intolerance of uncertainty, and moral disengagement during the coronavirus crisis. *Ethics & Behavior*, *32*(1), 1–11. https://doi.org/10.1080/10508422.2020.1843171

Mari, S., Gil de Zúñiga, H., Suerdem, A., Hanke, K., Brown, G., Vilar, R., Boer, D., & Bilewicz, M. (2022). Conspiracy theories and institutional trust: Examining the role of uncertainty avoidance and active social media use. *Political Psychology*, *43*(2), 277–296. https://doi.org/10.1111/pops.12754

Marques, M. D., Ling, M., Williams, M. N., Kerr, J. R., & McLennan, J. (2021). Australasian public awareness and belief in conspiracy theories: Motivational correlates. *Political Psychology*, *43*(1), 177–198. https://doi.org/10.1111/pops.12746

McBride, O., Murphy, J., Shevlin, M., Gibson-Miller, J., Hartman, T. K., Hyland, P., Levita, L., Mason, L., Martinez, A. P., McKay, R., Stocks, T. V., Bennett, K. M., Vallières, F., Karatzias, T., Valiente, C., Vazquez, C., & Bentall, R. P. (2021). Monitoring the psychological, social, and economic impact of the COVID-19 pandemic in the population: Context, design and conduct of the longitudinal COVID-19 psychological research consortium (C19PRC) study. *International Journal of Methods in Psychiatric Research*, *30*(1), e1861. https://doi.org/10.1002/mpr.1861

McKinney, W. (2011). Pandas: A foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, *14*(9), 1–9.

Melo, S., Corcoran, R., Shryane, N., & Bentall, R. P. (2009). The persecution and deservedness scale. *Psychology and Psychotherapy: Theory, Research and Practice*, *82*(3), 247–260. https://doi.org/10.1037/t20074-000

Miller, J. M. (2020). Psychological, political, and situational factors combine to boost COVID-19 conspiracy theory beliefs. *Canadian Journal of Political Science/Revue canadienne de science politique*, *53*(2), 327–334. https://doi.org/10.1017/s000842392000058x

Miller, J. M., Saunders, K. L., & Farhart, C. E. (2016). Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust. *American Journal of Political Science*, *60*(4), 824–844. https://doi.org/10.1111/ajps.12234

Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). https://christophm.github.io/interpretable-ml-book

Moulding, R., Nix-Carnell, S., Schnabel, A., Nedeljkovic, M., Burnside, E. E., Lentini, A. F., & Mehzabin, N. (2016). Better the devil you know than a world you don't? Intolerance of uncertainty and worldview explanations for belief in conspiracy theories. *Personality and Individual Differences*, *98*, 345–354. https://doi.org/10.1016/j.paid.2016.04.060

Newheiser, A.-K., Farias, M., & Tausch, N. (2011). The functional nature of conspiracy beliefs: Examining the underpinnings of belief in the Da Vinci Code conspiracy. *Personality and Individual Differences*, *51*(8), 1007–1011. https://doi.org/10.1016/j.paid.2011.08.011

OfficeforNationalStatistics. (2017). *Household disposable income and inequality in the UK: Financial year ending 2016*. OfficeforNationalStatistics. https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2016

Oliver, J. E., & Wood, T. J. (2014). Conspiracy theories and the paranoid style(s) of mass opinion. *American Journal of Political Science*, *58*(4), 952–966. https://doi.org/10.1111/ajps.12084

OpenScienceCollaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 910–911. https://doi.org/10.1126/science.aac4716

Parent, M., Vandebeek, C. A., & Gemino, A. C. (2005). Building citizen trust through e-government. *Government Information Quarterly*, *22*(4), 720–736. https://doi.org/10.1016/j.giq.2005.10.001

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Poon, K.-T., Chen, Z., & Wong, W.-Y. (2020). Beliefs in conspiracy theories following ostracism. *Personality and Social Psychology Bulletin*, *46*(8), 1234–1246. https://doi.org/10.1177/0146167219898944

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv*, preprint arXiv:1811.12808. https://doi.org/10.48550/arXiv.1811.12808

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*(2), 151–161. http://doi.org/10.1177/0146167201272002

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2009). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575. https://doi.org/10.1109/tpami.2009.187

Rothmund, T., Farkhari, F., Ziemer, C.-T., & Azevedo, F. (2022). Psychological underpinnings of pandemic denial - patterns of disagreement with scientific experts in the German public during the COVID-19 pandemic. *Public Understanding of Science*, *31*(4), 437–457. http://doi.org/10.1177/09636625211068131

Sapp, S. G., & Harrod, W. J. (1993). Reliability and validity of a brief version of Levenson's locus of control scale. *Psychological Reports*, *72*(2), 539–550. https://doi.org/10.2466/pr0.1993.72.2.539

Siddiqui, N. (2020). Who do you believe? Political parties and conspiracy theories in Pakistan. *Party Politics*, *26*(2), 107–119. https://doi.org/10.1177/1354068817749777

Ståhl, T., & Van Prooijen, J.-W. (2018). Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. *Personality and Individual Differences*, *122*, 155–163. https://doi.org/10.1016/j.paid.2017.10.026

Stempel, C., Hargrove, T., & Stempel, G. H. III (2007). Media use, social structure, and belief in 9/11 conspiracy theories. *Journalism & Mass Communication Quarterly*, *84*(2), 353–372. https://doi.org/10.1177/107769900708400210

Sternisko, A., Cichocka, A., Cislak, A., & Van Bavel, J. J. (2021). National narcissism predicts the belief in and the dissemination of conspiracy theories during the COVID-19 pandemic: Evidence from 56 countries. *Personality and Social Psychology Bulletin*, *Advance online publication*, https://doi.org/10.1177/01461672211054947

Stoica, C. A., & Umbreş, R. (2021). Suspicious minds in times of crisis: Determinants of Romanians' beliefs in COVID-19 conspiracy theories. *European Societies*, *23*(1), 246–261. https://doi.org/10.1080/14616696.2020.1823450

Stojanov, A., Bering, J. M., & Halberstadt, J. (2020). Does perceived lack of control lead to conspiracy theory beliefs? Findings from an online MTurk sample. *Plos One*, *15*(8), e0237771. https://doi.org/10.1371/journal.pone.0237771

Stojanov, A., & Halberstadt, J. (2020). Does lack of control lead to conspiracy beliefs? A meta-analysis. *European Journal of Social Psychology*, *50*(5), 955–968. https://doi.org/10.1002/ejsp.2690

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, *15*(3), 194–200. http://doi.org/10.1080/10705500802222972

Swami, V. (2012). Social psychological origins of conspiracy theories: The case of the Jewish conspiracy theory in Malaysia. *Frontiers in Psychology*, *3*, 280. https://doi.org/10.3389/fpsyg.2012.00280

Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. *Applied Cognitive Psychology*, *24*(6), 749–761. https://doi.org/10.1002/acp.1583

Swami, V., Coles, R., Stieger, S., Pietschnig, J., Furnham, A., Rehim, S., & Voracek, M. (2011). Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories. *British Journal of Psychology*, *102*(3), 443–463. https://doi.org/10.1111/j.2044-8295.2010.02004.x

Swami, V., & Furnham, A. (2012). Examining conspiracist beliefs about the disappearance of Amelia Earhart. *The Journal of General Psychology*, *139*(4), 244–259. https://doi.org/10.1080/00221309.2012.697932

Swami, V., Pietschnig, J., Tran, U. S., Nader, I. W., Stieger, S., & Voracek, M. (2013). Lunar lies: The impact of informational framing and individual differences in shaping conspiracist beliefs about the moon landings. *Applied Cognitive Psychology*, *27*(1), 71–80. https://doi.org/10.1002/acp.2873

Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, *133*(3), 572–585. https://doi.org/10.1016/j.cognition.2014.08.006

Tomás-Sábado, J., Gómez-Benito, J., & Limonero, J. T. (2005). The Death Anxiety Inventory: A Revision. *Psychological Reports*, *97*(3), 793–796. http://doi.org/10.2466/pr0.97.3.793-796

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22–30.

Van Prooijen, J. W. (2017). Why education predicts decreased belief in conspiracy theories. *Applied Cognitive Psychology*, *31*(1), 50–58. https://doi.org/10.1002/acp.3301

Van Prooijen, J.-W. (2019). An existential threat model of conspiracy theories. *European Psychologist*, *25*(1), 16–25. https://doi.org/10.1027/1016-9040/a000381

Van Prooijen, J. W., & Acker, M. (2015). The influence of control on belief in conspiracy theories: Conceptual and applied extensions. *Applied Cognitive Psychology*, *29*(5), 753–761. https://doi.org/10.1002/acp.3161

Van Prooijen, J. W., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, *48*(7), 897–908. https://doi.org/10.1002/ejsp.2530

Van Prooijen, J. W., Douglas, K. M., & De Inocencio, C. (2018). Connecting the dots: Illusory pattern perception predicts belief in conspiracies and the supernatural. *European Journal of Social Psychology*, *48*(3), 320–335. https://doi.org/10.1002/ejsp.2331

Van Prooijen, J. W., & Jostmann, N. B. (2013). Belief in conspiracy theories: The influence of uncertainty and perceived morality. *European Journal of Social Psychology*, *43*(1), 109–115. https://doi.org/10.1002/ejsp.1922

Van Prooijen, J.-W., Krouwel, A. P., & Pollet, T. V. (2015). Political extremism predicts belief in conspiracy theories. *Social Psychological and Personality Science*, *6*(5), 570–578. https://doi.org/10.1177/1948550614567356

Van Prooijen, J.-W., Spadaro, G., & Wang, H. (2021). Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships. *Current Opinion in Psychology*, *43*, 65–69. https://doi.org/10.1016/j.copsyc.2021.06.013

Van Prooijen, J.-W., & Van Lange, P. A. (2014). *The social dimension of belief in conspiracy theories*. Cambridge University Press. https://doi.org/10.1017/cbo9781139565417

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics [Electronic Resource]*, *7*(1), 1–8. https://doi.org/10.1186/1471-2105-7-91

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*(3), 457–469. https://doi.org/10.1177/2167702617691560

Wei, P., Lu, Z., & Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety*, *142*, 399–432. https://doi.org/10.1016/j.ress.2015.05.018

Wood, S. K., & Bhatnagar, S. (2015). Resilience to the effects of social stress: Evidence from clinical and preclinical studies on the role of coping strategies. *Neurobiology of Stress*, *1*, 164–173. https://doi.org/10.1016/j.ynstr.2014.11.002

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yi, J. P., Smith, R. E., & Vitaliano, P. P. (2005). Stress-resilience, illness, and coping: A person-focused investigation of young women athletes. *Journal of Behavioral Medicine*, *28*(3), 257–265. https://doi.org/10.1007/s10865-005-4662-1

## SUPPORTING INFORMATION

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

**How to cite this article:** Brandenstein, N. (2022). Going beyond simplicity: Using machine learning to predict belief in conspiracy theories. *European Journal of Social Psychology*, *52*, 910–930. https://doi.org/10.1002/ejsp.2859

# Appendix B1 - Manuscript II

Manuscript II: The key determinants of individual greenhouse gas emissions in Germany are mostly domain-specific

Note:
This is the final version of the open-access article published in *Nature Communications Earth and Environment*. The reproduction of this article in this dissertation is permitted under the respective Creative Commons Attribution 4.0 International License.

Brandenstein, N., Ackermann, K., Aeschbach, N., & Rummel, J. (2023). The key determinants of individual greenhouse gas emissions in Germany are mostly domain-specific. *Communications Earth & Environment*, *4*(1), 422. https://doi.org/10.1038/s43247-023-01092-x

communications earth & environment

ARTICLE

Check for updates

https://doi.org/10.1038/s43247-023-01092-x OPEN

# The key determinants of individual greenhouse gas emissions in Germany are mostly domain-specific

Nils Brandenstein [1,2]✉, Kathrin Ackermann[2,3], Nicole Aeschbach [2,4,5] & Jan Rummel[1,2]

Individual behavior plays a pivotal role in mitigating climate change but our understanding of the multifaceted, determining factors of sustainable behavior remains incomplete. Here we conducted a comprehensive, cross-sectional survey of German households in 2021 ($N = 10,813$), assessing various potential determinants and measuring behavior in greenhouse gas emissions across various life domains (shelter, mobility, consumption, and diet). Machine learning models were employed to predict emissions from determining factors and benchmarked against commonly used linear models. Our findings indicate that machine learning models excel in capturing complex relationships between personal and situational factors, offering a more nuanced understanding of how determinants interplay and contribute to emissions. Notably, some factors like perceived behavioral control or habits consistently affected emissions, while others like infrastructural barriers and pro-environmental attitudes were domain-specific. These insights about key determinants of sustainable behavior are valuable for policymakers crafting effective climate change strategies at the individual level.

[1] Psychological Institute, Heidelberg University, Heidelberg, Germany. [2] Heidelberg Center for the Environment, Heidelberg University, Heidelberg, Germany. [3] Department of Social Sciences, University of Siegen, Adolf-Reichwein-Str. 2, Siegen 57076, Germany. [4] TdLab Geography, Heidelberg University, Heidelberg, Germany. [5] Heidelberg School of Education, Heidelberg, Germany. ✉email: nils.brandenstein@psychologie.uni-heidelberg.de

Human caused climate change and the concerning new climate reality continues to threaten the way of living of the current and following generations. The development of effective mitigation strategies thus remains a key challenge in the 21st Century[1] which requires targeting various aspects of everyday life[2,3]. Mitigation strategies are often categorized as supply-side and demand-side solutions. While the more traditional supply-side solutions focus on technological advancements to, for instance, decarbonize supply chains and improve energy efficiency of appliances, demand-side solutions directly address individual behavior and needs, such as promoting reduced car usage or energy consumption[4,5]. Supply-side solutions are traditionally considered as crucial for reaching global climate goals; but there is also a growing acknowledgment of the importance of demand-side solutions in tackling climate change from various angles. This shift mirrors the understanding that relying solely on potential technological advancements on the supply-side is unlikely to effectively accomplish global climate and sustainability objectives in due time[5]. However, since demand-side solutions aim to address individual behavior, their effectiveness in the real world heavily relies on individuals' willingness and capacity to adopt these measures[3]. To devise effective regulations and policies in this area, it is thus crucial to gain a comprehensive understanding of the factors that drive or hinder individual sustainable behavior in the first place. Consequently, researchers have dedicated immense effort to identifying these factors in the past. To date, studies have recognized a wide range of individual and situational factors that are relevant to individual behavior in various life domains such as the role of social norms, values and structural barriers for electricity consumption or mobility behavior[3,6]. While identifying singular factors can advance our general knowledge about existing drivers of and barriers to sustainable behavior, the intricacies of their relationships and their absolute and relative strength of influence on sustainable behavior is still widely unknown. This issue has also been raised in the most recent IPCC report: Although a lot of evidence linking singular factors to individual sustainable behavior exists, there still is a high need for investigating and understanding the mutual interactions of factors and their relative importance for individual behavioral change[3].

In the present study we aim to close this knowledge gap by (1) considering a multitude of potential influencing factors of sustainable behavior and their interaction in an interdisciplinary setting, (2) measuring individual sustainable behavior in total greenhouse gas (GHG) emissions for different life domains, and (3) using machine learning (ML) models to analyze the data. Taking this specific methodological approach is crucial for several reasons. First, most of the potentially influential factors have been obtained in study designs investigating only a singular or handful of factors at a time, even though researchers have repeatedly argued that individual sustainable behavior is shaped by a multitude of factors and their interaction with each other[7–13]. Studying factors in isolation and thus not being able to control for the influence of other factors can lead to severe over- or underestimation of their relative importance for sustainable behavior[14]. Consequently, if policy makers base demand-side solutions on factors that do not (or only marginally) affect individual behavior, it will not be surprising that these measures fail to reach the desired impact on climate change in the real world[15–17]. Relatedly, even when study designs considered multiple factors at the same time in previous research, scholars often used simple linear models to analyze the data, effectively neglecting complex (e.g., non-linear) associations among the factors. In doing so, any obtained results likely have led to biased conclusions in how particular factors influence behavior in the real world[14] which – again – can result in poor regulative decision making. Beyond studying factors in isolation and neglecting their associations, previous study designs often did not capture impactful behavior, that is, behavior high in GHG emissions[18,19]. In detail, studies often investigated specific types of sustainable behavior that can be assessed rather easily but may only exert relatively low GHG emissions, such as self-reported recycling or water-saving behaviors. Notably, the determining factors of such low GHG behaviors can differ drastically from those influencing high GHG behaviors[8,20], which may have led researchers and policymakers alike to focus on less important factors for climate change mitigation. On the other hand, studies that did measure high GHG behaviors (such as mobility and consumption patterns or energy use) often only measured parts of the respective behaviors (e.g., traveled distance by car but not accounting for car type or fuel consumption) which may also affect the obtained relative importance of factors. As a result, measuring sustainable behavior and related GHG emissions comprehensively and focusing on behaviors with high GHG impacts seems necessary to devise solutions effectively battling climate change[8,20,21].

Right now, only a few studies exist which at least satisfy some of the criteria outlined above to a certain degree. For instance, studies have analyzed the influence of multiple factors on mobility behavior[22], accounted for complex interplays of factors using ML models to identify high emission household[23], or measured behavior in GHG emissions more comprehensively[18]. However, to the best of our knowledge, no previous work has yet fully met all these study requirements in a single analysis which would be necessary to comprehensively investigate drivers and barriers of sustainable behavior and derive effective demand-side solutions.

To provide a comprehensive view on the factors relevant for the demand-side, we first compiled previously identified factors influencing sustainable behavior in an extensive literature review. We then asked a large sample representative of the general German population to report on all these factors as well as their GHG emissions in different life domains. In a single analysis of this data set we used ML models to predict individual behavior measured in GHG emissions with the candidate factors. We also compare the performance of our ML models to simple linear models which are still predominantly used in the literature. Finally, we analyze the relative importance of all factors in our models to identify which factors should best be targeted to initiate changes in high GHG behaviors. Our findings demonstrate that employing ML models for predicting sustainable behavior not only enhances the accuracy of predictions compared to simple linear models, but also facilitates the identification of important factors among numerous contenders. While certain factors investigated in our study are relevant for sustainable behavior across various life domains (e.g., perceived behavioral control, behavioral habits), others exhibit domain-specific relevance (e.g., infrastructural barriers, pro-environmental attitudes). Our results regarding the relative importance of factors for sustainable behavior are largely in line with those of previous studies but we also uncover notable disparities from previous results regarding the importance of some crucial factors (e.g., easy access to public transportation, income, or availability of sustainable food options). We assert that demand-side solutions aimed at mitigating climate change must recognize the intricate interplay of behavioral drivers. Based on our results, we offer recommendations regarding which factors one may want to focus on to maximize the beneficial effects of behavioral change for climate change mitigation.

## Results

**Study design**. To collect data for our study, we ran surveys in German households from April to July 2021. The recruiting and sampling procedure was conducted by the panel provider Respondi AG who ensures representative panels on national population statistics. Our sample was quoted by gender, age, and federal state. The final sample size used in the analyses after performing several quality checks was $N = 10{,}993$. More information about the sampling procedure and sample composition can be found in the Supplementary Methods and Supplementary Table 1.

In preparation for constructing our survey, we first identified the relevant factors of individual sustainable behavior in a comprehensive literature review and selected factors which are grounded in established theories on sustainable behavior as well as those which were identified as relevant in meta analyses and large scale studies. In general, the identified drivers and barriers of sustainable behavior can be categorized into internal (i.e., person-related) and external (i.e., situation-related) factors[10]. Internal factors mainly include psychological factors such as individual beliefs, attitudes, values, and intentions. External factors subsume political, social, economic, and cultural conditions people find themselves in. A detailed description of the literature review and all factors included in our analysis can be found in the Methods and Supplementary Notes 1 and 2.

To measure impactful sustainable behavior, we asked participants to report on their past behavior in the most important domains of everyday life: 1. shelter (electricity & heating), 2. mobility, 3. consumption and 4. diet. We then calculated people's GHG footprint in $CO_2$ emission equivalents to quantify an individual's contribution to climate change based on validated calculation principles for the German population[24–26]. The detailed calculation and life domain selection principles can be found in the Methods. We then used all the collected internal and external factors to predict domain-specific footprints. To account for the multitude of factors and their associations with each other, we used ML models to analyze the data. To this end, we chose popular models used in previous studies investigating GHG emissions and survey data. In detail, we used Random Forests (RF), support vector machines (SVM) and Lasso Regression (LASSO) and compared their performance to a traditional linear ordinary least squares regression model (LM) which still represents the current practice of predicting climate-relevant behavior. More information about the ML model selection, working principles and advantages over more traditional models can be found in the Methods and Supplementary Methods.

**Footprint prediction capacity of internal and external factors**. In the first step of our analysis, we aimed to test whether we can predict GHG emissions in the different domains of life through the internal and external factors included in our study (for an overview of all predictors used in the models, see Tables 1–7). To this end, we evaluated prediction performance of all models on out-of-sample data (i.e., data the models were not fitted on). That is, we cross-validated all model fits on a training dataset (CV) and assessed the final prediction performance on a separate test set. The mean absolute error (MAE) was used as the general prediction error metric and the explained variance ($R^2$) was used to quantify the variance in the GHG emissions that is accounted for by the internal and external factors. A more detailed description of the analysis procedure and rationale can be found in the Methods and the Supplementary Methods.

The results show that all ML models outperformed the standard LM. In detail, the RF exhibited the highest prediction performance of all models across domains (on average), followed by the SVM, LASSO and LM (see Fig. 1 and Supplementary Discussion for full model description and discussion). Not only did the ML models outperform the LM in most domains, but they also consistently exhibited low variance in prediction performance (Fig. 1, Supplementary Table 4 and Supplementary Discussion). In the best performing RF, predicting individual GHG emissions was most successful for the domains mobility and diet. In these domains, prediction error on the test set was the lowest (Mobility: $RF_{MAE} = 0.59$, Diet: $RF_{MAE} = 0.68$) and explained variance the highest across domains (Mobility: $RF_{R2} =$ 33%, Diet: $RF_{R2} = 24\%$). Relatedly, the biggest performance gains of the RF over the LM could also be observed in domains mobility and diet, the latter domain showing the highest improvement. In all domains, test set performance of the RF matched the performance of the CV on the training set (test set prediction values are within 1 SD intervals of the CV, Fig. 1).

These results illustrate three important points: First, low overall prediction performance and high variance in performance in the LMs indicates that using standard LMs could have led to overestimation of model performance in the past, potentially leading to misjudgments in the relative importance of factors for predicting individual behavior. Second, the ML models managed to capture the relationships between internal and external factors with individual GHG emissions more accurately. Third, performance gains of more complex (non-linear) models like RF and SVM over linear models (LM and LASSO) indicate that internal and external factors have complex relations with and interactive effects on sustainable behavior (Fig. 1), which empirically supports the assumption that individual sustainable behavior is influenced by various interacting factors[12].

**Key drivers and barriers of impactful sustainable behavior in domains mobility and diet**. Besides overall GHG emissions prediction performance, we were also interested in evaluating the relative importance of the factors included in our models. Since the RF performed the best among all models and prediction of GHG emissions worked best in domains mobility and diet, we focused on analyzing factor importance of the RF in these domains. Further, domains mobility and diet represent the first and second highest contributing domains to households' overall GHG emissions in many European countries[27,28]. Thus, it seems promising to develop tailored GHG reduction programs for them. Note that due to the high number of factors investigated in our analysis and our study goal of identifying important factors, we only focus on those factors considerably contributing to the model prediction in the following (see Methods). All important factors for predicting mobility and diet related emissions in our models are depicted in Figs. 2 and 3, sorted by their relative importance.

Within the mobility domain, two of the most important demographic factors predicting GHG emissions are income and professional status. Employed and wealthier people exhibited much higher mobility related GHG emissions than people who are retired and/or have lower incomes (Supplementary Figs. 4 and 5). This finding is consistent with previous studies investigating impactful behavior and the environmental impact of affluent citizens[5,18,20,22,29]. Further analysis of these relationships indicated that higher-income individuals are more likely to possess cars (most important predictor of GHG emissions in our model) and tend to travel by car more frequently (Supplementary Table 5). Again, this finding is in accordance with other previous findings[30,31]. The relationship of GHG emissions with professional status, on the other hand, seemed to be exclusively driven by retired individuals, since all other groups were attributed a comparable amount of GHG emissions in our model (Supplementary Fig. 5). Whereas previous research seems to be inconclusive regarding travel activity at retirement age[32–34], our

**Table 1 Overview of previously investigated factors (predictors) in the models to predict sustainable behavior.**

| INTERNAL FACTORS | | EXTERNAL FACTORS | |
|---|---|---|---|
| **Name of predictor (factor type)** | **Description [number and type of items, example item]** | **Name of predictor (factor type) [domain]** | **Description [number and type of items, example item]** |
| Behavioral Beliefs (psychological), TPB | Attitudes towards sustainable behavior [mean score of 2 items, numerical "From my point of view, saving electricity is unimportant"] | Evaluation of state financial subsidies for (energetic) renovation (political) [Shelter: Heating] | Satisfaction, transparency, accessibility of financial subsidies for energetical renovation state [mean score of 3 items, numerical "In 2019 the financial subsidies of the state for an (energetic) renovation and installation of more efficient heating sources were insufficient"] |
| Personal Norms (psychological), TPB | Opinions and habits of peers towards sustainable behavior [mean score of 2 items, numerical "Most of the people who are close to me keep their consumption behavior and spending (e.g., on clothing, everyday objects) as low as possible."] | Evaluation of financial subsidies of city/ municipality for renovation (political) [Shelter: Heating] | Satisfaction, transparency, accessibility of financial subsidies for energetical renovation muncipality [mean score of 3 items, numerical "In 2019 the financial subsidies of my city/ municipality for an (energetic) renovation and installation of more efficient heating sources were insufficient"] |
| Perceived Control (psychological), TPB | Perceived behavorial control over sustainable behavior [mean score of 2 items, numerical "For me, reducing my consumer behavior and my consumption expenditure (e.g., on clothing, everyday objects) is hard"] | Evaluation of information about (energetic) renovation (political) [Shelter: Heating] | Satisfaction, transparency, accessibility of information regarding energetical renovation [mean of 3 items, numerical "Information for an (energetic) renovation and installation of more efficient heating sources are satisfactory"] |
| Behavioral Intention (psychological), TPB | Future intention to behave sustainably [1 item, numerical, "When you think about your future behavior (regardless of current pandemic-related limitations), how much do you agree with the following statements? - I intend going without a car and using alternative means of transport (bicycle, public transport, etc.)."] | Evaluation of information about sustainable heating behavior (political) [Shelter: Heating] | Accessibility and transparency of ecological heating behavior information in general [mean score of 2 items, numerical "Information on environmentally friendly heating behavior at home (e.g., setting the water temperature, ventilation behavior, thermostat regulation) are:"] |

Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Fig. 1, brackets behind predictor names indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. TPB: Theory of Planned Behavior. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in Supplementary Notes 2.

results suggest that retired individuals exhibited less mobility related GHG emissions. This may be due to overall low usage of high GHG transportation means such as cars, planes, and buses (Supplementary Table 8).

Of the internal factors, *perceived behavioral control* constituted the most important predictor of mobility-related GHG emissions. Low subjective control over using alternative transportation means to get to places of interest (e.g., work, grocery stores) was associated with higher GHG emissions (Fig. 2). This result illustrates that if individuals perceived alternative mobility options like public transportation or bike riding as not feasible, they exhibited higher GHG emissions, likely due to more frequent and extensive car use (Supplementary Table 5). Similarly, *negative emotions*, attitudes (*behavioral beliefs*) and low *personal norms* towards alternative transportation means as well as *dissatisfaction with alternative mobility* options were associated with higher GHG emissions (Fig. 2). That is, if individuals were dissatisfied with infrastructural circumstances of public transportation, simply do not like sitting in buses or riding a bike to work or if their family and friends also do not use alternative options, they may have been more likely to resort to the car more often (Supplementary Table 5). People's travel mode *habit* was another important predictor of GHG emissions in our analysis. Individuals who routinely chose alternative transportation means over driving had much lower GHG emissions than their counterparts. This finding was expected, since travel mode choices are very habitualized overall[35]. Although being less important for GHG emission prediction overall, behavioral beliefs, norms and perceived behavioral control regarding air travel seem to influence people's mobility footprint as well. Like car usage, if individuals and their social contacts held positive attitudes towards air travel and saw no feasible alternatives, they were more likely to fly, leading to higher emissions (Supplementary Table 6).

Although these findings suggest mobility related GHG emissions to be largely under individual control through travel mode choices, which is in line with some previous arguments[22,36], we also found external factors affecting travel mode choice to be particularly relevant in predicting GHG emissions. The relative importance of these factors in our models strongly highlight the role of mobility infrastructure and living circumstances. Most prominently, *added travel time* using alternative mobility options played a major role (most important predictor after car possession and income) in mobility behavior: If individuals had to spent considerably more time for their daily trips using alternative mobility options, their footprint increased substantially (Supplementary Fig. 6), presumably due to more extensive car usage (Supplementary Table 7). This result is in line with findings from another recent study showing relative travel time (among other travel mode attributes) to affect subsequent commuting behaviors[37]. As expected, people reporting low subjective control over and less habitual use of using alternative

**Table 2 Overview of previously investigated factors (predictors) in the models to predict sustainable behavior (continuation of Table 1).**

| INTERNAL FACTORS | | EXTERNAL FACTORS | |
|---|---|---|---|
| **Name of predictor (factor type)** | **Description [number and type of items, example item]** | **Name of predictor (factor type) [domain]** | **Description [number and type of items, example item]** |
| Values (psychological), VBN | General human values, Schwartz scales<br>*Subscales:* Self-Direction, Power, Universalism, Achievement, Security, Stimulation, Conformity/Tradition, Hedonism, Benevolence<br>[all subscales: mean value of 3 items, numerical] | Energy advice municipality (political)<br>[Shelter: Electricity] | Information about electricity savings behavior muncipality<br>[1 item, categorical<br>In 2019, my municipality/city offered free advice on energy efficiency, energetic renovation and/or energy saving.] |
| NEP Growth/ Technology (psychological), VBN | New ecological paradigm: Growth, Harmony and Technology<br>[sum score of 3 items, numerical<br>"In order to survive, people have to live in harmony with nature."] | Employer support of alternative mobility options (situational/ infrastructural)<br>[Mobility] | Financial and situational support of using alternative means of transportation at work<br>[1 item, categorical<br>"In 2019, my employer promoted the use of alternative mobility options (e.g., the possibility of a work bike, more vacation days without a flight, job tickets)."] |
| Awareness of Consequences (psychological), VBN | Awareness of consequences of climate change<br>[sum score of 4 items, numerical<br>"The consequences of climate change are being dramatized."] | Homeoffice (situational/ infrastructural)<br>[Shelter, Mobility] | Possibility and use of Homeoffice<br>[2 items, binary coded<br>"How many days a week did you work from home on average in 2019 (remote work / home office)?"] |
| Ascription of Responsibility (psychological), VBN | Acknowledgement of responsibility towards the environment/climate<br>[sum score of 2 items, numerical<br>"Climate change is mainly caused by humans."] | Tenant/owner of house/ flat (situational/ infrastructural)<br>[Shelter: Electricity] | Ownership structure of housing<br>[1 item, categorical<br>"Were you tenant or owner of your apartment / house in 2019?"] |
| Perceived Consumer Effectiveness (psychological), VBN | Perceived Effectiveness of one's own behavior for environment/climate change<br>[sum score of 2 items, numerical<br>"I can't do much to protect the environment on my own."] | Involved in care of relatives (situational/ infrastructural)<br>[Mobility] | [1 item, binary coded<br>"At the end of 2019, did you have relatives in need of care outside of your household that you looked after?"] |

Continued from Table 1. Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Fig. 1, brackets behind predictor names indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. VBN: Value Belief Norm Theory. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in Supplementary Notes 2.

options for their everyday trips also mostly reported high levels of added travel time for these options and vice versa (Supplementary Figs. 8 and 9). Furthermore, travel time requirements seemed to interact with some internal factors. The GHG reducing effect of positive attitudes and emotions towards travel modes like trains and buses seemed to be less pronounced for individuals who face substantially prolonged travel times when using these alternative means (Supplementary Figs. 10 and 11). Although to lesser extent, external factors like the physical *distance to a train station* and *living area* (city vs. rural) also seem to influence mobility decisions and resulting GHG emissions (Fig. 2, Supplementary Fig. 7, and Supplementary Table 7). This finding is partly consistent with previous study results[15,22,38].

Next, we focus on the diet domain (Fig. 3), which highlights the importance of internal factors. Dietary *habits* were the key determinant of individual GHG emissions. That is, self-reported automaticity of sustainable diet habits was most predictive of a GHG-emission-friendly diet. Although still exhibiting high relative importance, other internal factors like *perceived behavioral control*, *behavioral intentions*, and attitudes (*behavioral beliefs*) on dietary behavior which previously have been claimed to be the main predictors in this domain[39–41], deemed less important than dietary habits. Two main reasons arguably account for this result. First, our footprint calculation incorporated purchasing frequency of local, seasonal, and organically produced foods as well as individuals' dietary form. This comprises a more comprehensive behavioral measurement than many of the above-mentioned previous studies (e.g., buying frequency of "eco-friendly", "green", or organically produced food or meat consumption). Second, an individual's diet (e.g., omnivorous, vegetarian, vegan) seems to be comparatively stable over time and comprises the singular most important determinant of overall diet GHG emissions[42–45]. This is arguably why we found dietary habits to be very influential. Moreover, contemporary studies showed diet habits to significantly reduce the influence of other factors on diet behavior[44,46], which again speaks towards the necessity of investigating factors of sustainable behavior in unison to assess their relative importance.

Value orientations and social norms were also relevant in predicting diet related GHG emissions. People pursuing goals of social status (*power*) or pleasure and gratification (*hedonism*) exhibited higher diet related GHG emissions. Relatedly, politically conservative individuals showed higher diet related GHG emissions than liberals (*Ideology*) and individuals believing that behaving sustainably constitutes a good citizen (*Citizen norms: Sustainability*) showed lower diet-related GHG emissions overall. These results are in accordance with previous research on meat consumption and veganism showing that, for instance, conservative people and those valuing goals of power reported higher meat consumption and more lapses from a vegetarian and vegan diet[44,47,48]. Further, our findings suggest that individuals scoring high on *conscientiousness* and *neuroticism* eat more sustainable. In general, the literature on the relationship between dietary patterns and personality traits revealed mixed findings in the past[49]. Our results, however, align with more recent studies and meta-analyses, showing that high conscientiousness is associated

**Table 3 Overview of previously investigated factors (predictors) in the models to predict sustainable behavior (continuation of Tables 1–2).**

| INTERNAL FACTORS | | EXTERNAL FACTORS | |
|---|---|---|---|
| **Name of predictor (factor type)** | **Description [number and type of items, example item]** | **Name of predictor (factor type) [domain]** | **Description [number and type of items, example item]** |
| Personal Normative Beliefs (psychological), VBN | Personal normative beliefs on climate obligations [sum score of 2 items, numerical "I feel obliged to do everything in my power to stop climate change."] | Possibility to renovate apartment energetically (situational/ infrastructural) [Shelter: Heating] | [1 item, categorical "If I wanted, I could ensure that my (rented) apartment / house would be energetically renovated and/or a more efficient heating source would be installed."] |
| Habits (psychological) | Habit of behaving sustainably [1 item, numerical "Saving electricity is something I do automatically."] | Flight status (situational/ infrastructural) [Mobility] | [1 item, categorical "Did you fly at least once in the past three years (2018 to 2021)?"] |
| Emotions (psychological) | Emotions towards sustainable behavior (happiness and satisfaction): [mean score of 2 items, numerical "How do you think you would feel if you were to eat sustainably in the near future?"] | Car possession (situational/ infrastructural) [Mobility] | Car possession [1 item, categorical "Do you own a car?"] |
| Knowledge (psychological) | Knowledge about determinants/processes of climate change [sum of correct answers to 9 items, multiple choice format "Compared to the corresponding amount of vegetables in terms of calories, beef is… harmful for the environment"] | Satisfaction alternative mobility options (situational/ infrastructural) [Mobility] | Satisfaction with infrastructure using different mobility types (bike, public transport, walking) [mean score 3 items, numerical "How satisfied were you in 2019 with the connection, travel time and frequency (for public transport) as well as the transport infrastructure of the following mobility options in your residential area for typical destinations such as your work, supermarket or leisure"] |
| Political efficacy (psychological) | Internal and external political efficacy mean scores of subscales: *Internal:* [4 items, numerical] *External:* [4 items, numerical] ["I can understand and assess important political questions very well.", "The politicians strive to maintain close contact with the population."] | Availability of special mobility options (situational/ infrastructural) [Mobility] | Availability of car sharing, cycle paths, park & ride infrastructure, public transport discounts and other special mobility options [sum score of 5 items, categorical "Were there the following mobility options in your area in 2019? - Public transport offers (e.g., for professional groups, family cards, etc.)"] |

Continued from Tables 1–2. Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Fig. 1, brackets behind predictor names indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. VBN: Value Belief Norm Theory. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in Supplementary Notes 2.

with higher fruit and vegetable intake and more healthy diets in general and that vegetarians report higher levels of Neuroticism[49–51]. In contrast to many previous studies on ecological food consumption[52], we did not find income to be a limiting factor to pursue sustainable diets in general. Importantly, consumption of ecologically produced foods represent only one part of the diet GHG equation[44]. Pursuing a conventionally produced, mostly plant-based diet – on the other hand – is usually not more expensive than a meat-based diet[53]. Thus, lower income may not inevitably lead to high GHG emissions, an assumption corroborated by our results.

Surprisingly, and in contrast to the mobility-sector results, external factors did not seem to be overly relevant for diet related GHG emissions. The only two relevant external factors were the *availability of sustainable food in restaurants* or cantinas people usually eat at and if people do or do not *eat out regularly*. Those who eat out more regularly and reported lower availability of sustainable food options in their local restaurants exhibited higher GHG emissions (Fig. 3). Otherwise, knowledge about sustainable food options, the availability of local, seasonal, and organically produced foods in supermarkets or their identifiability, for instance, do not seem to be limiting factors. Notably, this finding is in line with only a few previous studies on sustainable food consumption[54,55].

In general, our results regarding the diet domain support claims of more general theories of sustainable behavior proposing human values and pro-environmental attitudes to be precursors of sustainable behavior[12,56,57]. That is, diet-related $CO_2$ emissions seem to not only be driven by diet-specific internal factors like habits or perceived control but also by individuals' personality and more general (ideological) beliefs, values, and norms revolving around sustainability which can form the motivational basis of dietary patterns. However, as argued above, the relative importance of the latter factors is lower in comparison to the former factors (see Fig. 3).

**Drivers and barriers of impactful sustainable behavior in domains shelter and consumption**. Unlike when predicting mobility and diet GHG emissions, attempts to predict people's shelter (electricity & heating) and consumption GHG emissions were only partly successful. Notably, previous studies predicting shelter related GHG emissions showed better performance and found demographic (e.g., income, age), external factors (e.g., community zone) and dwelling characteristics (e.g., fuel and apartment type, household size) to be relevant[58–60]. Although we found similar factors like *living area*, *income* and *age* to be among the most important factors in the shelter domain in our models

**Table 4 Overview of previously investigated factors (predictors) in the models to predict sustainable behavior (continuation of Tables 1–3).**

| INTERNAL FACTORS | | EXTERNAL FACTORS | |
|---|---|---|---|
| **Name of predictor (factor type)** | **Description [number and type of items, example item]** | **Name of predictor (factor type) [domain]** | **Description [number and type of items, example item]** |
| Political trust (psychological) | Trust in different official institutions [mean score of 9 items, numerical "Please indicate below to what extent you trust the respective institutions or groups of people. - German federal government"] | Satisfaction special mobility (situational/ infrastructural) [Mobility] | Satisfaction with special mobility options, such as car sharing, cycle paths, park & ride infrastructure, public transport discounts and other special mobility options [mean score of 5 items, numerical "How satisfied were you with the following mobility option (s) near you in 2019? - infrastructure for "park and ride""] |
| Big Five (psychological) | Big Five personality factors sum scores of Subscales: *Conscientiousness* [3 items, numerical] *Openness to Experience* [3 items, numerical] *Extraversion* [3 items, numerical] *Agreeableness* [3 items, numerical] *Neuroticism* [3 items, numerical] ["I am someone who..."] | Development of e-Mobility Infrastructure (situational/ infrastructural) [Mobility] | Development of charging infrastructure and conditions for e-Mobility [mean score of 2 items, numerical "How good were the infrastructural conditions in 2019 at your typical everyday destinations for electric cars (e.g., free parking spaces, extra lanes/parking spaces for electric cars)?"] |
| Voting Intention: Party (psychological) | Voting intention in the next German federal election [1 item, categorial "If there was a federal election next Sunday, which party would you vote for with your ZWEITSTIMME?"] | Distance to nearest public transport (situational/ infrastructural) [Mobility] | Distance to nearest bus and train station [2 items, integer number (Distance in km)] |
| Ideology (psychological) | Self-Placement in one-dimensional political Spectrum [1 item, numerical "Many people use the terms "left" and "right" to denote different political attitudes. We have a yardstick here that runs from left to right. When you think of your own political views, where would you place those views on this scale? Please tick the appropriate box."] | Added travel time alternative mobility (situational/ infrastructural) [Mobility] | Change in time spent commuting to when using alternative transportation [3 items, recoded to one item "How much more/less time did you have to spend on average if you had used alternative means of transport (e.g., public transport, bicycle, etc.) instead of a car to get to work?"] |

Continued from Tables 1–3. Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Fig. 1, brackets behind predictor names indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in Supplementary Notes 2.

(Fig. 4), the overall prediction accuracy of our models in this domain remains relatively low (see Fig. 1 and Supplementary Discussion).

An explanation for the lower prediction accuracy in our models may be that shelter related emissions are under lower personal control than emissions in other domains[18,61]. This is indicated by the relatively high factor importance of perceived control in our models (Fig. 4) and becomes even more apparent when reviewing the determining factors of shelter related GHG emissions: The amount of energy necessary for space and water heating (and subsequent GHG emissions) is largely determined by the installed heating source (i.e., fuel type), type and state of the dwelling as well as its size[62]. An oil heating system produces, all other things being equal, much more GHGs than, for instance, an electrical heat pump fueled by solar power[63]. Similarly, a newly built, low-energy house radiates much less heat to the outside than a mid-18th century building. At the same time, individuals living in large, single family homes produce much more GHGs than individuals living in a small flat[62]. Simply speaking, if people live in spacious, old buildings with bad isolation and high-emission energy sources, they can heat as frugally as they want and still produce large amounts of GHG emissions[18,25]. Regarding electricity, the most important determinant of GHG emissions is the generative source[61,64]. High GHG electricity generation can quickly undermine benefits of saving behaviors[61]. Therefore, consumers receiving electricity generated from renewable sources (e.g., solar, wind or hydro-electric) can use considerably more energy than their non-renewable counterparts and still produce less GHGs (see also current net avoidance factors in Germany[25,65]).

Of course, we could have simply included dwelling characteristics and electricity type in our model as additional external factors. However, in doing so we would have confounded predictor and outcome variables because these variables are already included in carbon footprint calculators (as the one we used) to estimate individual GHG emissions. For instance, people's heating behavior (e.g., thermostat setting, ventilation habits) is offset against their dwelling characteristics (e.g., fuel type, isolation, living space, household size) when their footprint is estimated (Supplementary Note 3). Since dwelling and electricity generation characteristics are the main determinants of individual GHG emission, other internal and external factors

**Table 5 Overview of previously investigated factors (predictors) in the models to predict sustainable behavior (continuation of Tables 1-4).**

| INTERNAL FACTORS | | EXTERNAL FACTORS | |
|---|---|---|---|
| **Name of predictor (factor type)** | **Description [number and type of items, example item]** | **Name of predictor (factor type) [domain]** | **Description [number and type of items, example item]** |
| Political participation (psychological) | Political participation regarding environmental issues<br>*Sustainability as Topic of Conversation:*<br>[1 item, categorical<br>"There are various ways in which one can try to improve something in Germany or to prevent something from getting worse. Have you done any of this over the past 12 months?<br>- Discussed the topic of "sustainability" with friends / family/other people"]<br>*Member of an Environmental Organization:*<br>[1 item, categorical<br>"Some people are members of different groups or associations. Please indicate whether you are a member of the following groups yourself and how you are involved.<br>- environmental organization"] | Care of relatives (situational/infrastructural) [Mobility] | Regular care work for relatives outside of own house [1 item, binary<br>"At the end of 2019, did you have relatives in need of care outside of your household that you looked after?" |
| Conspiracy belief: Climate Change (psychological) | Conspirational thinking about human-cause climate change<br>[1 item, numerical<br>"Man-made climate change is an excuse to patronize citizens or to tax them."] | Eating out regularly (situational/infrastructural) [Diet] | Regular visits to canteens/restaurants [1 item, binary<br>"Did you eat out regularly (at least once a week) in 2019 (e.g., canteen/ restaurant at work)?"] |

Continued from Tables 1-4. Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Fig. 1, brackets behind predictor names indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in Supplementary Notes 2.

in our models can only account for the effect of individual behavior regarding heat and electricity savings on GHG emissions. Individual savings behavior, however, affects the overall footprint to a much lesser extent, as argued before[25]. The fact that we did not include dwelling characteristics and electricity generation in our models may partly explain the lower prediction accuracies of our models compared to previous studies which did include factors like fuel type, household size, and apartment type to predict shelter related emissions. Nevertheless, we consider it important to not conceptually confound predictors with the measurement of the to-be-predicted behavior. In summary, regulatory efforts on the supply-side (e.g., implementation of renewable electricity generation and energetically efficient dwellings) as well as demand-side (e.g., target people's dwelling characteristic needs) seem to be the most promising approaches to notably reduce shelter GHG emissions.

A challenge for investigating consumption related GHG emissions is the tractability of purchased products' complete environmental life cycle. That is, calculating accurate footprints would require assessing every product bought in the year of interest, including a detailed description of the brand, type, and origin of goods to estimate product lifecycle emissions[66]. Even though tracking of singular products might be possible, assessing and calculating direct and indirect GHG emissions for all products that individuals bought over a year is virtually unattainable. As a result, we surveyed more general consumption patterns (e.g., frequency of buying second-hand goods, monthly spending on goods) based on recommendations in previous studies (Supplementary Note 3) to calculate consumption based GHG emissions. However, this estimation method might be more prone to memory errors, leading to less accurate GHG emission values. In fact, the GHG distribution of our sample in this domain differed the most from the expected distribution for the German population (Supplementary Fig. 3). Although in our models we found factors which were relevant for consumer behavior in previous studies (e.g., income[59] and habits[67]) to also be the most important predictors of consumption GHG emissions (Fig. 5), the overall model prediction performance remains relatively low (Fig. 1). For the present study, we found no feasible way of calculating consumption related GHG emissions more accurately. Future research could explore the utility and improvements in calculation accuracy of alternative approaches (e.g., longitudinal recordings of goods consumption via mobile apps).

## Discussion

Contemporary solutions to climate change mitigation target individual behavior and needs. Developing these solutions effectively, however, requires a comprehensive understanding of the factors influencing individual sustainable behavior in everyday life as well as their interactions. In this regard, our study provides important insights.

First, we were able to empirically demonstrate that individual sustainable behavior in different life domains is shaped by a multitude of different factors and their mutual interactions. The higher prediction accuracy of the ML models compared to traditional linear regression models commonly used to study factors of sustainable behavior shows that the former, more complex models are more appropriate for investigating predictors of GHG emissions.

Second, we were able to analyze the relative importance of predictors of GHG emissions more accurately than previous studies by considering multiple predicting factors for different life domains simultaneously, accounting for their complex interplay, measuring behavior comprehensively, and validating results on out-of-sample data. We found many important factors in our models for different life domains to resemble those identified in

**Table 6 Overview of previously investigated factors (predictors) in the models to predict sustainable behavior (continuation of Tables 1–5).**

| INTERNAL FACTORS | | EXTERNAL FACTORS | |
|---|---|---|---|
| **Name of predictor (factor type)** | **Description [number and type of items, example item]** | **Name of predictor (factor type) [domain]** | **Description [number and type of items, example item]** |
| Media consumption (psychological) | Number of regularly consumed (non-) mainstream media formats: Frequency of Use, Media for News Consumption, Mainstream Media Consumption, Non-Mainstream Media Consumption [3 items each, sum score "Please indicate how often you have used the medium mentioned on average over the past 12 months. - TV", "Please indicate how often you have used the medium mentioned on average over the past 12 months. - News programs on TV (also online via the media library)", "Which sources of information do you mainly use to find out about politics in the media? (Multiple choices possible)"] | Availability sustainable groceries (situational/ infrastructural) [Diet] | Availability of seasonal, organic or regional groceries [1 item, numerical "In order to be able to buy organic, regional or seasonal products, I had to go a long way and undergo disturbing circumstances in 2019."] |
| Citizen Norms: Sustainability (psychological) | Environmental citizenship norms [1 item, numerical (7 point scale) "There are different opinions about what constitutes a "good citizen". How important would you rate the following behaviors to be a good citizen? - Behave sustainably and environmentally friendly in everyday life"] | Availability sustainable food in restaurant (situational/ infrastructural) [Diet] | Price and options of vegetarian and vegan dishes in frequently visited restaurants/ canteens [sum score of 5 items, numerical "The restaurants/canteens I ate in regularly offer a wide range of vegetarian dishes."] |
| Citizen Norms: Political Consumerism (psychological) [*Consumption, Diet*] | Citizenship norms [1 item, numerical (7 point scale) "There are different opinions about what constitutes a "good citizen". How important would you rate the following behaviors to be a good citizen? - Select products according to political, ethical or ecological aspects, even if they are more expensive"] | Possibilities of sustainable consumption (situational/ infrastructural) [Consumption] | Possibility to rent and buy second hand [mean score of 2 items, categorical "In 2019 there were opportunities to borrow utensils (e.g., lawn mowers, drills, etc.) that I rarely need near my place of residence."] |

Continued from Tables 1–5. Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Fig. 1, brackets behind predictor names indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in Supplementary Notes 2.

previous studies, but we also found the importance of quite a few factors to deviate from previous literature.

For policy makers, our results suggest that reducing mobility related GHG emissions demand actions on multiple levels. First and foremost, alternative mobility options like public transportation, car sharing, or bike riding must become more attractive and widely available. As indicated by our results, people will not refrain from high GHG transportation means such as cars or national flights if alternative options are not accessible, inappropriate, overly time-consuming, or simply out of their behavioral control, even if they have the desire to use them. These findings are in line with previous studies on external factors of mobility behavior[15,38] and also point towards so-called "lock-in effects": The situational and infrastructural circumstances of individuals' living sectors may lock them in to certain behaviors (i.e., car usage) to fulfill their needs, like going to work or grocery shopping, which makes direct behavioral interventions less effective[68–70]. These circumstances can only be improved by investing in reliable, fast, and ample public transportation infrastructure and changes in city designs and land use[29,70]. Unlocking travel mode choices and satisfying individual needs through targeting mobility circumstances is likely to lead to changes in mobility behavior and may also reduce the strong effect of income on mobility GHG emissions found in our analysis and in previous studies[3,29]. However, when implementing such measures, policy makers must be aware of "rebound effects" (i.e., the saved time and money is spent on other consumables or increased leisure air travel) that may negate some of the GHG reductions[29]. An alternative approach to reduce the particularly strong effect of income on mobility related GHG emissions may also include more general socio-ecological transformations, such as reducing contracted working times and discussing ways of economic degrowth while still maintaining (or even improving) well-being. However, the wider implications of such measures have been debated[29].

Although situational and infrastructural barriers limit individual agency and the potential effectiveness of behavioral interventions, our results indicate internal factors (e.g., attitudes, norms, and habits) to be important for choosing travel modes. Therefore, directly targeting internal factors seems promising for reducing GHG emissions. This may be addressed by supporting habit changes[71,72], financial incentives to use and continue using alternative options[15,17,73,74], or promoting their benefits[75] to change people's attitudes and norms. The impact of more general

**Table 7 Overview of previously investigated factors (predictors) in the models to predict sustainable behavior (continuation of Tables 1-6).**

| INTERNAL FACTORS | | EXTERNAL FACTORS | |
|---|---|---|---|
| **Name of predictor (factor type)** | **Description [number and type of items, example item]** | **Name of predictor (factor type) [domain]** | **Description [number and type of items, example item]** |
| Gender (demographical) | Self-identified gender of respondent [1 item, categorical (male, female, other)] | Knowledge electricity (technological) [Shelter: Electricity] | *Knowledge of annual Electricity Consumption* [1 item, binary coded "I know my annual electricity consumption from 2019."] *Knowledge of Electricity Consumption of Household Devices* [1 item, numerical "In 2019 I knew exactly which devices in my household were using the most electricity."] |
| Federal home state (demographical) | 1 item, categorical | Traceability electricity (technological) [Shelter: Electricity] | *Power Consumption in own Household* [1 item, numerical "In 2019 I was able to see and understand my current power consumption exactly at all times. Information: For instance, using a smart meter can make electricity consumption traceable."] |
| Income (demographical) | Monthly income in € 1 item, numerical | Traceability heating (technological) [Shelter: Heating] | *Heating consumption traceability in own Household* [1 item, numerical "In 2019 I was able to see and understand my current heating consumption exactly at all times."] |
| Professional status (demographical) | Job status [1 item, categorical looking for a job, apprenticeship, student, employee, civil servant, self-employed, retired] | Traceability consumption (technological) [Consumption] | *Traceability of sustainability of new Products* [1 item, numerical "When I buy new everyday objects (e.g., furniture, electronics or clothing), I can see exactly how sustainably they have been produced."] |
| Highest educational qualification (demographical) | 1 item, categorical | Traceability diet (technological) [Diet] | *Identifiability of sustainable Food* [2 items, numerical "The labeling of ecological, regional and organic food is easy to understand."] |
| Living area (demographical) | Area of residency (urban/rural) 1 item, categorical | | |
| PLZ (demographical) | First three ZIP Code digits 1 item, numerical (integer number) | | |
| Age (demographical) | 1 item numerical (integer number) | | |

Continued from Tables 1–6. Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Fig. 1, brackets behind predictor names indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in Supplementary Notes 2.

factors of pro-environmentalism (e.g., environmental education, perceived consequences and responsibility or norms regarding climate change) on reducing mobility related GHG emissions seems negligible, which is in line with some previous studies[22,72,76]. Nevertheless, implementing regulative measures that target these factors (e.g., pointing out consequences or individual responsibility for climate change) may still lead to behavioral changes in the long run[3,17].

Regarding diet related GHG emissions, our results highlight the potential and need for demand-side solutions. Dietary habits seem to be the main component of diet-related GHG emissions. Therefore, extrapolating habit-breaking strategies that have been shown to, for instance, help reducing meat consumption or eating healthier to sustainable diets in general may be promising[46,77]. Although to a lesser extent, positive attitudes, norms and perceived control towards sustainable diets and sustainability in general are also predictive of diet behavior. Since these factors are precursors of intention formation[78] and motivate behavior[12], regulatory efforts may be best invested in promoting the multitude of positive effects of sustainable diets (e.g., climate and health benefits[17,44]). In contrast to previous suggestions, factors like availability of sustainable food options in supermarkets, their identifiability or individual income do not seem to contribute much to dietary choices and thus may not be promising targets for encouraging GHG-emission-friendly diets. Crucially, however, this does not imply that targeting availability and prices of conventional food options (meat products in particular) would also be inefficient in reducing GHG emissions. Quite the contrary, implementing policies that shift financial and structural power from conventional food lobbies and the meat industry to the organic sector may substantially complement the effectiveness of campaigns promoting dietary shifts[79].

Furthermore, our findings indicate that shelter-related emissions are strongly influenced by dwelling and electricity generation characteristics. Consequently, a combination of supply-side and demand-side solutions should have big effects on GHG emissions in this domain. Making use of technological advancements in dwelling solutions and energetically renovating buildings contribute more to GHG reductions than, for instance, information campaigns on saving energy or promoting environmental awareness. This notion is backed up by multiple studies in environmental science literature showing the importance and climate benefits of directly targeting dwelling characteristics, like house construction and heating systems[60,80–82]. Similar to

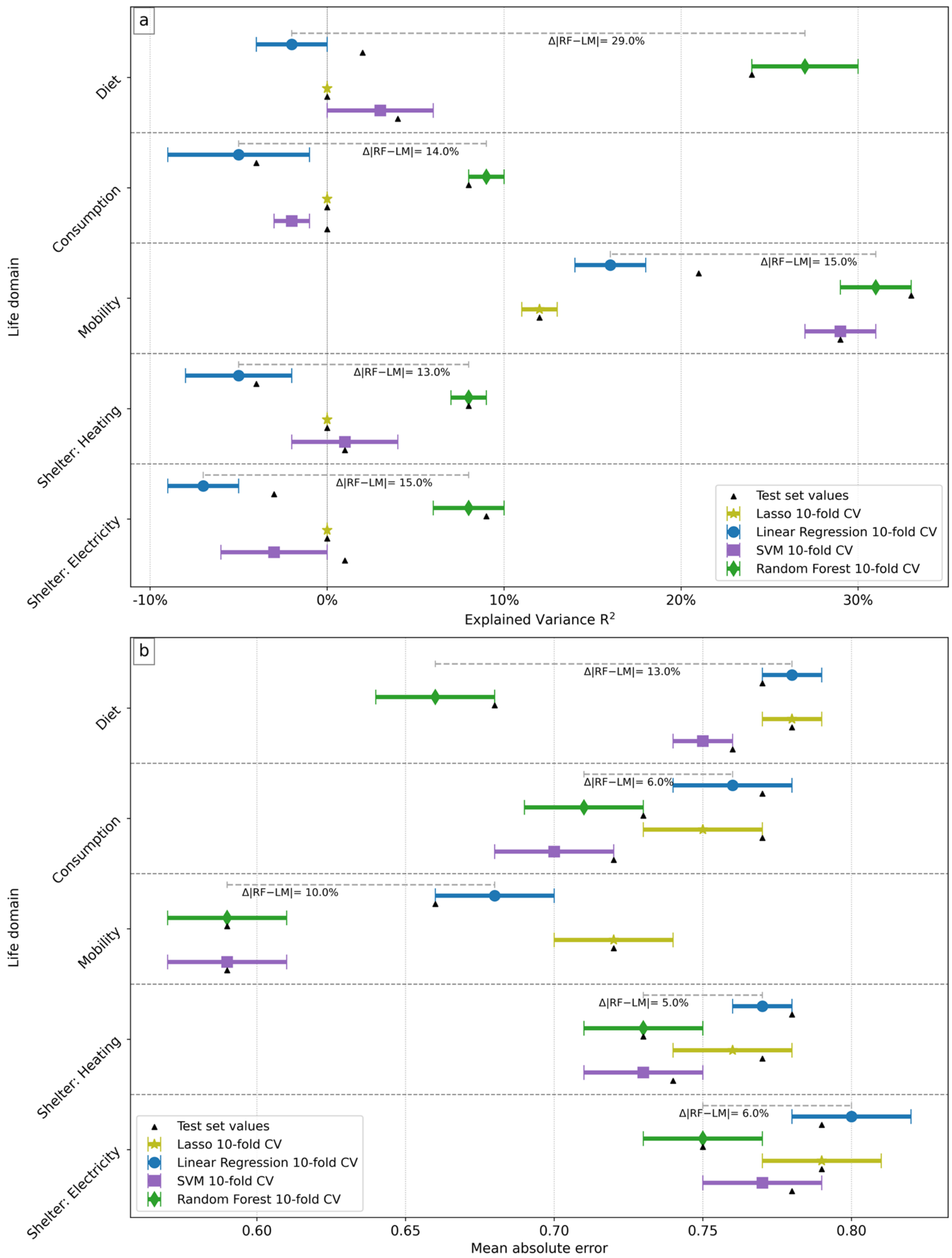



**Fig. 1 Prediction performance of individual GHG emissions in all life domains (cross-validation and independent test set). a** Explained variance and **b** mean absolute prediction error of the Random Forest (RF), Linear Regression (LM), Lasso Regression (LASSO) and Support vector machine (SVM) models. Colored error bars indicate performance in the 10-fold cross-validation (*CV*) on the training set, spanning a 1 SD interval. Star and square indicators represent average performance across folds. Upward pointing triangles below the error bars indicate prediction performance on the independent test set. Dashed lines indicate absolute performance difference between the LM and RF. Values are rounded to two decimal places. Negative $R^2$ values indicate a worse model fit than always predicting the mean GHG emission ($R^2 = 0$) (see Supplementary Methods).

heating, expending renewable electricity production has a more direct effect in reducing GHG emissions than just educate or encourage people to save electricity[61]. However, large-scale decarbonization of the electrical grid as well as building and adaptation of more energy efficient homes may not be enough to meet global climate goals. Within the demand-side framework, researchers have advocated for implementing policies that more directly target resource-intensive living standards. As mentioned previously, the size of dwellings and their type (single family vs. multifamily housing) particularly contribute to shelter-related GHGs. Thus, promoting broader societal changes by directly addressing city planning and individual lifestyles, for instance by supporting compact city designs and multifamily housings as well as incentivizing individuals to reduce their living space while still maintaining personal well-being, might also be necessary[3,5,68].

Our study comes with a few limitations. Although we did our best to identify all previously investigated internal and external factors influencing sustainable behavior, there is a possibility that certain factors may have eluded our investigation. Nonetheless, based on previous study results and our comprehensive review of existing literature, it seems likely that we managed to include the most important factors in our analyses. Further, whereas the

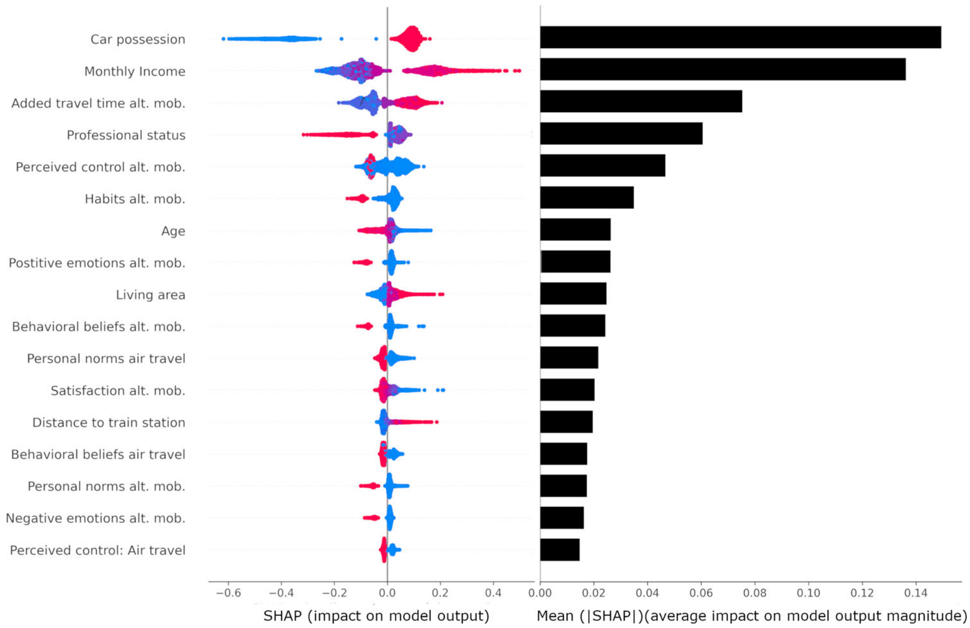



**Fig. 2 Predictor importance for individual GHG emissions for mobility domain in the Random Forest.** Summary (beeswarm) plot and predictor importance (black bars) for life domain mobility, predictor importance was calculated using Shapely values. The summary plot shows the relationship of individual predictor values with model prediction compared to the average prediction. Dots represent individuals in the dataset, overlapping points are jittered on the y-axis. Individual values on the respective predictors range from low (blue hue) to high (red hue). Positive SHAP values indicate a change in model prediction towards higher emissions. Black bars indicate the overall importance of the predictor for the model prediction performance. Predictors are sorted by their relative importance. Only predictors with average importance (mean SHAP value) above the mean importance of all predictors are shown (i.e., most important predictors) but the plot is based on including all predictors in the model. For more information, see Supplementary Methods.

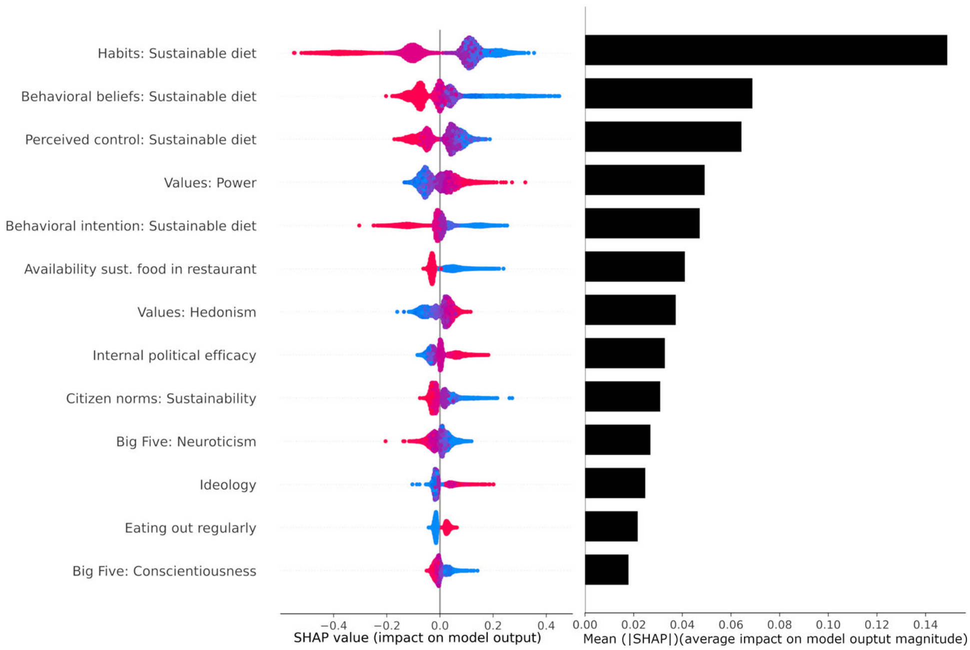



**Fig. 3 Predictor importance for individual GHG emissions for diet domain in the Random Forest.** Summary (beeswarm) plot and predictor importance (black bars) for life domain diet, predictor importance was calculated using Shapely values. The summary plot shows the relationship of individual predictor values with model prediction compared to the average prediction. Dots represent individuals in the dataset, overlapping points are jittered on the y-axis. Individual values on the respective predictors range from low (blue hue) to high (red hue). Positive SHAP values indicate a change in model prediction towards higher emissions. Black bars indicate the overall importance of the predictor for the model prediction performance. Predictors are sorted by their relative importance. Only predictors with average importance (mean SHAP value) above the mean importance of all predictors are shown (i.e., most important predictors) but the plot is based on including all predictors in the model. For more information, see Supplementary Methods.

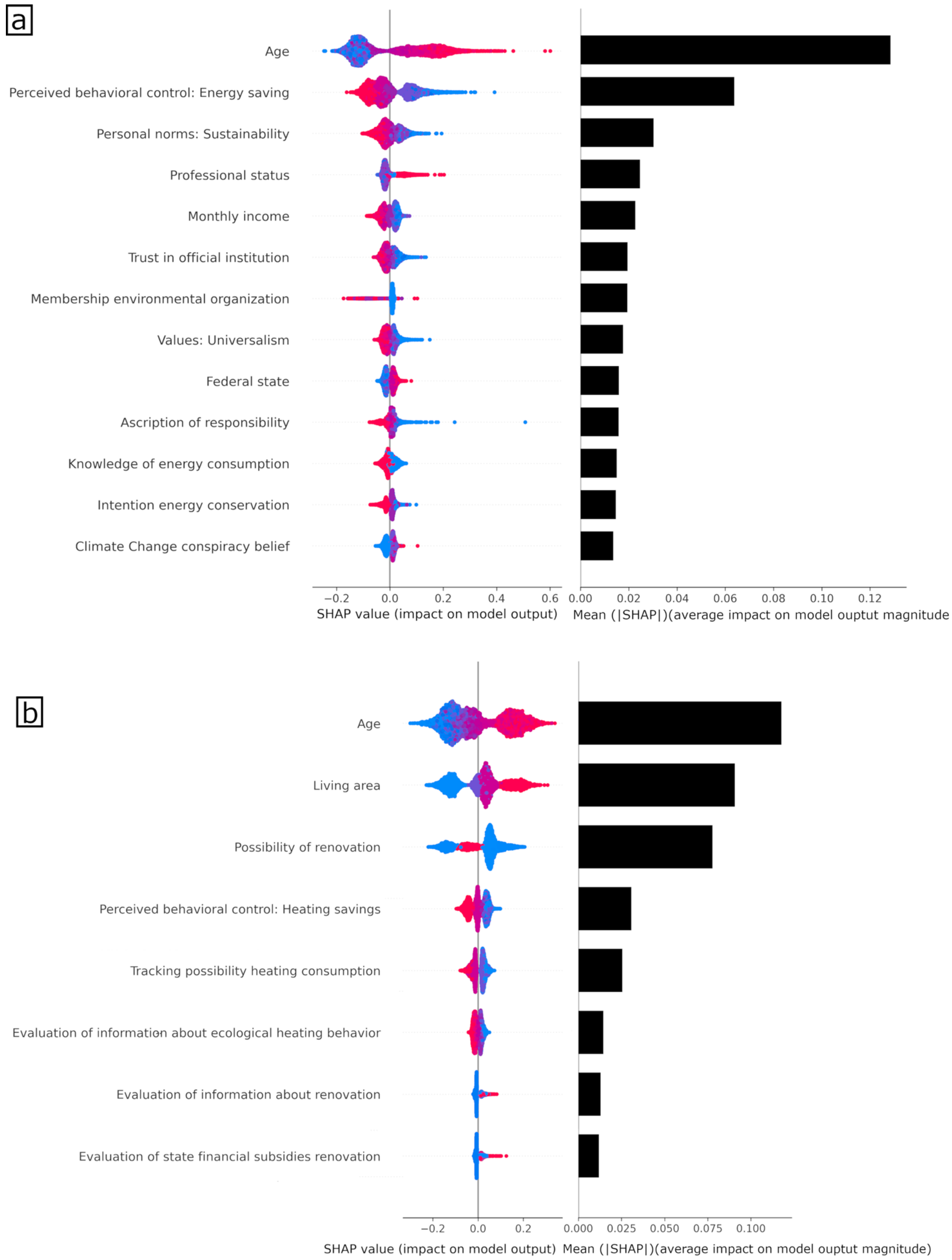



**Fig. 4 Predictor importance for individual GHG emissions for shelter domain in the Random Forest.** Summary (beeswarm) plot and predictor importance (black bars) for life domain shelter. **a** Domain Shelter: Electricity, **b** Domain Shelter: Heating. Predictor importance was calculated using Shapely values. The summary plot shows the relationship of individual predictor values with model prediction compared to the average prediction. Dots represent individuals in the dataset, overlapping points are jittered on the y-axis. Individual values on the respective predictors range from low (blue hue) to high (red hue). Positive SHAP values indicate a change in model prediction towards higher emissions. Black bars indicate the overall importance of the predictor for the model prediction performance. Predictors are sorted by their relative importance. Only predictors with average importance (mean SHAP value) above the mean importance of all predictors are shown (i.e., most important predictors) but the plot is based on including all predictors in the model. For more information, see Supplementary Methods.

present study investigated individual behavior, emissions directly caused by individuals in their daily life only constitute part of a population's overall GHG emissions, around 18% of total GHG emissions in Germany[65], which seems representative of many Western civilizations[28,83,84]. Therefore, mitigation efforts must be extended to infrastructural, agricultural, and industrial sectors which emit the remaining part of total GHGs. Notably, average footprints and the observed drivers and barriers of sustainable behavior might not be the same everywhere across the globe[85] and thus the relative impact of regulatory measures might also vary between countries as well as over time[18,28,86,87]. Therefore, future research should extend the proposed design to non-Western countries and also observe changes in behavior and their drivers and barriers over time, another knowledge gap identified in the recent IPCC report[3]. Being able to flexibly adjust regulative measures in different transition phases and contexts based on knowledge about the respective factors and their interaction could speed up climate change mitigation drastically[3]. Although we measured impactful behavior by using externally validated GHG footprint calculators, we were unable to examine the calculation

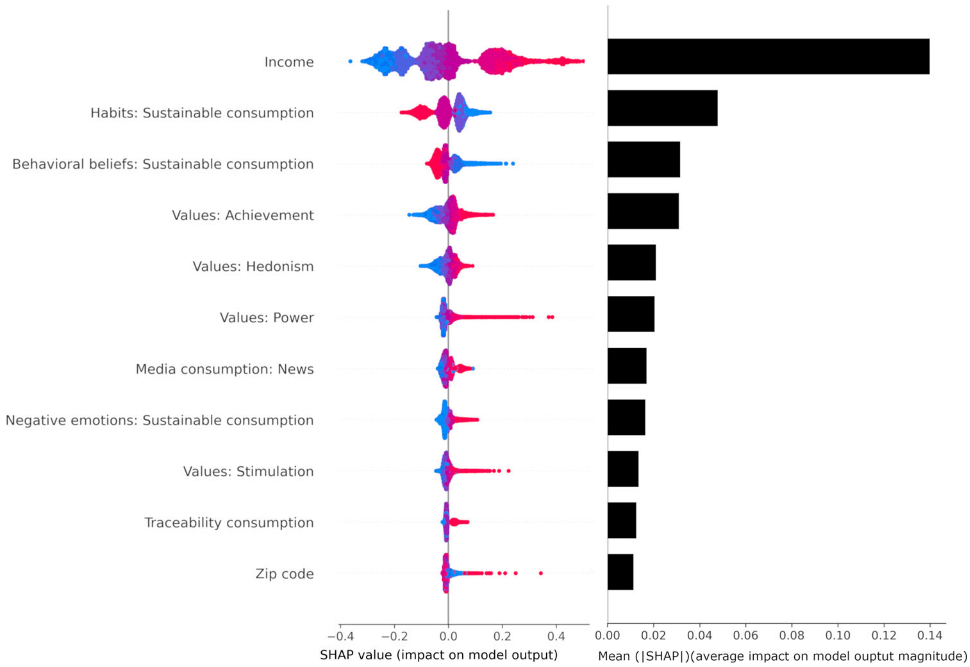



**Fig. 5 Predictor importance for individual GHG emissions for consumption domain in the Random Forest.** Summary (beeswarm) plot and predictor importance (black bars) for life domain shelter. Predictor importance was calculated using Shapely values. The summary plot shows the relationship of individual predictor values with model prediction compared to the average prediction. Dots represent individuals in the dataset, overlapping points are jittered on the y-axis. Individual values on the respective predictors range from low (blue hue) to high (red hue). Positive SHAP values indicate a change in model prediction towards higher emissions. Black bars indicate the overall importance of the predictor for the model prediction performance. Predictors are sorted by their relative importance. Only predictors with average importance (mean SHAP value) above the mean importance of all predictors are shown (i.e., most important predictors) but the plot is based on including all predictors in the model. For more information, see Supplementary Methods.

principles ourselves due to restrictions in code accessibility. Therefore, future research efforts may focus on further development of reliable GHG assessment tools and make their source code available to interdisciplinary research teams. Due to our overall study aim of identifying the most important factors driving and hindering sustainable behavior, we did not follow up on all potential interactions or relationships between predictors implicitly captured by our ML models. However, since there are likely to be more important interactions that could lead to even more tailored regulative strategies, future studies could use our data set and test for specific interaction effects.

Our findings demonstrate that the interplay of drivers and barriers to individual GHG emissions in everyday life is complex. We thus argue that factors influencing sustainable behavior should be investigated with approaches which are able to account for this complexity. We found an overall high effect of internal factors such as perceived behavioral control, habits, and attitudes on individual GHG-emission-friendly behavior. However, in some life domains, their impact can be altered or even extinguished by external factors such as infrastructural barriers or dwelling characteristics. Policy makers thus need to consider these complex interplays and may focus on the most important factors when designing demand-side solutions to climate change mitigation targeting individual behavior and needs.

## Methods

**Literature review and selection of influencing factors**. We first performed a systematic literature review to identify the relevant factors of individual sustainable behavior. We selected those factors which are grounded in established theories on sustainable behavior as well as those which were relevant in meta analyses, large scale studies and climate change mitigation reports. Since the final list of identified factors was extensive, we provide a summarized overview (see Tables 1–7). A more in-depth description of the literature review, identified factors and previous research results can be found in Supplementary Note 1, 2 and Supplementary Fig. 2.

In general, the identified drivers and barriers of sustainable behavior can be categorized into internal (i.e., person-related) and external (i.e., situation-related) factors[10]. Internal factors consisted of constructs derived from the Theory of Planned Behavior, Value-Belief Norm Theory, and habit formation. These include attitudes, (personal) norms and values, perceived behavioral control, behavioral intentions, climate change awareness and personal responsibility beyond others. We further considered environmental knowledge, emotions towards sustainable behaviors, political attitudes and voting intention as well as media consumption and demographics as internal factors. External factors subsume political, social, economic, and cultural conditions people find themselves in. They were highly domain-specific and included items like accessibility to information and feedback about energy and heating behavior, situational possibilities of eating sustainably or infrastructural mobility circumstances (e.g., access to public transportation).

**Measuring sustainable behavior**. We measured impactful sustainable behavior by calculating people's GHG footprint in $CO_2$ emission equivalents to quantify an individual's contribution to climate change. Unlike most previous studies, we assessed emissions separately for the most important domains of everyday life: 1. shelter (electricity & heating), 2. mobility, 3. consumption and 4. diet. This is crucial, since individuals who behave sustainably in one life domain (e.g., exclusively use public transportation), do not necessarily behave sustainably in another life domain (e.g., renounce from eating animal products) and identifying a factor as impactful in one life domain does not imply it having the same effect in other domains[11,58,88,89]. The respective (sub-) domains were derived from previous studies on GHG emission sectors[86] and guidelines proposed by the German Federal Environmental Office[24] in cooperation with two major non-profit organizations for investigating sustainable behavior in Germany ifeu gGmbH[25] and KlimAktiv gGmbH[26]. In their work, the authors identified important sectors of GHG emissions, presented suggested methods of assessment and footprint calculation principles, which we adapted. The authors further divided shelter related emissions into *electricity* and *heating* related emissions, which we also adapted. Doing so, we adhered to validated measurement strategies, following most of the best practice measurement principles for footprint calculators[90], account for the high region-specificity of individual footprints and emission factors[18,86] and assess drivers and barriers of electricity and heating behavior separately. For three out of the four life domains of interest, we used the official footprint calculation tool issued by the German Federal Environmental Office[24]. For the mobility domain, we calculated individual carbon footprints based on current insights of carbon emission budgeting and emission factors for the mobility domain in Germany. These calculations were performed by the TdLab Geography research group (for more information on the calculation principles, see Supplementary Note 4). Prior to the final calculation, values for each item of the respective calculators were plotted and inspected. Implausible values were identified and removed/recoded by using existing statistics about individual behavior or theoretical maxima of specific items (e.g., >30 liters of fuel consumption for a regular family sedan or the theoretical maximum duration of an inner-European flight). In some cases, we were not able to identify specific cut-off values for items and relied on using boxplot statistics to remove/recode values. After the final calculation, each participant was assigned their designated carbon footprints, represented in $CO_2$ equivalents. The respective mean $CO_2$ equivalents for all life domains of our sample (except from consumption) fell within the expected range of the official governmental report of Germany (Supplementary Fig. 3). Initial preprocessing was done using R programming language, final preprocessing and calculation steps were performed using the Python programming language. For more information on the specific preprocessing steps, calculation principles for each life domain and formulas, see Supplementary Notes 3 and 4.

**Model selection**. To analyze the relative importance of internal and external factors for individual sustainable behavior, we used ML models. Compared to traditional statistical models, a main advantage of ML models is that they are better able to quantify the impact of different internal and external factors in everyday life when many other potentially influential factors are present[23,91,92]. This is because ML models can learn complex associative patterns (e.g., non-linear relationships, higher order interaction effects and interdependence between variables) directly from the data, without the need to specify all potential patterns beforehand[93,94]. This feature is crucial, since manifold relationships between internal and external factors influencing sustainable behavior exist in the real world (as argued before) and the models used to predict sustainable behavior must be able to capture this complexity.

Despite their increasing popularity in environmental and social sciences, ML models have not yet been widely applied to identify factors influencing climate-relevant behavior. Although a few recent studies deployed ML models to analyze influencing factors of emissions, the used models either were not able to account for complex interactions between factors[60], focused on specific intervention effects on GHG emissions[95] or were used to identify overall household emission clusters[23]. In our study, however, we aim to predict individual GHG emissions (in all relevant life domains) through a multitude of internal and external factors and account for their complex interplay. Doing so, we applied different ML models. First, we chose the popular Random Forest ML model due to its ability to approximate any input-output mapping function, ability to cope with small to medium-sized dataset and – on average – higher prediction accuracies on survey data compared to other ML or traditional models[91,94,96]. We also directly compared the RFs performance to two other popular ML models, a linear LASSO regression and non-linear SVM, which delivered promising results in related work on predicting GHG emissions mentioned earlier[60,95]. We compared the ML models' performance to a traditional linear ordinary least squares linear regression model which represents the current practice of predicting climate-relevant behavior. For detailed information about the used ML models and full model results, see Supplementary Methods and Supplementary Discussion.

**Analysis procedure**. All analysis steps were performed using the Python programming language version 3.8, as well as computational libraries such as scikit-learn[97], numpy[98], pandas[99] and scikit-optimize[100]. Prior to the analysis, the complete dataset was shuffled and randomly split into 80% training set and 20% testing set. The final sample sizes for the training and testing set per domain can be found in Supplementary Table 2. This allows us to assess the resulting models' prediction performance on out-of-sample data. We did so because interpreting models that don't generalize well to out-of-sample data (i.e., over-/underfitted models) can lead to biased conclusions regarding the relationships of predictors with the outcome. Focusing on out-of-sample prediction performance is thus crucial to assess how robustly a model captures patterns in the data for a studied population[93]. In turn, focusing on prediction performance on out-of-sample data aids our main study goal of assessing the relative importance of individual factors for sustainable behavior and their complex relationships with each other more accurately.

All internal and external factors (Tables 1–7) were entered as predictors into the ML and LM models with the respective footprints for each life domain as the dependent variables (see Supplementary Note 2 for more detailed information about the predictors). To ensure prediction suitability of the used constructs[101], we examined the scale reliability which revealed good overall internal consistency ($\alpha > 0.70$) with only a few exceptions (see Supplementary Table 3). Continuous predictor variables and all footprint values were standardized (z-transformed) to ensure comparability across life domains. Categorical predictors were dummy coded (i.e., one-hot encoded) for the LM, LASSO and SVM to work with. To simultaneously get a first estimate of the models' prediction capability and find the best hyperparameters for the ML models, we employed a nested 10-fold cross-validation technique (CV) on the training set. This procedure represents the current best practice of evaluating model performance in ML settings, since finding the optimal setting for the model and prediction performance are not mixed[102–104]. Hyperparameters in the ML models during nested CV were tuned using Bayesian Optimization[100] or Grid Search. After the CV, all models were once again fitted on the complete training set and evaluated on the held-out test set to validate the estimated prediction performance from the CV. A graphical depiction and detailed description of the whole analysis procedure can be found in Supplementary Methods and Supplementary Fig. 1. The mean absolute error and explained variance $R^2$ were used as evaluation metrics. The respective model prediction accuracies are depicted in Fig. 1, a tabular version can be found in Supplementary Table 4. To calculate the predictor importance, we used Shapeley Additive explanation (SHAP) values[105], which represent the contribution (i.e., importance) of each predictor to the final model output for each single observation. In the main text, we only focused on predictors with mean SHAP values greater than the mean SHAP values of all predictors. For more information on SHAP values, see Supplementary Methods.

## Data availability

The data set underlying the results presented in this article as well as the raw survey data set are freely accessible on the Open Science Framework at https://doi.org/10.17605/OSF.IO/WNFMB.

## Code availability

The code used to analyze the data is freely accessible on the Open Science Framework at https://doi.org/10.17605/OSF.IO/WNFMB.

Received: 23 March 2023; Accepted: 8 November 2023;
Published online: 20 November 2023

## References

1. Assembly U. G. Transforming Our World: The 2030 Agenda for Sustainable Development. https://www.refworld.org/docid/57b6e3e44.html (2015).
2. Scoones, I. The politics of sustainability and development. *Annu. Rev. Environ. Resour.* **41**, 293–319 (2016).
3. Creutzig F., et al. Demand, services and social aspects of mitigation. in *IPCC, 2022: Climate Change 2022: Mitigation of Climate Change*. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change). https://doi.org/10.1017/9781009157926.007 (2022).
4. Creutzig, F. et al. Towards demand-side solutions for mitigating climate change. *Nat. Clim. Change* **8**, 260–263 (2018).
5. Wiedmann, T., Lenzen, M., Keyßer, L. T. & Steinberger, J. K. Scientists' warning on affluence. *Nat. Commun.* **11**, 3107 (2020).
6. Clayton, S. et al. Psychological research and global climate change. *Nat. Clim. Change* **5**, 640–646 (2015).
7. Black, J. S., Stern, P. C. & Elworth, J. T. Personal and contextual influences on household energy adaptations. *J. Appl. Psychol.* **70**, 3 (1985).
8. Gifford, R., Kormos, C. & McIntyre, A. Behavioral dimensions of climate change: drivers, responses, barriers, and interventions. *Wiley Interdiscip. Rev. Clim. Change* **2**, 801–827 (2011).
9. Guagnano, G. A., Stern, P. C. & Dietz, T. Influences on attitude-behavior relationships. *Environ. Behav.* **27**, 699–718 (1995).
10. Kollmuss, A. & Agyeman, J. Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior? *Environ. Educ. Res.* **8**, 239–260 (2002).
11. Steg, L. & Vlek, C. Encouraging pro-environmental behaviour: an integrative review and research agenda. *J. Environ. Psychol.* **29**, 309–317 (2009).
12. Stern, P. C. New environmental theories: toward a coherent theory of environmentally significant behavior. *J. Soc. Issues* **56**, 407–424 (2000).
13. van Valkengoed, A. M. & Steg, L. Meta-analyses of factors motivating climate change adaptation behaviour. *Nat. Clim. Change* **9**, 158–163 (2019).
14. James G., Witten D., Hastie T., Tibshirani R. *Statistical learning*. Springer (2021).
15. Abou-Zeid, M., Witter, R., Bierlaire, M., Kaufmann, V. & Ben-Akiva, M. Happiness and travel mode switching: findings from a Swiss public transportation experiment. *Transp. Policy* **19**, 93–104 (2012).
16. Hao, Y., Deng, Y., Lu, Z.-N. & Chen, H. Is environmental regulation effective in China? Evidence from city-level panel data. *J. Clean. Produc.* **188**, 966–976 (2018).
17. van Valkengoed A. M., Abrahamse W., Steg L. To select effective interventions for pro-environmental behaviour change, we need to consider determinants of behaviour. *Nat. Human Behav.* **6**, 1482–1492 (2022).
18. Kennedy, E. H., Krahn, H. & Krogman, N. T. Egregious emitters: disproportionality in household carbon footprints. *Environ. Behav.* **46**, 535–555 (2014).
19. Moser, S. & Kleinhückelkotten, S. Good intents, but low impacts: diverging importance of motivational and socioeconomic determinants explaining pro-environmentalbehavior,energy use, and carbon footprint. *Environ. Behav.* **50**, 626–656 (2018).
20. Huddart Kennedy, E., Krahn, H. & Krogman, N. T. Are we counting what counts? A closer look at environmental concern, pro-environmental behaviour, and carbon footprint. *Local Environ.* **20**, 220–236 (2015).
21. Stern, P. C. Contributions of psychology to limiting climate change. *Am. Psychol.* **66**, 303–314 (2011).
22. Hunecke, M., Haustein, S., Grischkat, S. & Böhler, S. Psychological, sociodemographic, and infrastructural factors as determinants of ecological impact caused by mobility behavior. *J. Environ. Psychol.* **27**, 277–292 (2007).
23. Chen, P., Wu, Y., Zhong, H., Long, Y. & Meng, J. Exploring household emission patterns and driving factors in Japan using machine learning methods. *Appl. Energy* **307**, 118251 (2022).
24. Umweltbundesamt. UBA Carbon Calculator. https://uba.co2-rechner.de/en_GB/start#panel-calc (2022).
25. Schächtele K., Hertle H. Die CO2 Bilanz des Bürgers Recherche für ein internetbasiertes Tool zur Erstellung persönlicher CO2 Bilanzen. https://www.umweltbundesamt.de/publikationen/co2-bilanz-des-buergers (2007).
26. Schunkert S., et al. Der UBA-CO2-Rechner für Privatpersonen: Hintergrundinformationen. 44. https://www.umweltbundesamt.de/publikationen/der-uba-co2-rechner-fuer-privatpersonen (2022)
27. Moberg, K. R. et al. Mobility, food and housing: responsibility, individual consumption and demand-side policies in European deep decarbonisation pathways. *Energy Effic.* **12**, 497–519 (2019).
28. Ivanova, D. et al. Mapping the carbon footprint of EU regions. *Environ. Res. Lett.* **12**, 054013 (2017).
29. Wiedenhofer, D., Smetschka, B., Akenji, L., Jalas, M. & Haberl, H. Household time use, carbon footprints, and urban form: a review of the potential contributions of everyday living to the 1.5 C climate target. *Curr. Opin. Environ. Sustain.* **30**, 7–17 (2018).
30. Druckman, A. & Jackson, T. The carbon footprint of UK households 1990–2004: a socio-economically disaggregated, quasi-multi-regional input–output model. *Ecolo. Econ.* **68**, 2066–2077 (2009).
31. Sovacool, B. K. & Brown, M. A. Twelve metropolitan carbon footprints: a preliminary comparative global assessment. *Energy Policy* **38**, 4856–4869 (2010).
32. Blazey, M. A. Travel and retirement status. *Ann. Tour. Res.* **19**, 771–783 (1992).
33. Newbold, K. B., Scott, D. M., Spinney, J. E., Kanaroglou, P. & Páez, A. Travel behavior within Canada's older population: a cohort analysis. *J. Transp. Geogr.* **13**, 340–351 (2005).
34. Siren, A. & Haustein, S. How do baby boomers' mobility patterns change with retirement? *Ageing Soc.* **36**, 988–1007 (2016).
35. Verplanken, B., Aarts, H., Van Knippenberg, A. & Van Knippenberg, C. Attitude versus general habit: antecedents of travel mode choice 1. *J. Appl. Soc. Psychol.* **24**, 285–300 (1994).
36. Lorek, S. & Spangenberg, J. H. Indicators for environmentally sustainable household consumption. *Int. J. Sustain. Dev.* **4**, 101–120 (2001).
37. Lunke, E. B., Fearnley, N. & Aarhaug, J. Public transport competitiveness vs. the car: impact of relative journey time and service attributes. *Res. Transp. Econ.* **90**, 101098 (2021).
38. De Witte, A., Macharis, C. & Mairesse, O. How persuasive is 'free'public transport?: a survey among commuters in the Brussels Capital Region. *Transp. Policy* **15**, 216–224 (2008).
39. Nguyen, H. V., Nguyen, C. H. & Hoang, T. T. B. Green consumption: closing the intention-behavior gap. *Sustain. Dev.* **27**, 118–129 (2019).
40. Vermeir, I. & Verbeke, W. Sustainable food consumption: exploring the consumer "attitude–behavioral intention" gap. *J. Agric. Environ. Ethics* **19**, 169–194 (2006).
41. Tanner, C. & Wölfing Kast, S. Promoting sustainable consumption: determinants of green purchases by Swiss consumers. *Psychol. Mark.* **20**, 883–902 (2003).
42. Clark, M. & Tilman, D. Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. *Environ. Res. Lett.* **12**, 064016 (2017).
43. Rabès, A. et al. Greenhouse gas emissions, energy demand and land use associated with omnivorous, pesco-vegetarian, vegetarian, and vegan diets accounting for farming practices. *Sustain. Prod. Consum.* **22**, 138–146 (2020).
44. Vermeir, I. et al. Environmentally sustainable food consumption: a review and research agenda from a goal-directed perspective. *Front. Psychol.* **11**, 1603 (2020).
45. Borland, S. E., Robinson, S. M., Crozier, S. R. & Inskip, H. M. Stability of dietary patterns in young women over a 2-year period. *Eur. J.Clin. Nutr.* **62**, 119–126 (2008).

46. Rees, J. H. et al. Breaking the habit: on the highly habitualized nature of meat consumption and implementation intentions as one effective way of reducing it. *Basic Appl. Soc. Psychol.* **40**, 136–147 (2018).
47. Hayley, A., Zinkiewicz, L. & Hardiman, K. Values, attitudes, and frequency of meat consumption. Predicting meat-reduced diet in Australians. *Appetite* **84**, 98–106 (2015).
48. Hodson, G. & Earle, M. Conservatism predicts lapses from vegetarian/vegan diets to meat consumption (through lower social justice concerns and social support). *Appetite* **120**, 75–81 (2018).
49. Esposito, C. M., Ceresa, A. & Buoli, M. The association between personality traits and dietary choices: a systematic review. *Adv. Nutr.* **12**, 1149–1159 (2021).
50. Forestell, C. A. & Nezlek, J. B. Vegetarianism, depression, and the five factor model of personality. *Ecol. Food Nutr.* **57**, 246–259 (2018).
51. Lunn, T. E., Nowson, C. A., Worsley, A. & Torres, S. J. Does personality affect dietary intake? *Nutrition* **30**, 403–409 (2014).
52. Aschemann-Witzel, J. & Zielke, S. Can't buy me green? A review of consumer perceptions of and behavior toward the price of organic food. *J. Consum. Aff.* **51**, 211–251 (2017).
53. Springmann, M., Clark, M. A., Rayner, M., Scarborough, P. & Webb, P. The global and regional costs of healthy and sustainable dietary patterns: a modelling study. *Lancet Planet.Health* **5**, e797–e807 (2021).
54. Magnusson, M. K., Arvola, A., Hursti, U. K. K., Åberg, L. & Sjödén, P. O. Attitudes towards organic foods among Swedish consumers. *Br. Food J.* **103**, 209–227 (2001).
55. Reipurth, M. F., Hørby, L., Gregersen, C. G., Bonke, A. & Cueto, F. J. P. Barriers and facilitators towards adopting a more plant-based diet in a sample of Danish consumers. *Food Qual. Prefer.* **73**, 288–292 (2019).
56. Brunsø, K., Scholderer, J. & Grunert, K. G. Testing relationships between values and food-related lifestyle: results from two European countries. *Appetite* **43**, 195–205 (2004).
57. Schösler, H., de Boer, J. & Boersema, J. J. Fostering more sustainable food choices: can self-determination theory help? *Food Qual. Prefer.* **35**, 59–69 (2014).
58. Gatersleben, B., Steg, L. & Vlek, C. Measurement and determinants of environmentally significant consumer behavior. *Environ. Behav.* **34**, 335–362 (2002).
59. Wilson, J., Tyedmers, P. & Spinney, J. E. An exploration of the relationship between socioeconomic and well-being variables and household greenhouse gas emissions. *J. Indus. Ecol.* **17**, 880–891 (2013).
60. Shi, X., Wang, K., Cheong, T. S. & Zhang, H. Prioritizing driving factors of household carbon emissions: an application of the LASSO model with survey data. *Energy Econ.* **92**, 104942 (2020).
61. Goldstein, B., Gounaridis, D. & Newell, J. P. The carbon footprint of household energy use in the United States. *Proc. Natl. Acad. Sci.* **117**, 19122–19130 (2020).
62. Santin, O. G., Itard, L. & Visscher, H. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy Build.* **41**, 1223–1232 (2009).
63. Deetjen, T. A., Walsh, L. & Vaishnav, P. US residential heat pumps: the private economic potential and its emissions, health, and grid impacts. *Environ. Res. Lett.* **16**, 084024 (2021).
64. Pehl, M. et al. Understanding future emissions from low-carbon power systems by integration of life-cycle assessment and integrated energy modelling. *Nat. Energy* **2**, 939–945 (2017).
65. Umweltbundesamt. Indicator: Greenhouse gas emissions. https://www.umweltbundesamt.de/en/data/environmental-indicators/indicator-greenhouse-gas-emissions#at-a-glance (2022).
66. Matthews, H. S., Hendrickson, C. T. & Weber, C. L. The importance of carbon footprint estimation boundaries. *Environ. Sci. Technol.* **42**, 5839–5842 (2008).
67. Verplanken, B. & Roy, D. Empowering interventions to promote sustainable lifestyles: testing the habit discontinuity hypothesis in a field experiment. *J. Environ. Psychol.* **45**, 127–134 (2016).
68. Akenji L., et al. *1.5-degree lifestyles: Towards a fair consumption space for all.* Hot or Cool (2021).
69. Brand-Correa, L. I., Mattioli, G., Lamb, W. F. & Steinberger, J. K. Understanding (and tackling) need satisfier escalation. *Sustain. Sci. Pract. Policy* **16**, 309–325 (2020).
70. Mattioli, G., Roberts, C., Steinberger, J. K. & Brown, A. The political economy of car dependence: a systems of provision approach. *Energy Res. Soc. Sci.* **66**, 101486 (2020).
71. Bamberg, S. The promotion of new behavior by forming an implementation intention: results of a field experiment in the domain of travel mode choice 1. *J. Appl. Soc. Psychol.* **30**, 1903–1922 (2000).
72. Lanzini, P. & Khan, S. A. Shedding light on the psychological and behavioral determinants of travel mode choice: a meta-analysis. *Transp. Res. Part F: Traffic Psychol. Behav.* **48**, 13–27 (2017).
73. De Witte, A. et al. The impact of "free" public transport: the case of Brussels. *Transp. Res. Part A: Policy Pract.* **40**, 671–689 (2006).
74. Steenberghen T., Lannoy P., Macharis C. Impact of free public transport on travel behaviour: a case study. https://www.belspo.be/belspo/organisation/Publ/pub_ostc/CPtrans/rappCP63_en.pdf (2006).
75. Gärling T., Fujii S. Travel behavior modification: Theories, methods, and programs. In: *The expanding sphere of travel behaviour research* (eds R K., T Y., Yamamoto T.). Emerald (2009).
76. Bamberg, S. & Schmidt, P. Incentives, morality, or habit? Predicting students' car use for university routes with the models of Ajzen, Schwartz, and Triandis. *Environ. Behav.* **35**, 264–285 (2003).
77. Wiedemann, A. U., Lippke, S., Reuter, T., Ziegelmann, J. P. & Schwarzer, R. How planning facilitates behaviour change: additive and interactive effects of a randomized controlled trial. *Eur. J. Soc. Psychol.* **41**, 42–51 (2011).
78. Ajzen, I. The theory of planned behavior. *Organiz. Behav. Hum. Decis. Process.* **50**, 179–211 (1991).
79. Fuchs, D. et al. Power: the missing element in sustainable consumption and absolute reductions research and action. *J. Clean. Produc.* **132**, 298–307 (2016).
80. Firth S. K., Lomas K. J., Wright A. Investigating CO2 emission reductions in existing urban housing using a community domestic energy model. In: *Eleventh International IBPSA Conference*) (2009).
81. Pietrapertosa, F., Tancredi, M., Giordano, M., Cosmi, C. & Salvia, M. How to prioritize energy efficiency intervention in municipal public buildings to decrease CO2 emissions? A case study from Italy. *Int. J. Environ. Res. Public Health* **17**, 4434 (2020).
82. Zhang, J. & Smith, K. R. Household air pollution from coal and biomass fuels in China: measurements, health impacts, and interventions. *Environ. Health Perspect.* **115**, 848–855 (2007).
83. European Environment Agency. Greenhouse gas emissions by aggregated sector. https://www.eea.europa.eu/data-and-maps/daviz/ghg-emissions-by-aggregated-sector-5#tab-dashboard-02 (2019).
84. United States Environmental Protection Agency. Sources of Greenhouse Gas Emissions. https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions (2020).
85. Noll, B., Filatova, T., Need, A. & Taberna, A. Contextualizing cross-national patterns in household climate change adaptation. *Nat. Clim. Change* **12**, 30–35 (2022).
86. Ivanova, D. et al. Environmental impact assessment of household consumption. *J. Indus. Ecol.* **20**, 526–536 (2016).
87. Markandya, A. et al. Public health benefits of strategies to reduce greenhouse-gas emissions: low-carbon electricity generation. *Lancet* **374**, 2006–2015 (2009).
88. Green-Demers, I., Pelletier, L. G. & Ménard, S. The impact of behavioural difficulty on the saliency of the association between self-determined motivation and environmental behaviours. *Can. J. Behav. Sci./Revue Can. des sciences du comportement* **29**, 157 (1997).
89. Bratt, C., Stern, P. C., Matthies, E. & Nenseth, V. Home, car use, and vacation: the structure of environmentally significant individual behavior. *Environ. Behav.* **47**, 436–473 (2015).
90. Birnik, A. An evidence-based assessment of online carbon calculators. *Int. J. Greenhouse Gas Control* **17**, 280–293 (2013).
91. Kern, C., Klausch, T. & Kreuter, F. Tree-based machine learning methods for survey research. *Surv. Res. Methods* **13**, 73–93 (2019).
92. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learning Res.* **3**, 1157–1182 (2003).
93. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
94. Breiman, L. Random forests. *Mach. learning* **45**, 5–32 (2001).
95. Wang, A., Lam, J. C., Song, S., Li, V. O. & Guo, P. Can smart energy information interventions help householders save electricity? a svr machine learning approach. *Environ. Sci. Policy* **112**, 381–393 (2020).
96. Liaw, A. & Wiener, M. Classification and regression by random Forest. *R news* **2**, 18–22 https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf (2002).
97. Pedregosa F. et al. Scikit-learn: Machine learning in Python. In: *Journal of Machine Learning Research*) (2011).
98. Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).
99. McKinney, W. pandas: a foundational Python library for data analysis and statistics. *Python Perform. Sci. Comput.* **14**, 1–9 (2011).
100. Head T., Kumar M., Nahrstaedt H., Louppe G., Shcherbatyi I. Scikit-optimize/scikit-optimize. *Zenodo*, 4014775 (2021).
101. Jacobucci, R. & Grimm, K. J. Machine learning and psychological research: the unexplored effect of measurement. *Perspect. Psychol. Sci.* **15**, 809–816 (2020).
102. He, J. & Chalise, P. Nested and repeated cross validation for classification model with high-dimensional data. *Revista Colombiana de Estadística* **43**, 103–125 (2020).
103. Krstajic, D., Buturovic, L. J., Leahy, D. E. & Thomas, S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminform* **6**, 10–10 (2014).

104. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **7**, 1–8 (2006).
105. Lundberg S. M., Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems* **30**, https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (2017).

## Acknowledgements
This research was funded by the Federal Ministry of Education and Research (BMBF) and the Ministry of Science Baden-Württemberg within the framework of the Excellence Strategy of the Federal and State Governments of Germany. For the publication fee we acknowledge financial support by German Research Foundation (Deutsche Forschungsgemeinschaft - DFG) within the funding program “Open Access Publications” and Heidelberg University.

## Author contributions
N.B., K.A. and J.R. designed the study. N.B. conducted data collection and analyzed the main results. N.B. wrote the first draft of the article and all other authors provided critical feedback. N.A. contributed to calculating mobility footprints. All authors endorsed the final version of the article.

## Funding
Open Access funding enabled and organized by Projekt DEAL.

## Competing interests
The authors declare no competing interests.

## Ethics approval
This research was approved by the Faculty of Behavioral and Cultural Studies Ethics Committee at Heidelberg University on March 9th 2021.

## Additional information
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s43247-023-01092-x.

**Correspondence** and requests for materials should be addressed to Nils Brandenstein.

**Peer review information** *Communications Earth & Environment* thanks Peipei Chen and Marta Baltruszewicz for their contribution to the peer review of this work. Primary Handling Editors: Niheer Dasandi and Clare Davis, Aliénor Lavergne. A peer review file is available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Appendix C1 - Manuscript III

Manuscript III: To follow or not to follow: Estimating political opinion from Twitter data using a network-based machine learning approach

Note:
This is the final version of the open-access article published in *Social Science Computer Review*. The reproduction of this article in this dissertation is permitted under the respective Creative Commons Attribution 4.0 International License.

Brandenstein, N., Montag, C., & Sindermann, C. (2024). To follow or not to follow: Estimating political opinion from Twitter data using a network-based machine learning approach. *Social Science Computer Review*, *0*(0). https://doi.org/10.1177/08944393241279418

Original Manuscript

Social Science Computer Review
2024, Vol. 0(0) 1–22

DOI: 10.1177/08944393241279418
journals.sagepub.com/home/ssc
Sage

# To Follow or Not to Follow: Estimating Political Opinion From Twitter Data Using a Network-Based Machine Learning Approach

**Nils Brandenstein[1], Christian Montag[2], and Cornelia Sindermann[3]**

## Abstract

Studying political opinions of citizens stands as a fundamental pursuit for both policymakers and researchers. While traditional surveys remain the primary method to investigate individual political opinions, the advent of social media data (SMD) offers novel prospects. However, the number of studies using SMD to extract individuals' political opinions are limited and differ greatly in their methodological approaches and levels of success. Recent studies highlight the benefits of analyzing individuals' social media network structure to estimate political opinions. Nevertheless, current methodologies exhibit limitations, including the use of simplistic linear models and a predominant focus on samples from the United States. Addressing these issues, we employ an unsupervised Variational Autoencoder (VAE) machine learning model to extract individual opinion estimates from SMD of $N$ = 276 008 German Twitter (now called 'X') users, compare its performance to a linear model and validate model estimates on self-reported opinion measures. Our findings suggest that the VAE captures Twitter users' network structure more precisely, leading to higher accuracy in following decision predictions and associations with self-reported political ideology and voting intentions. Our study emphasizes the need for advanced analytical approaches capable to capture complex relationships in social media networks when studying political opinion, at least in non-US contexts.

## Keywords

social media, political opinion, estimation, network structure, machine learning

[1]Heidelberg University, Germany
[2]Ulm University, Germany
[3]University of Stuttgart, Germany

**Corresponding Author:**
Nils Brandenstein, Institute of Psychology, Heidelberg University, Hauptstr. 47-51, Heidelberg 69117, Germany.
Email: nils.brandenstein@psychologie.uni-heidelberg.de

## Introduction

Monitoring the political opinions of the general public is a valuable asset in today's world. Tracking individuals' political views and attitudes enables policymakers to develop tailored regulatory measures and supports researchers in studying political trends and factors that shape political opinions and behavior (Dong & Lian, 2021; Schober et al., 2016). Today, various methods exist to study public- and individual-level political opinions. Although self-reports integrated into surveys still dominate the opinion research space (Berinsky, 2017), the utilization of publicly available social media data (SMD) has gained extensive popularity in recent years (Rousidis et al., 2020). Arguably, SMD offers key benefits: Spontaneous expressions of opinions can be captured, accessed and analyzed in real-time without the limitations of predefined response options (e.g., Reveilhac et al., 2022; Schober et al., 2016). Further, the pool of available participants is high with a total of 4.8 billion unique user identities and 400 million active users worldwide across the major platforms as of July 2023 (Kemp, 2023). Although SMD represents a promising tool in social science, the number of studies using SMD to study sociopolitical issues—although increasing—remains relatively low compared to other data sources like self-reports (c.f. Dong & Lian, 2021), especially in countries other than the US. Additionally, existing studies differ greatly in their overall study aim, methodologies and study characteristics (Dong & Lian, 2021). The existing work extracting information about political opinions (e.g., attitudes and ideology) and behavior of users from SMD on an individual level can be broadly categorized based on what type of SMD is analyzed. Most of the work either deployed text analysis or network analysis (or in rare cases a combination of both).

Text analysis aims to transform textual (social media) content into quantitative data using (automated) lexical analyses (Schober et al., 2016). A variety of methods exists within text analysis (for an overview see Eichstaedt et al., 2021), all relying on the theoretical notion that individuals freely share their thoughts and opinions (e.g., in posts) and that derived estimates are based on psycho-linguistic properties of the texts (e.g., Kumar & Sebastian, 2012). A popular text analysis method to analyze SMD is sentiment analysis. Here, researchers first cluster semantically similar words within a text (e.g., social media posts) into categories or contexts which are then used to further categorize the content into positive, negative or neutral (and infrequently more detailed) sentiments. The contextualization of words in these models to extract sentiments is done using dictionary-based approaches, word embedding models or (nowadays increasingly popular) transformer-based models (Widmann & Wich, 2023). Finally, the estimated sentiment of posts is then used to predict sociopolitical outcomes (see Skoric et al., 2020, for an overview of recent studies). Notably, not all studies on SMD explicitly model sentiments. Some works also use machine learning models (ML) to directly use the contextualized words to predict sociopolitical outcomes (Skoric et al., 2020). Although using text analysis—specifically sentiment analysis—has yielded promising results in the past (e.g., Chung & Zeng, 2016; Lansdall-Welfare et al., 2012; Tumasjan et al., 2010), recent meta-analyses and reviews have revealed significant divergences in their prediction accuracies of sociopolitical outcomes (e.g., Rousidis et al., 2020; Skoric et al., 2020). This may not only be due to large differences in the analyzed political outcomes, data sources and political systems but also due to the model types used, with ML models yielding the highest prediction performance of political opinions and sociopolitical outcomes (Skoric et al., 2020).

Further, recent articles also showed that so-called network analysis approaches often outperformed text analysis approaches in estimating (or predicting) political opinions, with the highest accuracy observed when using both approaches in combination (e.g., Skoric et al., 2020). Based on these results and as argued by others, network analysis presents a promising way to study political opinions and sociopolitical outcomes using SMD (e.g., Gayo-Avello, 2013; Kwak & Cho, 2018; Livne et al., 2021; Pallavicini et al., 2017), which is why we will focus on this approach in our work. Network approaches use the connections (i.e., follows) and interactions (i.e., sharing, likes and comments) between users, so-called relational edges, to extract information about individuals' political opinions and behavior. Such approaches rely on the theoretical notion that interactions and following decisions of users represent signals of their political opinions (e.g., attitudes and ideology) (e.g., Barberá et al., 2015). In line with that, research has shown that individuals tend to be more likely to connect with individuals who are similar to them, for instance, those who share their (political) opinions (McPherson et al., 2001), a phenomenon known as homophily. Especially on social media, such a tendency can be strengthened by friend recommenders and other algorithms (Aral, 2020). Although the homophily assumption describes a general tendency and does not imply that all individuals only connect with like-minded others, most empirical studies have shown that—much like in the real world—like-minded users (e.g., similar opinions) are more often than not connected on social media and, consequently, are also more likely to form homogeneous opinion networks (e.g., Aiello et al., 2012; Cinelli et al., 2021; Figeac & Favre, 2023; Khanam et al., 2023; Lee & Brusilovsky, 2010; McPherson et al., 2001; Pallavicini et al., 2017). Building on the homophily assumption, we also aim to estimate individuals' political opinions from their social networks.

Although manifold approaches and model classes exist to analyze social networks, an essential method is dimensionality reduction (DR). DR presents a family of tools to condense the network structure (i.e., user and their connections) to create a more compact and informative representation of the network (Chikhi et al., 2007; Nishana & Surendran, 2013). In previous research, DR has been used to enable network models to work with high-dimensional SMD through creating embeddings (e.g., Chikhi et al., 2007; Grover & Leskovec, 2016; Yan et al., 2007) or was indirectly built into (advanced) models (e.g., layer transformations & pooling in Graph Convolutional Neural Networks, Bronstein et al., 2016; Zhang et al., 2019). However, DR methods have also been used standalone in network analysis for user clustering and community detection (e.g., Al-Omairi et al., 2021; Zarzour et al., 2018), uncovering hidden structures (e.g., Chikhi et al., 2007) or extracting global network characteristics like religious beliefs (e.g., Kurucz et al., 2008). Notably, and relevant to our approach, many studies on estimating (individual-level) political opinions from SMD have also used standalone DR models. For instance, Barberá et al. (2015) successfully used DR to analyze ideological segregation and cross-ideological communication in social networks. Wojcieszak et al. (2022) used DR to analyze ideological congruency in the interaction of social media users with politicians and news organizations, while Barberá (2015a) and Bond and Messing (2015) used DR to estimate political ideology of Facebook and Twitter users (renamed to "X"; we use the name "Twitter" throughout the work like it was called during data collection) which correlated with self-reported political opinions and was predictive of voting turnout.

Although studies using standalone DR have yielded promising results, they exhibit a few limitations that we will address in the present study. First, the DR techniques used so far to extract individual-level political opinions from SMD (e.g., ideology, beliefs and attitudes) tend to be relatively simplistic. That is, studies mostly used models such as Correspondence Analysis (CA) or linear latent space models (e.g., Barberá, 2015b; Barberá et al., 2015; Eady et al., 2019; Tausanovitch & Warshaw, 2017). However, these models focus on linear relationships in the data and thus are likely to perform badly in instances when more complex, non-linear relationships exist (e.g., De Backer et al., 1998; Nanga et al., 2021). We argue that similar to SMD sentiment analysis, where more complex ML models increased estimation performance (Anjaria & Guddeti, 2014; Skoric et al., 2020), network-based analysis of SMD would also benefit from using complex ML models. This is because individual political opinions (such as ideology) are complex phenomena and a linear combination of factors (i.e., followed accounts) seems unlikely to explain this manifoldness. Social networks—including those on platforms like Twitter—can be understood as complex systems (Boccaletti et al., 2006; Tunstel et al., 2021). As such, the different nodes (i.e., Twitter accounts) in a social network may interact, communicate, provide feedback to each other, and adjust accordingly. Looking at social networks from the angle of cybernetics, causal feedback loops continuously impact nodes in their thinking, attitudes, and behaviors (Tilak et al., 2023). In such networks, next to linear and non-linear effects, also synergetic effects can occur between nodes. Therefore, posts, likes, shares, etc. from various accounts a user follows, might exhibit synergetic effects that only occur at a specific constellation and/or complex non-linear, non-additive effects might shape the opinions of the user, including ideological views. Such effects, however, are unlikely to be caught by simplistic linear models. Complex ML models therefore seem more fitting to examine complex network structures and ultimately individuals' opinions (Silva & Zhao, 2016). Notably, however, they have not been widely applied to extract individual-level public opinions from SMD.

Further, most studies so far have been conducted in the US political context (e.g., Barberá et al., 2015; Bond & Messing, 2015; Brito et al., 2021; Eady et al., 2019). It is questionable if their results would translate to other parts of the world, since estimating individual political opinions of users in bipartisan political systems (such as the US) seems much simpler than in more complex, multiparty systems like most European countries (Traber et al., 2023; Wagner, 2021). Further, previous studies analyzing cross-country network compositions revealed that European Twitter users usually have more ideologically heterogeneous networks than US users, making it potentially harder to pinpoint their political opinion from their network structure (Barberá, 2015b). To our knowledge, only one study by Barberá (2015b) explicitly estimated individuals' political opinions from social media networks in countries other than the US. However, validation of the opinion estimates was limited to the correlation of political elites (i.e., followed political accounts by individual users) with expert ratings. Therefore, the feasibility and accuracy of estimating individuals' political opinions from social media networks in non-US contexts is still understudied.

In the current study, we aim to address the mentioned limitations. Our primary research objective is to examine if general political opinions of non-US social media users can be estimated based solely on their Twitter network structure. Further, we aim to test if a complex, non-linear DR model outperforms a widely used linear DR model in extracting these opinions. Specifically, we employ an unsupervised Variational Autoencoder (VAE) ML model and a linear CA. Thereby, we obtain a point estimate for each social media account in latent space which we assume to represent their general political opinion. To test the capability of both models to estimate individuals' political opinions, we first compare their DR performance on the same dataset and inspect the distribution of point estimates to validate how closely they resemble expected opinion distributions (from left-leaning to right-leaning). In the final validation step, we correlate the models' point estimates with self-reported opinion data (symbolic ideology and voting intention) for a subsample of users from a dataset surveyed in 2021.

## Methods

### *Data Collection and Pre-Processing*

To facilitate individual-level opinion estimation, we analyzed the connections of regular, politically interested Twitter users with political accounts which present a clear political stance. For the most part, we followed the approach of Barberá et al. (2015). First, we identified the Twitter accounts of German politicians and their followers. To accomplish this, we created a list of

German politicians in office in all sixteen federal states or the national parliament and matched each politician with their Twitter account. A detailed description of this procedure can be found in the supplemental materials, I.I. In total, we retrieved $N = 1434$ active Twitter accounts of German politicians. Next, using their Twitter user IDs, we queried the complete follower list of each account. Data collection took place in July 2022. In the following, we refer to the retrieved followers of political accounts as users. Similar to previous studies and to improve data quality (e.g., Gayo-Avello, 2013; Kwak & Cho, 2018), we pre-processed the user lists by removing politicians following each other and user accounts created after our survey data collection (see supplemental materials, I.I for full procedure). After the initial pre-processing, the list of users following political accounts resulted in $N = 13\ 306\ 769$ unique users. Next, analogous to Barberá et al. (2015), we only kept users, who followed at least 10 political accounts to reduce the number of inactive accounts and focus on politically interested users. The final user count after these steps resulted in $N = 276\ 008$ unique users. Although this final list only encompassed a subset of all followers obtained in the first step, we expected this reduction based on similar following rates of political accounts (number of users following ten or more political accounts) in social media networks reported in previous research (e.g., Barberá et al., 2015; Wojcieszak et al., 2022).

Analogous to Barberá et al. (2015), we considered the political opinion of a user as a position (i.e., point estimate) in a multidimensional, latent space which can be obtained from the following structure of Twitter users with the respective political accounts. For our models to calculate these positions, we first arranged users and political accounts in an N × m adjacency matrix, representing the following structure of each Twitter user $i \in \{1, ..., n\}$ (row) for a target political account $j \in \{1, ..., m\}$ (column), with $Y_{ij} = 1$ indicating a following decision and $Y_{ij} = 0$, otherwise. Doing so, our adjacency matrix represented a bipartite, directional graph structure (see Tabassum et al., 2018) with initial network matrix dimensions of $Y = [276\ 008 \times 1434]$. After a further check of this following matrix, we removed seven additional political accounts that were either no longer followed by any user or were mistakenly identified as accounts of politicians previously, yielding the final following matrix $Y = [276\ 008 \times 1427]$. To validate our models on out-of-sample data, we divided the final matrix into a separate training and test set (90% train and 10% test), resulting in $Y_{train} = [248\ 407 \times 1427]$, $Y_{test} = [27601 \times 1427]$ matrices. Further information about the adjacency matrix and dataset splitting procedure can be found in the supplemental materials, subsection I.I.

For our final model checks and the relation of point estimates to self-reported political opinion and behavioral intentions, we used survey data (collected between December 2020 and February 2021) from a study conducted by one of the authors. In total, $N = 780$ individuals reported on different personality and political opinion items. Relevant to the current study, this survey assessed self-reported symbolic political ideology via the one-item left-right self-placement and voting intentions. Participants were also asked to provide their Twitter user name voluntarily (if they had an account). Of all participants in the study, $N = 173$ provided a (valid) user name. In the present study, we used the self-report data of these participants to validate our models and refer to it as the self-report dataset in the following. To create this dataset, we first used the reported Twitter user names and queried the Twitter API to retrieve their user IDs. In total, $N = 163$ unique accounts could be retrieved. Then, we matched user IDs to the IDs in the follower lists of political accounts acquired previously. Since one of our intended models (CA) only works if an individual user follows at least one political account (see Model Selection and Description), we removed all users not following any political account (i.e., all zero values). After applying these procedures, the self-report dataset matrix $Y_{self\text{-}report} = 119 \times 1427$ was obtained. Further information about the self-report dataset (including deployed scales, study procedure and ethics approval) can be found in the supplemental materials, subsection I.III and in (Sindermann et al., 2021, 2022, 2023).

### *Model Selection and Description*

As mentioned in the Introduction, manifold network analysis approaches could be applied to our research questions. However, based on our data structure (bipartite, directed graph) and similar to many previous studies on individual-level political opinion (e.g., Barberá, 2015a; Barberá et al., 2015; Bond & Messing, 2015; Eady et al., 2019; Kurucz et al., 2008; Wojcieszak et al., 2022), we used standalone DR models. Specifically, to test whether individual political opinions of German Twitter users can be inferred solely from their network structure and whether a more complex ML model would outperform a simple linear model, we deployed two DR models: Correspondence Analysis (CA) and Variational Autoencoder (VAE).

CA is conceptually related to Principal Component Analysis and can be used to analyze relationships between multiple categorical variables. Specifically, CA uses linear combinations (i.e., transformations) of the original input data to project the rows and columns onto a new, lower-dimensional subspace. Similar to previous studies (e.g., Barberá et al., 2015; Eady et al., 2019), we used CA to reduce the full following matrix Y to a lower-dimensional subspace. Afterward, we analyzed its DR capabilities by reversing this transformation to reconstruct (i.e., approximate) the input data from this subspace and compared it to the original input data. We further checked if projected subspace row coordinates represented Twitter users' opinion point estimates and if column coordinates represent the overall political positioning of the respective political accounts. The CA was set to reduce the following matrix to two latent dimensions. This was done for several reasons. First, we were interested in obtaining a single point estimate on one latent dimension for each user that is expected to represent their political opinion. In a CA, the first dimension extracted incorporates the highest eigenvalue (i.e., the highest amount of variance captured) of all dimensions, which we assumed to represent overall political opinion (see Barberá et al., 2015, for a similar approach and in-depth description). Second, fitting a CA on large datasets can be computationally expensive. Thus, mapping inputs to fewer dimensions reduces the computation time substantially (e.g., Halko et al., 2011). Lastly, using two dimensions simplifies data visualization by creating a 2D plot of the point estimates, even though we were only interested in the first CA latent dimension. A more detailed description of the CA and its working principles can be found in the supplemental materials, subsection I.II and in Barberá et al. (2015).

Similar to CA, VAE also presents a DR model. However, in contrast to CA, VAEs compress data with greater complexity using non-linear and probabilistic mappings. VAEs are unsupervised ML models based on neural-network autoencoders. They comprise an encoder network for projecting data into a latent distribution and a decoder network for reconstructing the original input from the latent codes. The encoder typically consists of multiple, fully connected layers with the last layer approximating the latent representations (posterior) using a multivariate Gaussian distribution from which the individual latent codes of users and political accounts (much like the projected row and column coordinates in the CA) can be sampled. The decoder part of the network then reconstructs (i.e., approximates) the original input matrix from the sampled latent codes. From an intuitive standpoint, VAEs learn to represent the essence of a dataset (i.e., network structure) in a condensed and structured way. Using a probabilistic, non-linear, neural network-based DR approach, their encoding of users and political accounts should lead to a more nuanced representation and differentiation of political stances in a multiparty system compared to a linear, deterministic DR model like CA. Although adaptations of VAE for network-based data (geometric ML) exist (Variational Graph Autoencoder, see Kipf & Welling, 2016), we used a regular VAE which does not explicitly preserve relational structures between users but only between users and political accounts. This decision was made based on our data structure and the overall goal of estimating users' opinions from their decisions to (not) follow political accounts analogous to previous studies. To enable model comparison with the CA, we mapped inputs to two latent dimensions in the VAE as well. We assumed that one dimension would represent individuals' and politicians' overall political opinions. The VAE used in our study comprised three hidden, fully connected neural layers for the encoder and decoder. The general model architecture is depicted in Figure 1, a detailed description of the VAE (latent code extraction, differences to CA, training procedure and hyperparameter setting) can be found in the supplemental materials, subsection I.II.

## *Analysis Procedure*

Both the CA and VAE were first trained on the training dataset. For an initial check of how accurately the models compressed the original network (DR performance), we reconstructed the following matrix from the calculated latent dimensions and compared it with the original matrix of the training dataset. Since both the original and reconstructed matrix contained binary data (following status), we used typical classification performance metrics to check model performance. In detail, we calculated Precision, Recall, F1 score, Matthew's correlation coefficient (MCC) and Balanced Accuracy (BACC). The latter two are particularly well-suited metrics for sparse, imbalanced datasets, common in many network analysis applications (Chen et al., 2024). Imbalance also applies to our study's data, since only ∼1.7% of all values in the training and test datasets represented "follows" (i.e., ones). Despite not opting for a data resampling technique before model fitting to balance the distribution, MCC and BACC still allowed us to evaluate the models' performance and capability to represent the minority class appropriately. Also, we assumed follow decisions to be equally important/signaling as non-follow decisions to estimate individual political opinions. We were thus interested in the classification performance for both follow and non-follows, which are better represented by MCC and BACC compared to the aforementioned metrics. Additionally, we calculated the overall proportion of correctly predicted cases (PCP) and the Brier score to enable performance comparison with previous studies (see

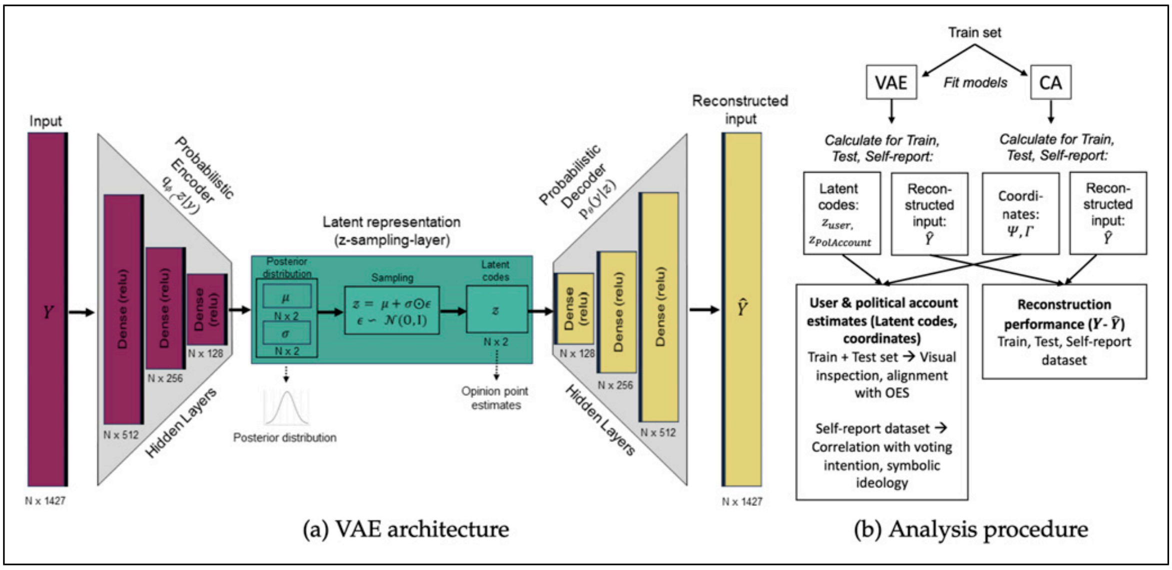



**Figure 1.** VAE architecture and analysis procedure. Note: (a) Depiction of the general VAE architecture. Left part (red bars) shows the input and encoder model. Right part (yellow bars) shows the decoder model. (Hidden) layer sizes (data shape) of the VAE are depicted under the respective bars. Green box depicts the sampling layer, in which the user and political account point estimates are extracted (sampled) from the two latent dimensions. (b) Study analysis procedure and workflow for VAE and CA. Detailed information about the VAE model and latent representations can be found in the supplementary material, subsection I.I and I.II.

supplemental materials, subsection II.II). As an uninformed classification baseline and to benchmark both the CA and VAE, we further calculated a Naive Classifier (NCF), which always predicts the positive majority class ("no follow"). After training and evaluating the models on the training set via the reconstruction approach, we applied and evaluated the trained models on the test set using the same classification performance metrics.

Next, we visually inspected our models' point estimates and latent dimensions. First, we inspected the overall distribution of user point estimates on the two latent dimensions. As mentioned, the latent dimensions in the models should represent a condensed version of the overall structure in the original network matrix. Therefore, we expected one of the two dimensions in our models to resemble the general political opinions of individual users. In doing so, we plotted the point estimates of each user (training and test set) on the two latent dimensions and highlighted them based on the percentage of political accounts from a respective party the user followed in relation to the total number of followed accounts. To do so, we first matched all political accounts in our dataset to their respective party membership for all parties that were part of the German national parliament at the time of the study: Parties on the left side of the political spectrum: Alliance '90/Greens (Green party), Social Democratic Party (SPD) and The Left (Linke); and parties on the right side: Alternative for Germany (AfD), Christian Democratic Union/Christian Social Union (CDU/CSU), Free Democratic Party (FDP). The left-right categorization was based on the Open Expert Survey 2021 (OES), in which experts rated each party for their political stance on different political dimensions and scales (Jankowski et al., 2022). We used data from this work instead of the often used Manifesto report (Lehmann et al., 2023) because the OES surveyed considerably more experts, potentially leading to more robust ratings of parties' political positions. As described, we expected that the more accounts from a specific party the users follow (percentage-wise), the more they are likely to hold similar opinions to the party based on the homophily assumption (McPherson et al., 2001), which was in detail introduced before and has been investigated and supported by previous research (e.g., Aiello et al., 2012; Barberá, 2015a; Barberá et al., 2015; Bond & Messing, 2015; Cinelli et al., 2021; Lee & Brusilovsky, 2010; McPherson et al., 2001; Pallavicini et al., 2017; Wojcieszak et al., 2022). Therefore, individuals following a relatively high number of politicians from left-leaning parties (Linke, Green party, SPD) were expected to be on the opposite spectrum compared to individuals following right-leaning parties (AfD, CDU/CSU, FDP). The total and average counts of followed accounts per party and dataset as well as the final vote shares of the German federal election in 2021 (Wilko, 2021) are depicted in Table 1. Vote shares were added to enable a comparison of the party following distribution in our datasets against the respective party popularity in the German population.

Afterward, we checked if the overall estimated political positioning of parties corresponded to their overall political positioning judged by experts. To this end, we again used data from the 2021 OES (Jankowski et al., 2022). Specifically, we used expert ratings positioning each party on a typically used left-right ideology scale ranging from 0–20. To check alignment, we plotted these scores against the median party point estimates of both the CA and VAE from our models. In detail, similar to Barberá et al. (2015), we collated the model column coordinates of political accounts from the same party on the first dimension (matching individual political accounts to their party) and used the median of all these column coordinates as a measure for the models' overall party positioning. Further information about the extraction of column coordinates can be found in the supplementary materials, subsection I.II. All estimates were standardized to ensure comparability of party positionings in the models and the scores of the OES report. As a final check of our models capturing political opinions, we compared user point estimates with their self-report data in our self-report data set. After applying our trained models on the self-report dataset to get users' point estimates, we first calculated Spearman correlations between users' point

**Table 1.** Overview of party accounts followed per dataset.

| | Dataset | | | | | | |
|---|---|---|---|---|---|---|---|
| | Train | | Test | | SR | | Federal election |
| Party | % follows | *Mdn (IQR)* | % follows | *Mdn (IQR)* | % follows | *Mdn (IQR)* | % votes |
| Linke | 11.3 | 2 (1, 4) | 11.4 | 2 (1, 4) | 10.0 | 2 (1, 5) | 4.9 |
| SPD | 27.0 | 5 (1, 8) | 26.9 | 5 (1, 8) | 33.3 | 3 (2, 5) | 25.7 |
| Greens | 24.9 | 4 (1, 7) | 24.9 | 4 (1, 7) | 34.9 | 2 (1.5, 5) | 14.8 |
| FDP | 11.0 | 2 (1, 3) | 11.0 | 2 (1, 3) | 12.0 | 2 (1, 5) | 11.5 |
| CDU/CSU | 19.3 | 4 (1, 6) | 19.1 | 4 (1, 6) | 7.8 | 2 (1, 3) | 24.1 |
| AfD | 6.5 | 3 (1, 9) | 6.7 | 3 (1, 9) | 2.0 | 3.5 (2, 10.3) | 10.3 |

Note. % follows = Percentage of follows relative to all followed party accounts, excluding non-follows. Mdn = Median number of party accounts followed calculated for all users following at least one of the respective party accounts. IQR = Interquartile range of 25% and 75%. % votes indicate overall party vote shares in the 2021 German federal election. Train: Training dataset, Test = Test dataset, SR = Self-report dataset. Values are rounded to one decimal. Sample size per dataset: $N_{train}$ = 248 407, $N_{test}$ = 27601, $N_{self\text{-}report}$= 119, total count of accounts followed per dataset: $N_{train}$ = 5 801 440, $N_{test}$ = 643 563, $N_{self\text{-}report}$ = 1759.

estimates and their self-reported political (symbolic) ideology assessed on a scale ranging from 1 (left) to 10 (right). We expected a positive relationship between point estimates and symbolic ideology. Next, we plotted users' point estimates conditioned on their reported party voting intention in the next German Federal election (Sonntagsfrage). Analogous to self-reported ideology, we expected a gradient in overall point estimates. That is, median point estimates of users signaling to vote for more left-leaning parties should be lower than users intending to vote for more right-leaning parties. Further, we used the opinion point estimates to predict self-reported voting intention using a multinomial logistic regression model (see supplemental materials, subsection I.III). However, due to the highly unequal cell sizes with much less participants indicating to vote for more right-leaning parties (see section "Opinion Estimate Validation Using Self-Report Data"), we focus on the visual inspection of point estimate distributions in the main text. Finally, we calculated the respective classification metrics of the reversed (i.e., reconstructed) matrices for the self-report set (see supplemental materials, subsection II.I). Figure 1 provides a visual depiction of the full analysis procedure and workflow.

## Results

### *Reconstruction Performance*

To check the models' capacity to represent users' network structure (i.e., following decisions) in a latent space, we inspected the adjacency matrix reconstruction performance. The overall results are depicted in Table 2. Across all calculated metrics and the training and test datasets, the VAE consistently achieved a higher reconstruction performance than both the CA and the NCF. The VAE managed to classify the majority of the positive class (non-follows) more accurately (higher Recall) and misclassified less of the following decisions (i.e., negative class) as non-follows (higher Precision). However, since both datasets were highly imbalanced (~98.3% of all values are in the positive, non-follow class), the NCF seems to (almost) match the performance of the VAE on the F1 score, Precision and Recall metrics and even outperformed the CA on the former two. Therefore, we focused on comparing model performance on the MCC and BACC which evenly take the classification performance for both classes into account. Again, the VAE showed

**Table 2.** Model reconstruction performance.

| | Metrics | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | BACC | | MCC | | F1 score | | Precision | | Recall | |
| Models | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test |
| VAE | **0.69** | **0.69** | **0.52** | **0.52** | **0.99** | **0.99** | **0.99** | **0.99** | **1.00** | **1.00** |
| CA | 0.52 | 0.52 | 0.12 | 0.12 | 0.79 | 0.79 | 0.66 | 0.65 | 0.99 | 0.99 |
| NCF | 0.50 | 0.50 | 0.00 | 0.00 | **0.99** | **0.99** | 0.98 | 0.98 | **1.00** | **1.00** |

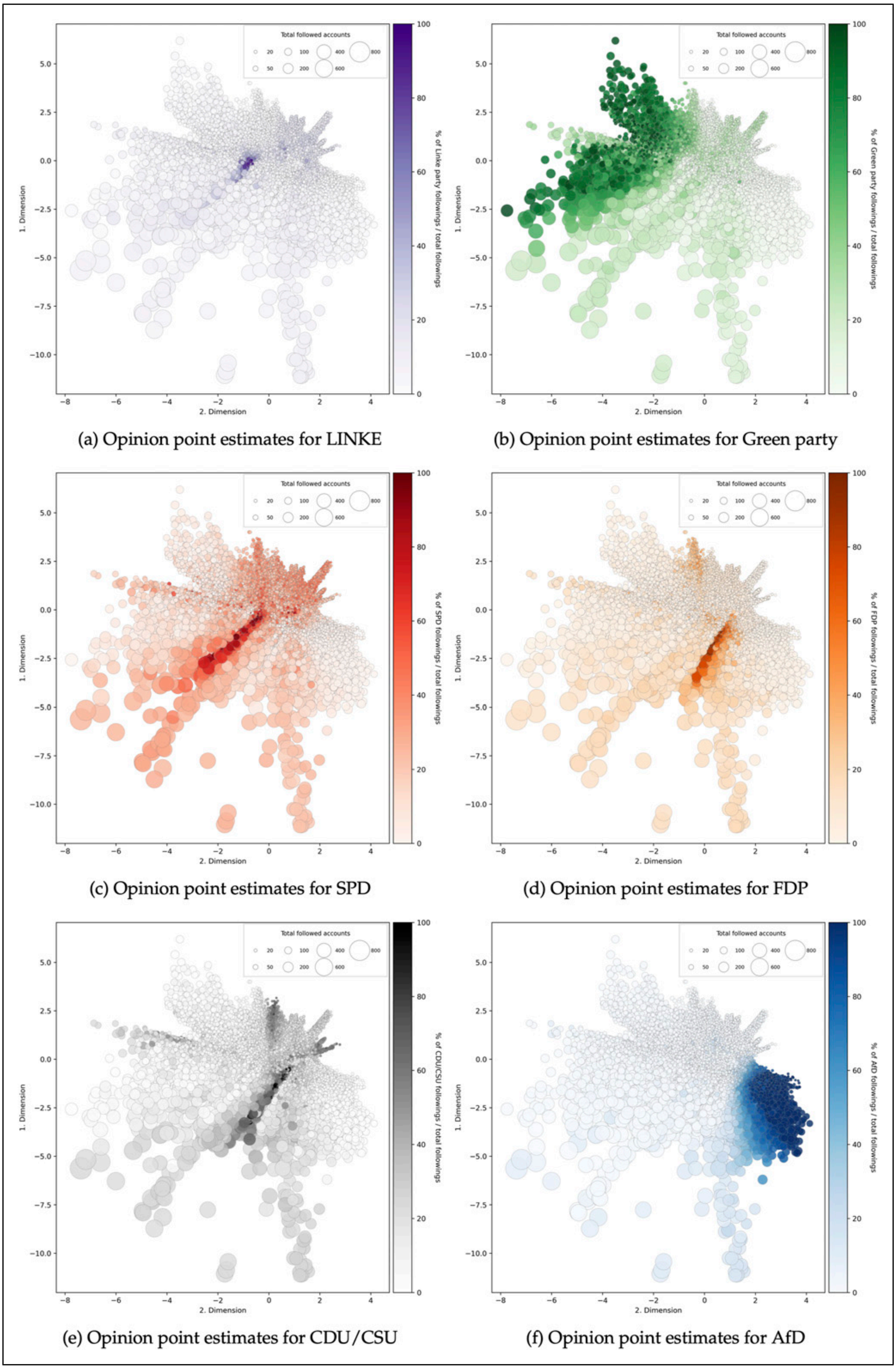
Note. Performance rounded to two decimals. NCF: Naive classifier (always predicting majority class, "No follow"). Positive class: "No follow," negative class: "Follow." BACC: Balanced accuracy, MCC: Matthew's correlation coefficient. Train: Training dataset, Test: Test dataset. Preferred model for each metric and dataset in bold.

an overall higher classification performance for the follow and non-follow decisions (higher BACC and MCC), thus reconstructing the following matrix more accurately than the other two models. Although performing worse than the VAE and the NCF on unbalanced metrics, the CA outperformed the NCF on the balanced BACC and MCC metrics. A similar results pattern emerged for the additionally calculated Brier score and PCP (see supplemental materials, subsection II.II).

## Inspection of Opinion Estimates

Next, we visually inspected the latent dimensions and user opinion point estimates. Since the VAE outperformed the CA regarding matrix reconstruction performance, we only show the point estimates for this model in the following (see supplemental materials, subsection II.III for CA point estimates). Figure 2 shows individual user point estimates from the VAE on the two latent dimensions plotted separately for each party. As expected, individuals following a proportionally high number of left-wing parties (Green party, Linke and SPD) appear to have low values on the second latent dimension and vice versa for followers of right-wing parties (AfD, CDU/CSU, FDP). This finding indicates that the second latent dimension in the VAE symbolizes individuals' general political opinions. The first latent dimension in the VAE seems to represent the magnitude of users' following decisions. In detail, the more political accounts users followed overall (total follows), the lower their scores in the first latent dimension seem to be. This is indicated by a gradient of large to small points from negative to positive values on the first latent dimension (Figure 2). Supporting this visual inspection, we found a substantial correlation between the total number of followed accounts and user values on the first latent dimension ($r = -.54, p < .001$). These observations are, however, purely post-hoc interpreted, because we had no prior assumptions about what this latent dimension in the VAE might represent before data analysis.

Next, we used the party estimates (i.e., column coordinates) on the second latent dimension in the VAE and the first dimension in the CA to check the overall estimated party alignment with expert ratings from the OES. The results are depicted in Figure 3. Broadly, the VAE more closely matched the expected overall party positions from the OES. The distance to the reference opinion estimates (OES) was lower for most of the parties compared to the CA (except SPD). Additionally, the overall distinction of estimated party positions (i.e., separating more left from more right-leaning parties) in the VAE was higher. In fact, party estimates in the CA were almost identical for five out of six parties (except AfD). Notably, the party estimates in both the VAE and CA for the party LINKE seem to deviate the most from the OES.




(a) Opinion point estimates for LINKE
(b) Opinion point estimates for Green party
(c) Opinion point estimates for SPD
(d) Opinion point estimates for FDP
(e) Opinion point estimates for CDU/CSU
(f) Opinion point estimates for AfD

**Figure 2.** Latent dimensions in the VAE. Note: Opinion point estimates in the VAE for all parties and users (combined training and test set). Each point represents a user, sequentially colored by the percentage of accounts followed per party (total number of accounts followed from a specific party divided by overall followed political accounts) with a brighter hue indicating a lower and a darker hue indicating a higher percentage of accounts followed from the respective party. Point sizes refer to the total number of followed accounts (independent of party) per user. Larger points indicate more followed accounts overall.

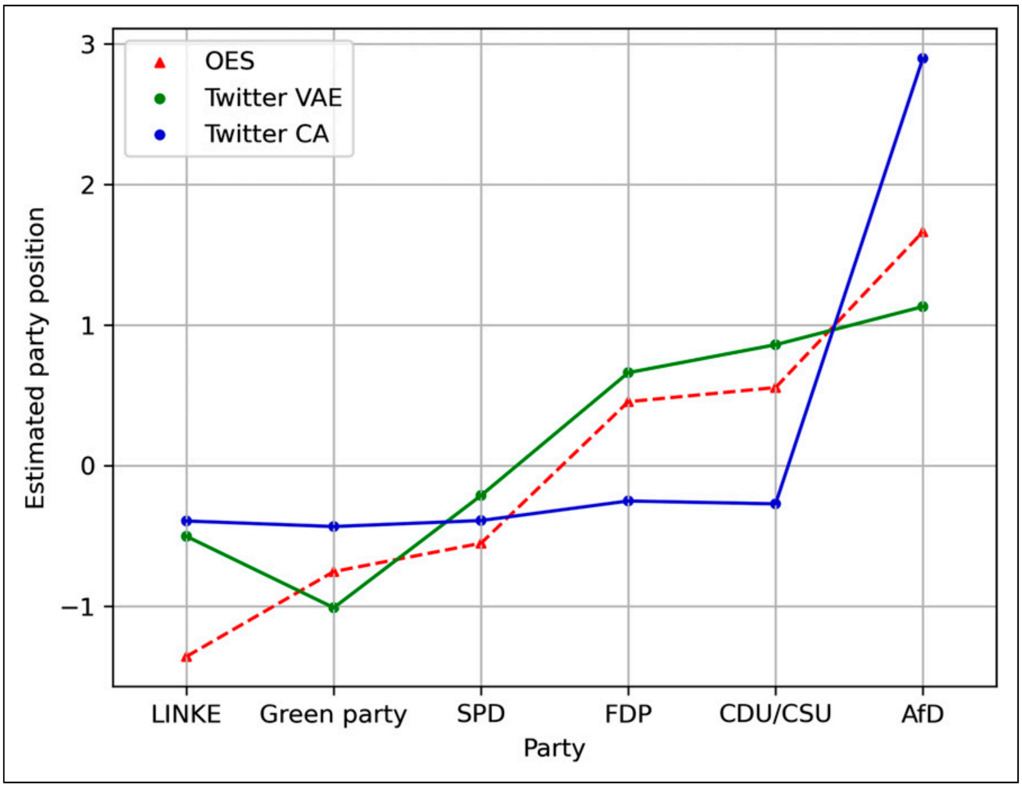



**Figure 3.** Comparison of left-right alignment parties. Note: OES estimates represent the mean estimates of party positioning by experts from 2021. The party ordering runs from most left to most right party as identified in the OES. Negative opinion point estimates on the Y-axis represent left-leaning parties and positive ones represent right-leaning parties. Points for VAE and CA represent party opinion estimates, calculated by using the median of all political accounts' column coordinates for each party. All estimates are standardized.

## *Opinion Estimate Validation Using Self-Report Data*

In the last step, we validated opinion estimates from the VAE and CA with our self-report dataset. First, we looked at the distributions of self-reported ideology and voting intentions of our sample. On average, our sample's self-reported ideology levels were left-leaning, and not a single individual reported strong right-leaning ideologies ($M_{ideology}$ = 3.29, $SD_{ideology}$ = 1.33, $MinMax_{ideology}$ = [1, 8]). The same was true for voting intentions. The majority of individuals reported intending to vote for left-leaning parties (Green party: 54, Linke: 15, SPD: 19) and only twelve individuals in total for the (more) right-leaning parties (FDP: 7, CDU/CSU: 4, AfD: 1). The remaining reported to vote for other, non-major parties (Other: 18) or to not vote at all (no vote: 1). These self-report results align only partly with the overall following rates of political accounts in our work since users in the self-report dataset followed even more politicians from left-leaning parties percentage-wise than users in the training and test set (Table 1).

Looking at the distribution of opinion point estimates of both the VAE and CA, their mean and standard deviation seemed to align with this overall left-leaning trend of the self-report data, with the VAE exhibiting a lower mean and higher standard deviation compared to the CA ($M_{CA} = -0.24$, $SD_{\mathrm{CA}} = 0.33$, $M_{\mathrm{VAE}} = -0.72$, $SD_{\mathrm{VAE}} = 0.76$). Looking at the relationship between opinion point estimates of users with their self-reported symbolic political ideology (left-right), we found a medium correlation in both models (VAE: $r = .46$, $p < .001$, CA: $r = .46$, $p < .001$). Although exhibiting similar positive correlation coefficients, the opinion point estimate distributions and absolute values seemed to differ between models (see Figure 4). In the CA, most point estimates were closely grouped except for a significant outlier on the highest self-reported ideology scale point. In comparison, the general pattern of point estimates and self-reported ideology in the VAE seemed more nuanced, showing a bigger variance in point estimates within and between self-reported ideology scores.

Next, we analyzed the capacity of our models' point estimates to predict party voting intentions. Figure 5 shows the distribution of point estimates by self-reported voting intentions for each party and model. As expected, the ordering of opinion estimates in the VAE was mostly in line with the expected ordering per party. That is, individuals intending to vote for more

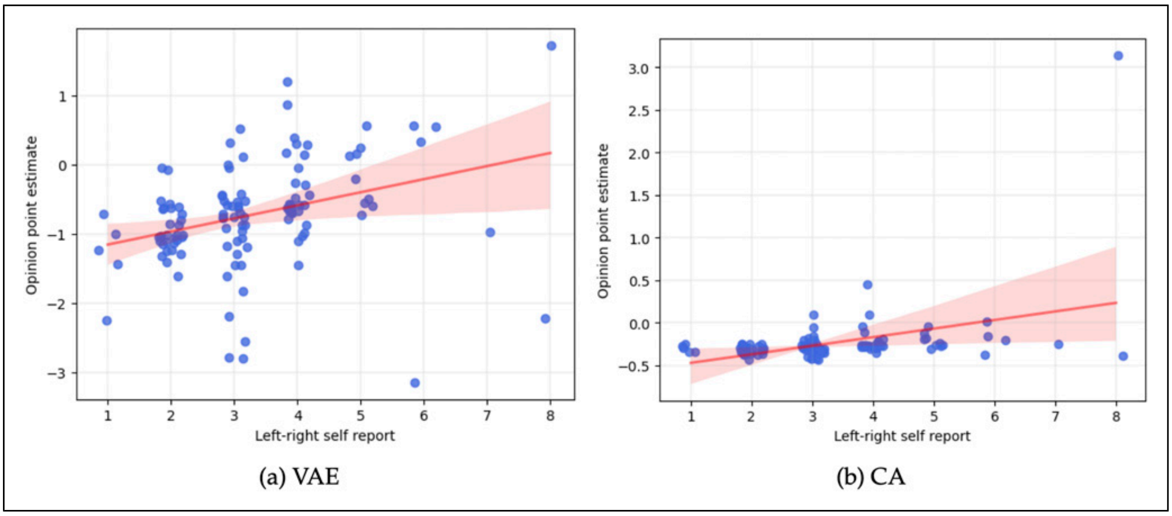



**Figure 4.** Correlation plot of opinion point estimates and self-reported ideology. Note: Opinion point estimates for all participants in the self-report dataset are represented by individual dots jittered on the x-axis with a bold red regression line. Error bars around the regression line indicate 95% confidence interval.

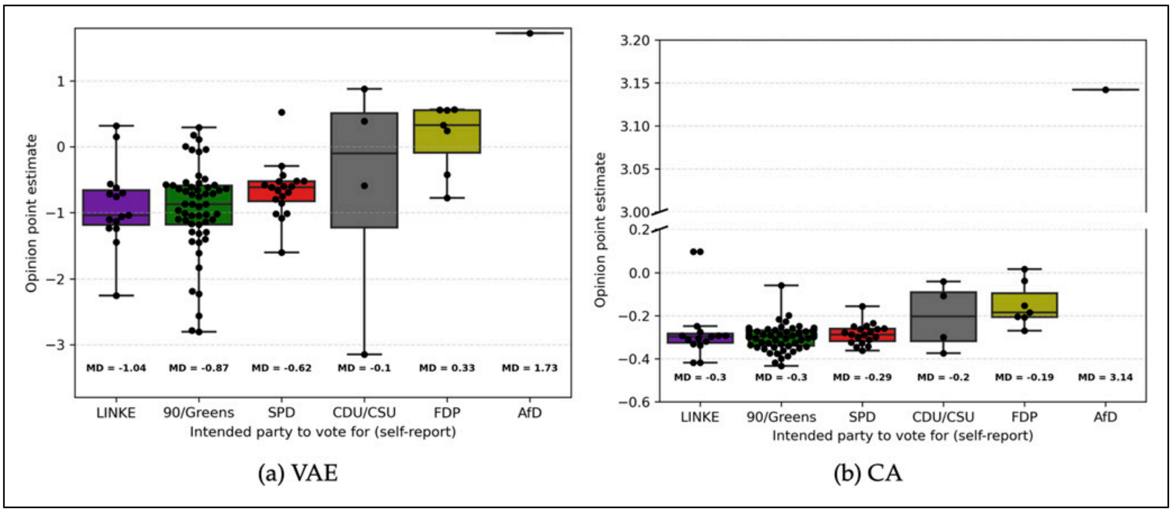



**Figure 5.** Self-reported voting intention and estimated ideological position. Note: Individual boxplots calculated for all participants intending to vote for respective party with an interquartile range of IQR = 3. Black dots indicating single observations. Points outside of Whiskers indicate outliers. Boxplots for AfD only showing single point estimate ($N$ = 1). Median point estimates per intended voting behavior shown underneath individual boxplots. CA graph (b) includes y-axis break to facilitate plotting.

left-leaning parties had lower, negative median opinion estimates ($Md_{Linke}$ = −1.04, $Md_{90/Greens}$ = −0.87, $Md_{SPD}$ = −0.62), whereas those intending to vote for more right-leaning parties had higher, (mostly) positive opinion estimates ($Md_{CDU/CSU}$ = −0.1, $Md_{FDP}$ = .33, $Md_{AfD}$ = 1.73). Notably, individuals intending to vote for CDU/CSU exhibited an overall lower median point estimate than expected (i.e., lower median estimates than the ones intending to vote for FDP). However, the point estimate variance for the CDU/CSU was relatively large compared to the other parties and only four individuals reported voting intentions for this party. Compared to the VAE, the CA median opinion estimates were less in line with the expected party ordering. Similar to the overall party positioning (Figure 3), the variance between median point estimates for the different parties was much smaller compared to the VAE with only the most right-leaning party (AfD) exhibiting a high difference (and high absolute value) in median point estimates. Median estimates for parties LINKE and Greens were identical (*Md* = −0.3) with the median estimate for SPD (*Md* = −0.29) being the lowest of all parties (although only marginally). Additionally, estimates for more right-leaning parties FDP and CDU/CSU were only slightly higher than for the previously mentioned left-leaning parties and also almost identical to one another ($Md_{FDP}$ = −0.19, $Md_{CDU/CSU}$ = −0.2). Put simply, the CA only showed small differences in estimated opinion estimates between individuals voting for different parties. This may, in turn, have led to a higher divergence of expected opinion estimate ordering based on the voting intention compared to the VAE. In sum, the relationship between the point estimates and voting intention in the VAE seemed more nuanced and in line with the expected distribution compared to the CA. This result was further supported by the multinomial logistic regression model, in which the VAE showed a superior model fit compared to the CA in predicting voting intentions through opinion point estimates (see supplemental materials, subsection II.IV).

Lastly, we also checked the CA and VAE following matrix reconstruction performance on the self-report dataset. Similar to the train and test set, the VAE consistently outperformed the CA and NCF across calculated metrics (see supplemental materials, section subsection II.I). This again indicates that the learned latent representations in the VAE more closely captured the network relationships compared to the CA.

## Discussion

In the present study, we explored whether social network data can be used to infer individuals' political opinions in countries other than the US. In doing so, we utilized dimensionality reduction (DR) models to analyze the network of German Twitter users to obtain individual-level political opinion estimates.

Our results not only corroborate findings from previous studies showing that estimating individual political opinions from social media data (SMD) using individuals' following decisions is feasible (e.g., Barberá, 2015a; Barberá et al., 2015; Bond & Messing, 2015) but that this estimation also works in a multiparty political system like Germany; and thus potentially in other countries outside of the US as well. This result is noteworthy since many countries other than the US have a multiparty system and—in the special case of Germany—ordering political parties on a general left-right continuum is a complex issue with no universally agreed-upon solution. Depending on the used dimension (social, economic, etc.) and applied method, researchers have come to different party orderings in the past (Jankowski et al., 2022; Lehmann et al., 2023). This ordering issue in combination with a higher number and unequal distribution of voting intention classes in our study (six parties following different political agendas in the German context) thus presents a more challenging prediction endeavor than in a bipartisan context like the US.

Despite these challenges, our study shows that opinion estimates of users following accounts from German politicians broadly align with the respective overall party stance judged by experts.

Further, the results suggest that the more accounts portraying a specific political stance social media users follow, the more likely they are to hold similar opinions. Particularly, the more accounts from “extreme” parties, like the right-leaning AfD, individuals follow, the more extreme their opinion point estimates become, reflecting a left-right dimension. In line with previous studies on social media and offline social networks, this supports the assumption of homophily (e.g., Aiello et al., 2012; McPherson et al., 2001).

Moreover, we find that using complex ML models can benefit the estimation accuracy of individuals’ public opinions from SMD. First, the unsupervised VAE ML model captured the intricacies of the following structure more precisely compared to an uninformed baseline model and a linear DR algorithm (CA). Possibilities to compare our models’ matrix reconstruction performance to the literature are somewhat limited since—to our knowledge—only one related, previous study (conducted in the US) reported the (unbalanced) reconstruction performance (e.g., Barberá et al., 2015). Compared to this study, however, our VAE exhibited higher absolute reconstruction performance and greater improvements over an uninformed baseline. The nuanced network representation in the VAE is further corroborated by a closer match of the ideological party estimates with expert ratings compared to the CA. The individual-level opinion estimates in the VAE also showed stronger relationships with self-reported opinions compared to the linear CA. Not only did the user opinion estimates show strong correlations with self-reported symbolic ideology but also a more accurate relationship with individuals voting intentions. All these results empirically support the assumption that complex, non-linear relationships in the following structure of social media networks exist and need to be considered in the modeling phase to estimate individual-level opinions more accurately. Being able to extract individual-level political opinions solely from SMD following decisions yields several practical implications. Using SMD seems to provide a promising alternative to cost and labor-intensive surveys. The extracted opinion estimates may further be used by researchers and policymakers alike to predict political outcomes like elections, opinion networks and political polarization in social media networks.

Nevertheless, our study comes with a few limitations. We found slight differences in the estimated political stance of German parties in our models and their expected positions rated by experts. One of the reasons might be structural biases in our data, even though we adhered to previous best practices of data cleaning as far as possible. We might not have excluded all inactive and/or bot accounts due to restrictions in users’ geo-location, and other meta information. However, despite the potentially resulting noise in our datasets, the VAE still captured meaningful relationships as shown by our results. Judging from the following distribution of political accounts compared to the overall vote shares in the 2021 German federal election and the fact that most social media users do not seem to follow political elites, our data may not be representative of the general population, a common problem of studies using SMD (Mellon & Prosser, 2017; Wojcieszak et al., 2022). The representativeness of our sample for the Twitter user space might further be limited since we only analyzed data of users who followed at least ten political accounts (similar to previous approaches). Importantly however, users who do follow political elites, seem to be more aligned with the opinions of the followed accounts which conversely aids our study’s assumption of homophilic networks (Wojcieszak et al., 2022). On a similar note, our self-report dataset is comparatively small and the following frequencies of political accounts deviated from the bigger training and test datasets. As shown by our results, the overall distribution of voting intentions and self-reported ideology are rather left-leaning and limited in their variance; only a few participants reported voting for right-leaning parties. This may have led to slightly biased correlations of self-reported opinions with point estimates and may thus also explain similar correlations comparing the VAE and CA. Nevertheless, a small to medium correlation of opinion estimates with symbolic ideology and a fit with voting intention patterns indicate that the VAE captured important patterns in the self-report dataset users' networks. Finally, researchers planning to adapt our method should be aware that VAEs require more careful planning, fine-tuning (i.e., hyperparameter and architecture selection) and computation time in contrast to simple linear DR models. Also, the interpretation of latent dimensions may require additional effort as noted in the Methods section. However, when trained and applied properly, they can capture the intricacies of the network in more detail, as shown by our results.

Based on our findings and study limitations, we pose several recommendations for future studies: While a ML DR model yielded promising results in the context of German SMD network-based analysis, future research should apply these models to other countries, political systems and politically uninterested users to test generalizability. Future studies may also test other existing, more traditional network analysis models and approaches which, for instance, focus on node-link prediction (instead of DR like our model, e.g., GNNs) to estimate individual-level political opinions. On a similar note, since our dataset exhibits a directed bipartite network structure (users →political accounts), future studies may explore the potential of other types of network data structures and representations (e.g., undirected user-user/user-political accounts networks, attributed graphs) to estimate political opinions from SMD. Doing so could, for instance, facilitate the application of resampling techniques and the application of other network analysis and benchmark models. Considering our rather small self-report dataset, future research may validate point estimates with self-report measures using larger samples. While we found substantial correlations between opinion estimates and self-reported symbolic ideology, future studies may investigate whether manifestations of dimensions of more complex models of political views (see for example Fatke, 2017; Feldman & Johnston, 2014; Gerber et al., 2010; Jost et al., 2009; Treier & Hillygus, 2009) can be predicted by these data as well. In this case, fitting models with more latent dimensions may be advisable to capture the manifoldness of users' political opinions in European countries even more accurately (Barberá, 2015a). On that note, future studies could explore whether our results can be generalized to measures of operational ideology assessing individuals' attitudes on specific topics and policy issues. Moreover, although our VAE ML model surpasses the network reconstruction performance of previous studies, linear (CA) and uninformed models (Naive), it is debatable what constitutes "good" dimensionality reduction performance. Therefore, future research could explore potential thresholds and benchmarks. Shortly after our data collection (July 2022), Twitter's leadership changed and user counts decreased substantially (Alex Hern, 2024). Although we do not expect this circumstance to affect (homophilic) behavior, future research may still investigate periodic effects of such events on social media networks and replicate our findings. Interrelated with this, Twitter has become less attractive to researchers and political actors as a data source since it does not share data with independent researchers (like most social media platforms) and sets comparatively high pricing for acquiring big datasets (Calma, 2023). This hinders and may even prevent further in-depth examinations in this research area (Bruns, 2019). We hope that with the Digital Services Act, researchers will (re)gain access to conduct research on online platforms in the future (European Commission, 2024).

Using social media data to assess the political opinions of individuals offers valuable benefits over traditional opinion surveys for researchers and policymakers alike. However, finding reliable and accurate methods to extract information from unstructured SMD in different sociopolitical contexts remains one of the main challenges. Our study adds new insights to this endeavor, showing that using ML models capable of capturing intricate user network characteristics in a multiparty system is crucial to estimating individual-level political opinions more accurately. Opinion estimates from SMD may then be used to monitor political trends, capture ideological shifts in societies, or predict political behavior like voting patterns or policy support, ultimately benefiting the democratic process in modern societies.

## Data Availability Statement

The data underlying this research project constitute “special categories of personal data” (“personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, [. . . ]” GDPR, Chapter 2, Art. 9). Data processing was based on the fact that “processing relates to personal data which are manifestly made public by the data subject” (GDPR, Chapter 2, Art. 9, §2e; training set, test set) or “the data subject has given explicit consent to the processing of those personal data [...]” (GDPR, Chapter 2, Art. 9, §2a; set). Participants of the survey to recruit the self-report dataset additionally provided consent that their data might be shared if re-identification is impossible. Given the exponential growth in available data analysis strategies (algorithms, etc.), we did not find an appropriate data masking or data aggregation strategy to prevent re-identification of participants and still provide the whole dataset for replicability of the findings to the public. Thus, if researchers are interested in replicating the results reported in the present work, we ask them to contact us (cornelia.sindermann@iris.uni-stuttgart.de) and we will provide access to part of the data, aggregated data, or the like. Note that every request will need to undergo thorough examination first to ensure compliance with the GDPR and that participants cannot be re-identified in the dataset access is provided to. All questionnaires used in the survey to recruit the self-report dataset will be made publicly available at the OSF upon acceptance of the manuscript. The analysis code used to produce the results is openly accessible on the OSF: https://osf.io/2ft8d/?view_only=e56f7e9b7fdc4b4f8188084584141793.

### Acknowledgments

We acknowledge the support of the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (MWK, Ministry of Science, Research and the Arts Baden-Württemberg under Az. 33-7533-9-19/54/5) in “Künstliche Intelligenz & Gesellschaft: Reflecting Intelligent Systems for Diversity, Demography and Democracy (IRIS3D)” and the support by the “Interchange Forum for Reflecting on Intelligent Systems” (IRIS) at the University of Stuttgart. Further, we acknowledge the support by the Stuttgart Center for Simulation Science (SimTech).

### Author Contributions

**Nils Brandenstein:** Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation; Data Curation; Writing - Original Draft; Visualization; Administration
**Christian Montag:** Writing - Review & Editing
**Cornelia Sindermann:** Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation; Data Curation; Writing - Review & Editing; Supervision; Administration

### Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: For reasons of transparency Dr. Montag mentions that he has received (to Ulm University and earlier University of Bonn) grants from agencies such as the German Research Foundation (DFG). Dr. Montag has performed grant reviews for several agencies; has edited journal sections and articles; has given academic lectures in clinical or scientific venues or companies; and has generated books or book chapters for publishers of mental health texts. For some of these activities he received royalties, but never from gaming or social media companies. Dr. Montag mentions that he was part of a discussion circle (Digitalität und Verantwortung: https://about.fb.com/de/news/h/gespraechskreis-digitalitaet-und-verantwortung/) debating ethical questions linked to social media, digitalization and society/democracy at Facebook. In this context, he received no salary for his activities. Also, he mentions that he currently functions as independent scientist on the scientific advisory board of the Nymphenburg group (Munich, Germany). This activity is financially compensated. Moreover, he is on the scientific advisory board of Applied Cognition (Redwood City, CA, USA), an activity which is also compensated.

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Nils Brandenstein https://orcid.org/0000-0003-2663-8793
Christian Montag https://orcid.org/0000-0001-8112-0837
Cornelia Sindermann https://orcid.org/0000-0003-1064-8866

## Supplemental Material

Supplemental material for this article is available online.

## References

Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web*, *6*(2), 1–33. https://doi.org/10.1145/2180861.2180866

Alex Hern. (2024). Twitter usage in US 'fallen by a fifth' since Elon Musk's takeover. Retrieved June 24, 2024, from. https://www.theguardian.com/technology/2024/mar/26/twitter-usage-in-us-fallen-by-a-fifth-since-elon-musks-takeover#:~:text=Use_of_Twitter_in_the,app%2Dmonitoring_company_Sensor_Tower

Al-Omairi, L. J., Abawajy, J., Chowdhury, M. U., & Al-Quraishi, T. (2021). An empirical analysis of graph-based linear dimensionality reduction techniques. *Concurrency and Computation: Practice and Experience*, *33*(5), e5990. https://doi.org/10.1002/cpe.5990

Anjaria, M., & Guddeti, R. M. R. (2014). Influence factor based opinion mining of Twitter data using supervised learning. 2014 sixth International Conference on communication systems and networks (COMSNETS), (pp. 1–8). https://doi.org/10.1109/COMSNETS.2014.6734907

Aral, S. (2020). *The hype machine: How social media disrupts our elections, our economy, and our health–and how we must adapt*. Crown. https://books.google.de/books?id=oH7JDwAAQBAJ

Barberá, P. (2015a). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, *23*(1), 76–91. https://doi.org/10.1093/pan/mpu011

Barberá, P. (2015b). How social media reduces mass political polarization. Evidence from Germany, Spain and the US. *Working Paper, New York University*. https://pablobarbera.com/static/barbera_polarization_APSA.pdf

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Berinsky, A. J. (2017). Measuring public opinion with surveys. *Annual Review of Political Science*, *20*(1), 309–329. https://doi.org/10.1146/annurev-polisci-101513-113724

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4), 175–308. https://doi.org/10.1016/j.physrep.2005.10.009

Bond, R., & Messing, S. (2015). Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook. *American Political Science Review*, *109*(1), 62–78. https://doi.org/10.1017/S0003055414000525

Brito, K. D. S., Filho, R. L. C. S., & Adeodato, P. J. L. (2021). A systematic review of predicting elections based on social media data: Research challenges and future directions. *IEEE Transactions on Computational Social Systems*, *8*(4), 819–843. https://doi.org/10.1109/TCSS.2021.3063660

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2016). Geometric deep learning: Going beyond Euclidean data. [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.1611.08097

Bruns, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, *22*(11), 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447

Calma, J. (2023). Twitter just closed the book on academic research. *The Verge*. Retrieved September 21, 2023, from. https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research

Chen, Z., Gan, W., Wu, J., Hu, K., & Lin, H. (2024). Data scarcity in recommendation systems: A survey [publisher: ACM New York, NY]. *ACM Transactions on Recommender Systems*.

Chikhi, N. F., Rothenburger, B., & Aussenac-Gilles, N. (2007). A comparison of dimensionality reduction techniques for web structure mining, 116–119. https://doi.org/10.1109/WI.2007.86

Chung, W., & Zeng, D. (2016). Social-media-based public policy informatics: Sentiment and network analyses of U.S. Immigration and border security. *Journal of the Association for Information Science and Technology*, *67*(7), 1588–1606. https://doi.org/10.1002/asi.23449

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), Article e2023301118. https://doi.org/10.1073/pnas.2023301118

De Backer, S., Naud, A., & Scheunders, P. (1998). Non-linear dimensionality reduction techniques for unsupervised feature extraction. *Pattern Recognition Letters*, *19*(8), 711–720. https://doi.org/10.1016/S0167-8655(98)00049-X

Dong, X., & Lian, Y. (2021). A review of social media-based public opinion analyses: Challenges and recommendations. *Technology in Society*, *67*(101724). https://doi.org/10.1016/j.techsoc.2021.101724

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *Sage Open*, *9*(1), 215824401983270. https://doi.org/10.1177/2158244019832705

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K., Buffone, A., Iwry, J., Seligman, M. E. P., & Ungar, L. H. (2021). Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, *26*(4), 398–427. https://doi.org/10.1037/met0000349

European Commission. (2024, January). DSA: Very large online platforms and search engines Shaping Europe's digital future. Retrieved from 26 January 2024. https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops

Fatke, M. (2017). Personality traits and political ideology: A first global assessment. *Political Psychology*, *38*(5), 881–899. https://doi.org/10.1111/pops.12347

Feldman, S., & Johnston, C. (2014). Understanding the determinants of political ideology: Implica-tions of structural complexity. *Political Psychology*, *35*(3), 337–358. https://doi.org/10.1111/pops.12055

Figeac, J., & Favre, G. (2023). How behavioral homophily on social media influences the perception of tie-strengthening within young adults' personal networks. *New Media & Society*, *25*(8), 1971–1990. https://doi.org/10.1177/14614448211020691

Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*, *31*(6), 649–679. https://doi.org/10.1177/0894439313493979

Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., & Ha, S. E. (2010). Personality and political. attitudes: Relationships across issue domains and political contexts [Publisher: Cambridge University Press]. *American Political Science Review*, *104*(1), 111–133. https://doi.org/10.1017/S0003055410000031

Grover, A., & Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. [Version Number: 1]. https://doi.org/10.48550/ARXIV.1607.00653

Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, *53*(2), 217–288. https://doi.org/10.1137/090771806

Jankowski, M., Kurella, A.-S., Stecker, C., Blätte, A., Bräuninger, T., Debus, M., Müller, J., & Pickel, S. (2022). Die Positionen der Parteien zur Bundestagswahl 2021: Ergebnisse des Open Expert Surveys. *Politische Vierteljahresschrift*, *63*(1), 53–72. https://doi.org/10.1007/s11615-022-00378-7

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual Review of Psychology*, *60*(1), 307–337. https://doi.org/10.1146/annurev.psych.60.110707.163600

Kemp, S. (2023, July). Digital 2023 july global statshot report (tech. Rep.). Datareportal. Retrieved August 23, 2023, from. https://datareportal.com/reports/digital-2023-july-global-statshot

Khanam, K. Z., Srivastava, G., & Mago, V. (2023). The homophily principle in social network analysis: A survey. *Multimedia Tools and Applications*, *82*(6), 8811–8854. https://doi.org/10.1007/s11042-021-11857-1

Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. [Version Number: 1]. https://doi.org/10.48550/ARXIV.1611.07308

Kumar, A., & Sebastian, T. M. (2012). Sentiment analysis: A perspective on its past, present and future. *International Journal of Intelligent Systems and Applications*, *4*(10), 1–14. https://doi.org/10.5815/ijisa.2012.10.01

Kurucz, M., Benczúr, A., & Pereszlényi, A. (2008). Large-scale principal component analysis on livejournal friends network [Publisher: Citeseer]. *Proceedings of SNAKDD, 2008*.

Kwak, J.-a., & Cho, S. K. (2018). Analyzing public opinion with social media data during election periods: A selective literature review. *Asian Journal for Public Opinion Research*, *5*(4), 285–301. https://doi.org/10.15206/AJPOR.2018.5.4.285

Lansdall-Welfare, T., Lampos, V., & Cristianini, N. (2012). Effects of the recession on public mood in the UK. Proceedings of the 21st International Conference on World Wide Web, 1221–1226. https://doi.org/10.1145/2187980.2188264

Lee, D. H., & Brusilovsky, P. (2010). Social networks and interest similarity: The case of CiteULike. Proceedings of the 21st ACM conference on Hypertext and hypermedia (pp. 151–156). https://doi.org/10.1145/1810617.1810643

Lehmann, P., Franzmann, S., Burst, T., Regel, S., Riethmüller, F., Volkens, A., Weßels, B., & Zehnter, L. (2023). The Manifesto data collection. *Manifesto Project (MRG/CMP/MARPOR)*. Version 2023a [Place: Berlin/Göttingen]. https://doi.org/10.25522/manifesto.mpds.2023a

Livne, A., Simmons, M., Adar, E., & Adamic, L. (2021). The party is over here: Structure and content in the 2010 election. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), 201–208. https://doi.org/10.1609/icwsm.v5i1.14129

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 205316801772000. https://doi.org/10.1177/2053168017720008

Nanga, S., Bawah, A. T., Acquaye, B. A., Billa, M.-I., Baeta, F. D., Odai, N. A., Obeng, S. K., & Nsiah, A. D. (2021). Review of dimension reduction methods. *Journal of Data Analysis and Information Processing*, *9*(3), 189–231. https://doi.org/10.4236/jdaip.2021.93013

Nishana, S., & Surendran, S. (2013). Graph embedding and dimensionality reduction-a survey. *International Journal of Computer Science & Engineering Technology (IJCSET)*, *4*(1), 29–34.

Pallavicini, F., Cipresso, P., & Mantovani, F. (2017). Beyond sentiment. In *Sentiment analysis in social networks* (pp. 13–29). Elsevier. https://doi.org/10.1016/B978-0-12-804412-4.00002-4

Reveilhac, M., Steinmetz, S., & Morselli, D. (2022). A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion. *Multimedia Tools and Applications*, *81*(7), 10107–10142. https://doi.org/10.1007/s11042-022-12101-0

Rousidis, D., Koukaras, P., & Tjortjis, C. (2020). Social media prediction: A literature review. *Multimedia Tools and Applications*, *79*(9–10), 6279–6311. https://doi.org/10.1007/s11042-019-08291-9

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social media analyses for social measurement. *Public Opinion Quarterly*, *80*(1), 180–211. https://doi.org/10.1093/poq/nfv048

Silva, T. C., & Zhao, L. (2016, January). *Machine learning in complex networks* (1st ed.). Springer.

Sindermann, Cornelia, Kannen, Christopher, & Montag, Christian, et al. (2021). The degree of heterogeneity of news consumption in Germany—Descriptive statistics and relations with individual differences in personality, ideological attitudes, and voting intentions. *News Media & Society*, *26*(2), 711–731. https://doi.org/10.1177/14614448211061729

Sindermann, Cornelia, Kannen, Christopher, & Montag, Christian (2022). Longitudinal data on (political) news consumption and political attitudes in a German sample collected during the election year 2021. *Data Brief*, *43*, 108326. https://doi.org/10.1016/j.dib.2022.108326

Sindermann, Cornelia, Kannen, Christopher, & Montag, Christian (2023). Linking primary emotional traits to ideological attitudes and personal value types. *PLoS One*, *18*(1), e0279885. https://doi.org/10.1371/journal.pone.0279885

Skoric, M. M., Liu, J., & Jaidka, K. (2020). Electoral and public opinion forecasts with social media data: A meta-analysis. *Information*, *11*(4), 187. https://doi.org/10.3390/info11040187

Tabassum, S., Pereira, F. S. F., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *WIREs Data Mining and Knowledge Discovery*, *8*(5), Article e1256. https://doi.org/10.1002/widm.1256

Tausanovitch, C., & Warshaw, C. (2017). Estimating candidates' political orientation in a polarized congress. *Political Analysis*, *25*(2), 167–187. https://doi.org/10.1017/pan.2017.5

Tilak, S., Evans, M., Wen, Z., & Glassman, M. (2023). Social network analysis as a cybernetic modelling facility for participatory design in technology-supported college curricula. *Systemic Practice and Action Research*, *36*(5), 691–724. https://doi.org/10.1007/s11213-022-09625-9

Traber, D., Stoetzer, L. F., & Burri, T. (2023). Group-based public opinion polarisation in multi-party systems. *West European Politics*, *46*(4), 652–677. https://doi.org/10.1080/01402382.2022.2110376

Treier, S., & Hillygus, D. S. (2009). The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly*, *73*(4), 679–703. https://doi.org/10.1093/poq/nfp067

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the International AAAI Conference on Web and Social Media*, *4*(1), 178–185. https://doi.org/10.1609/icwsm.v4i1.14009

Tunstel, E., Cobo, M. J., Herrera-Viedma, E., Rudas, I. J., Filev, D., Trajkovic, L., Chen, C. L. P., Pedrycz, W., Smith, M. H., & Kozma, R. (2021). Systems science and engineering research in the context of systems, man, and cybernetics: Recollection, trends, and future directions [Conference Name: IEEE transactions on systems, man, and cybernetics: Systems]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *51*(1), 5–21. https://doi.org/10.1109/TSMC.2020.3043192

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, *69*(102199). https://doi.org/10.1016/j.electstud.2020.102199

Widmann, T., & Wich, M. (2023). Creating and comparing dictionary, word embedding, and transformer-based models to measure discrete emotions in German political text. *Political Analysis*, *31*(4), 626–641. https://doi.org/10.1017/pan.2022.15

Wilko, Z. (2021, October). Results of the German federal election 2021. (tech. rep.). https://www.wahlrecht.de/ergebnisse/bundestag.htm

Wojcieszak, M., Casas, A., Yu, X., Nagler, J., & Tucker, J. A. (2022). Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. *Science Advances*, *8*(39), Article eabn9418. https://doi.org/10.1126/sciadv.abn9418

Yan, S., Xu, D., Zhang, B., Zhang, H.-j., Yang, Q., & Lin, S. (2007). Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(1), 40–51. https://doi.org/10.1109/TPAMI.2007.250598

Zarzour, H., Al-Sharif, Z., Al-Ayyoub, M., & Jararweh, Y. (2018). A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques [Journal abbreviation: 2018 9th International Conference on Information and communication Systems (ICICS)]. 2018 9th International Conference on information and communication systems (ICICS), (pp. 102–106). https://doi.org/10.1109/IACS.2018.8355449

Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: A comprehensive review. *Computational Social Networks*, *6*(1), 11. https://doi.org/10.1186/s40649-019-0069-y

## Author Biographies

**Nils Brandenstein** is currently a PhD student in Psychology at Heidelberg University, Germany, where he also earned his Masters' degree. His research interest lies in the field of Political Psychology and he uses Machine Learning models to investigate a variety of topics, including belief in conspiracy theories, sustainable behavior and political attitudes/behavior.

**Dr. Christian Montag** is Professor for Molecular Psychology at Ulm University, Ulm, Germany. He works at the intersection of psychology, neuroscience, computer science and behavioral economics. At the moment his work focuses on the research-questions how AI impacts on society and how social media can be improved.

**Dr. Cornelia Sindermann** did her Ph.D. in Psychology at Ulm University, Ulm, Germany. Currently, she is the Independent Research Group Leader of the Computational Digital Psychology team within the Interchange Forum for Reflecting on Intelligent Systems, University of Stuttgart. She is interested in how interactions between individual differences and technological innovations shape how information is presented and processed, and how this, in turn, impacts political opinion formation and behavior.

# Declaration in accordance to § 8 (1) c) and d) of the doctoral degree regulation of the Faculty

**FAKULTÄT FÜR VERHALTENS- UND EMPIRISCHE KULTURWISSENSCHAFTEN**

**UNIVERSITÄT HEIDELBERG** ZUKUNFT SEIT 1386

**Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg / Doctoral Committee of the Faculty of Behavioural and Cultural Studies of Heidelberg University**

**Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / Declaration in accordance to § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies**

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe. / I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

**Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / Declaration in accordance to § 8 (1) d) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies**

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe. / I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

| | |
|---|---|
| Vorname Nachname / First name Family name | Nils Brandenstein |
| Datum / Date | 10.04.2025 |
| Unterschrift / Signature | N. Brandenstein |