ANNUAL MEETING

DVPW Section Methods of Political Science

MAR 27-28, 2025
MANNHEIM

UNIVERSITÄT MANNHEIM

The open-source LaTeX template, AMCOS_booklet, used to generate this booklet is available at https://github.com/maximelucas/AMCOS_booklet

The cover page was generated using a template provided by TemplateLAB, which is available at https://templatelab.com/cover-page-templates.

The picture shown on the cover page is a Wikimedia Commons File, licensed under CC-BY-3.0 and available at https://commons.wikimedia.org/wiki/File:MZES_Uni_Mannheim_A5.jpg. It was taken by Nikolaus Hollermeier and shows the conference venue, A5, 6 in Mannheim.

Contents

About **4**
DVPW Section "Methods of Political Science" . . . 4
Annual Meeting . . . 4
Organizing committee . . . 4

Timetable **5**
General information . . . 5
Wednesday, March 26, 2025 . . . 5
Thursday, March 27, 2025 . . . 6
Friday, March 28, 2025 . . . 9

Presentations **11**
Thursday, March 27 . . . 11
Panel 1: Text as data: methodological issues . . . 11
Panel 2a: Elections, scandals, multilingualism, and diversity . . . 15
Panel 2b: Networks, AI, and conflict analysis . . . 19
Panel 3a: Groups, identities, and voting behavior . . . 23
Panel 3b: All politics is local . . . 27
Panel 4a: Images as data . . . 31
Panel 4b: Survey experiments . . . 35
Friday, March 28 . . . 38
Panel 5a: Measuring positions and salience from text . . . 38
Panel 5b: Parliamentary debate . . . 41
Panel 6: Studying public opinion . . . 45

List of Participants **49**

Locations and Useful Information **51**
Hotel and conference venue . . . 51
Meals . . . 51

Acknowledgments **52**

About

DVPW Section "Methods of Political Science"

The Section *"Methods of Political Science"* (SekMethoden) of the German Political Science Association (DVPW) facilitates the exchange between scholars who deal with methodological aspects of political science research. Through the organization of conferences, the operation of the Section's homepage and the regular information of the Section's members, it promotes the networking of the participating scholars and is committed to the further development of the subfield in research and teaching. The Section is also a member society of the German Statistical Association (DAGStat).

Website (DE/EN): dvpw.de/gliederung/sektionen/methoden-der-politikwissenschaft

Mailing list: lists.rrz.uni-hamburg.de/mailman/listinfo/sekmethoden

Annual Meeting

The Annual Meeting of SekMethoden is jointly organized by the Section Co-Speakers and a local organizing committee.

Organizing committee

Cohen, Denis	Section speaker & local organizer	MZES, University of Mannheim
Rajski, Hannah	Local organizer	MZES, University of Mannheim
Rittmann, Oliver	Local organizer	MZES, University of Mannheim
Gessler, Therea	Section speaker & co-organizer	Viadrina University Frankfurt/Oder
Leininger, Arndt	Section speaker & co-organizer	Chemnitz University of Technology

Timetable

T1 Paper Track **T2 Presentation Track** *** Presenting author**

General information

- Upload papers to the conference cloud by **Friday, March 21**. Password: `methods2025`.
- Please limit presentations to **10 minutes** and discussants comments to **5 minutes**.

Wednesday, March 26, 2025

Informal pre-conference dinner (self-pay) at HiKoo Noodle Kitchen (M2), 19:00-21:00. Spaces limited – available to presenters who signed up during registration .

Thursday, March 27, 2025

09:00-09:10	**Welcome** A230/231		
09:10-10:50	**Panel 1: Text as data: methodological issues** A230/231		
	Chair: Theresa Gessler		
	Discussant: Ruben Bach (MZES Data and Methods Unit)		
	T1	C. Oswald L. Braun*	Automated information extraction from text variables in event datasets with large language models
	T2	H. Licht* H. H. V. Nguyen	Metadata-aware fine-tuning: Mitigating regression-to-the-mean bias in text classification
	T1	M. Ratkovic* T. Gschwend L. Wendering K. Bauer A. Kurella O. Rittmann M. Sauter N. Schwitter	Harnessing GPT for Enhanced Academic Writing: Evidence from a Field Experiment with Early-Career Researchers
	T1	I. Sen* M. Faulborn M. Pellert A. Spitz D. Garcia	Only a Little to the Left: A Theory-grounded Measure of Political Bias in Large Language Models
10:50-11:10	**Coffee break** Lobby 2nd floor		
11:10-12:30	**Panel 2a: Elections, scandals, multilingualism & diversity** A230		
	Chair: Arndt Leininger		
	T2	R. Passerotti*	Automated Detection of Electoral Pledges in Political and Media Discourses: A Case Study of the 2022 French Presidential Election
	T2	S. Sangiovanni* A. Ceron G. M. Invernizzi	Investigating Political Scandals: Combining Conjoint Analysis and Audio-Based Experiments
	T2	T. Gessler* L. Kraft D. Morgenstern	Opposing autocratization across regimes
	T2	S. Morgenstern* R. Bach	Social Inequality and Super-diversity: The impact of minority-contact on political preferences. A field experiment with text-as-data insights.

11:10-12:30	**Panel 2b: Networks, AI & conflict analysis** A231	
	Chair: Oliver Rittmann	
T2	E. Wolfes-Wenker*	Leveraging LinkedIn data for career trajectory clustering of interest group representatives
T2	K. Seng* J. Garritzmann D. Broska	Parameter Stability in Semi-parametric Neural Networks
T2	J. Bernauer*	Bridging the Gap Between User and Hardware: Research-Ready AI Infrastructure for the Social Sciences
T2	J. Hellinger* R. Bhavnani	Why Discriminate? The Logic of Targeting in Gaza's Aerial Occupation
12:30-13:30	**Lunch break** Lobby 2nd floor	
13:30-15:00	**Panel 3a: Groups, identities, and voting behavior** A230	
	Chair: Denis Cohen	
T1	M. Kollberg* P. Dinesen K. Sonderskov	What Advances the Causes of Disadvantaged Groups in Identity Politics? Experimental Evidence on Particularistic and Universalistic Arguments *Discussant:* P. Kaminski
T2	N. Buntfuß*	The Social Foundations of Political Hostility: Disentangling Political and Social Identities through a Factorial Survey Experiment
T1	L. Erhard P. Kaminski* Z. O. Dunivin S. Šarenkapa E. Deiss-Helbig I. Guinaudeau T. Matthieß R. H. Heiberger	Automating ambiguous text annotation? LLMs struggle to detect social group references *Discussant:* M. Kollberg
T2	P. Selb* J. Wiederspohn	Seven Decades of Vote Validation Research. A Systematic Review.
13:30-15:00	**Panel 3b: All politics is local** A231	
	Chair: Hannah Rajski	
T2	S. Ziaja* A. May	Community-wide effects of a randomised citizen council

T1	M. C. Duarte* J. M. Fernandes M. M. Pereira	The Effects of Participatory Democracy on Political Attitudes: Causal Evidence from Citizens' Assemblies *Discussant:* S. Eschenwecker
T1	S. Eschenwecker* T. König	Whose Friend, Whose Foe? The Causal Effect of a Populist Challenger Party's Candidacy on Party Competition *Discussant:* M. C. Duarte
T2	K. Lindemann* L. Schramm	Irregular immigration and the reintroduction of national border controls: Evidence from Germany
15:00-15:30	**Coffee break** Lobby 2nd floor	
15:30-17:10	**Panel 4a: Images as data** A230	
	Chair: Oliver Rittmann	
T2	I. E. Acquarone* A. C. Cinar G. D. Landro N. Zilberstein	AI-Generated Visuals: A Solution for Image Demand in Social Science Research
T1	S. Scholz* N. Weidmann*	Manipulating Protest Size Estimates with AI-generated Images *Discussant:* C. Arnold
T1	C. Arnold* A. Küpfer	Structuring Quantitative Image Analysis with Object Prominence *Discussant:* N. Weidmann
T2	A. Aísa* B. González S. Scholz C. v. Soest N. B. Weidmann	Tactical Diversity Within Protest Events: An image-to-data analysis
15:30-17:10	**Panel 4b: Survey experiments** A231	
	Chair: Arndt Leininger *Discussant: Alexander Wenz (MZES Data and Methods Unit)*	
T1	F. Haggerty* P. W. Thurner L. Rudolph	Detecting heterogeneity in conjoint experiments on arms trade preferences in 5 major Western export countries: An application of two recent methods using survey data
T1	W. El-Ajou G. Brückmann* I. Stadelmann-Steffen	Is it worth to employ a serious game? An experimental analysis of information provision and public opinion formation

T1	L. Rudolph*	Validating Stated against Actual Baviour in Experiments – Comparing Vignette and Quasi-Experimental Evidence from Bangladesh
T1	C. Sajons* B. Müller M Samoiliuk	When statistical truth serum is not enough – Lessons from a failed list experiment to measure corruption in Ukraine
17:10-17:15	**Short break**	
17:15-18:00	**DVPW section members' meeting**	
19:00-22:00	Conference dinner	

Friday, March 28, 2025

09:00-10:40	**Panel 5a: Measuring positions and salience from text** A230	
	Chair: Theresa Gessler	
T1	B. Zafer*	Favor Exchange between Government and Media Owners *Discussant:* H. H. V. Nguyen
T1	J. Haak* S. Morét* L. Schwarz	From Weather to Words – The Effect of Extreme Weather on Political Elite Communication *Discussant:* L. Siefken
T1	H. H. V. Nguyen*	Improving Cross-Domain Text Classification When Measuring Multidimensional Party Positions *Discussant:* B. Zafer
T1	L. Siefken*	Applying Large Langue Models for the Measure of Populism in Election Manifestos *Discussant:* J. Haak
09:00-10:40	**Panel 5b: Parliamentary debate** A231	
	Chair: Hannah Rajski	
T1	M. C. Duarte R. Kittel*	Only Hot Air? National Parliamentary Discourse and Its Effect on Public Opinion on the European Union *Discussant:* T. Widmann
T1	V. Kunz*	Distinct Rhetoric? Predicting Voting Defection from Parliamentary Speech Using Language Models and Machine Learning *Discussant:* C. Leonhardt

	T1 T. Widmann*	Measuring 'Democratic Health' in Political Text: Evidence from 150 years of German Parliamentary Debates *Discussant:* R. Kittel
	T1 C. Leonhardt* A. Blätte	Linking Textual Data - Potentials and Pitfalls of using Named Entity Linking and Record Linkage in Political Science Research *Discussant:* V. Kunz
10:40-11:10	**Coffee Break** Lobby 2nd floor	
11:10-12:50	**Panel 6: Studying public opinion** A230/231	
	Chair: Denis Cohen	
	T1 A. Wuttke* M. Aßenmacher C. Klamm M. M. Lang Q. Würschinger F. Kreuter	AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers *Discussant:* K. Bogatyrev
	T1 P. Koc* M. Steenbergen	Assessing Policy-Driven Shifts in Public Opinion Across Nations: The Challenges of Staggered Adoption, Measurement Error, and Limited Sample Size *Discussant:* M. Shiraef
	T1 K. Bogatyrev*	Shifting Ground: The 2023 Earthquake and Electoral Accountability in Turkey *Discussant:* A. Wuttke
	T1 M. Shiraef*	The Tests of Time and Fluid Concepts: An Empirical Validation of Identity Measures Across Regime Shifts *Discussant:* P. Koc
12:50-13:00	**Farewell**	
13:15-14:30	*Optional:* Closing lunch	

Presentations

T1 Paper Track

T2 Presentation Track

Thursday, March 27

Panel 1: Text as data: methodological issues

Automated information extraction from text variables in event datasets with large language models

C. Oswald[1], **L. Braun**[1]

T1

University of the Bundeswehr

Scholars began studying subnational conflict processes with the advent of conflict event data. However, such event data are mostly confined to information about the type of violence and actors, the number of casualties, time, and location, whereas the text sources upon which they rely, oftentimes provided in a notes variable for individual observations, contain much more information. Using abductions and forced disappearances in the Armed Conflict Location and Event Data, we generate additional variables about the number of victims, subdivided by gender and whether they are of age or underage, and whether there was a ransom demand. We demonstrate that large language models can extract additional information about events with accuracy of over 90 percent. This opens up new, previously impossible research agendas such as analyzing government abductions and near-real time kidnappings. The approach put forward could easily be extended to gather other, additional information, for example about tactics and purported aims of the perpetrators. Importantly, we also show that open-weight models perform at least as well as closed-weight ones for this particular task.

Metadata-aware fine-tuning: Mitigating regression-to-the-mean bias in text classification

H. Licht[1], **H. H. V. Nguyen**[2]

[1] University of Innsbruck
[2] Universität Bremen Deutsches Institut für Erwachsenenbildung

T2

Researchers increasingly fine-tune Transformer text classifiers on human-annotated examples to generate measurements for comparative analysis. However, blindly applying such fine-tuned classifiers to unseen data from different contexts (e.g., countries, languages, parties, and/or periods) can result in systematic biases. This study focuses on regression-to-the-mean bias. Across a wide range of text datasets and classification tasks, we show that regression-to-the-mean bias arises when the relative frequency of a label class varies systematically between subgroups' training examples but the classifier's predictions converge to the overall label class distribution in training data. Specifically, we demonstrate that classifiers fine-tuned without any context information tend to systematically under- or overestimate a label class's prevalence in held-out data from a subgroup depending on whether the label class is over- or underrepresented in the subgroup's training data. We hypothesize that such regression to the mean stems from models' reliance on spurious features during fine-tuning that correlate with label distribution patterns, such as vocabulary and syntactic cues. Accordingly, we propose to leverage subgroup metadata indicators during fine-tuning to mitigate regression-to-the-mean bias in classifiers' predictions. Through systematic experiments with cross-sectional, multilingual, and temporally diverse datasets, we assess the effectiveness of metadata-aware fine-tuning in improving subgroup performance parity and cross-lingual equivalence. Our results and our open-source Python library for metadata-aware Transformer fine-tuning contribute to making text classification-based measurement in comparative research more reliable and valid.

Only a Little to the Left: A Theory-grounded Measure of Political Bias in Large Language Models

I. Sen[1], **M. Faulborn**[2], **M. Pellert**[3], **A. Spitz**[2], **D. Garcia**[2]

T1

[1] University of Mannheim
[2] University of Konstanz
[3] Barcelona Super Computing Centre

Prompt-based language models like GPT4 and LLaMa have been used for a wide variety of use cases such as simulating agents, searching for information, or for content analysis. For all of these applications and others, political biases in these models can affect their performance. Several researchers have attempted to study political bias in language models using evaluation suites based on surveys, such as the Political Compass Test (PCT), often finding a particular leaning favored by these models. However, there is some variation in the exact prompting techniques, leading to diverging findings and most research relies on constrained-answer settings to extract model responses. Moreover, the Political Compass Test is not a scientifically valid survey instrument. In this work, we contribute a political bias measured informed by political science theory, building on survey design principles to test a wide variety of input prompts, while taking into account prompt sensitivity. We then prompt 11 different open-source and commercial models, differentiating between instruction-tuned and non-instruction-tuned models, and automatically classify their political stances from 88,110 responses. Leveraging this dataset, we compute political bias profiles across different prompt variations and find that while PCT exaggerates bias in certain models like GPT3.5, measures of political bias are often unstable, but generally more left-leaning for instruction-tuned models.

Harnessing GPT for Enhanced Academic Writing: Evidence from a Field Experiment with Early-Career Researchers

M. Ratkovic[1], **T. Gschwend**[1], **L. Wendering**[1], **K. Bauer**[1], **A. Kurella**[1], **O. Rittman**[1], **M. Sauter**[1], **N. Schwitter**[1]

T1

University of Mannheim

In an era of rapid digital transformation, emerging AI technologies are increasingly influencing academic practice. This study examines the impact of GPT on academic writing quality and metacognitive engagement among early-career researchers in the social sciences. In a pre-registered field experiment involving 22 participants, researchers were stratified by writing skill and project development, then randomly assigned to either a GPT-assisted group or a control group during a focused three-day seminar. The results indicate that GPT assistance significantly improves the clarity and coherence of academic writing ($p < 0.1$) without adversely affecting originality. Principal Components Analysis shows that the primary impact of GPT is on enhancing the structural organization of written content rather than on its substantive analytical depth. Additionally, linguistic analysis using LIWC reveals that GPT-assisted participants incorporated more process-oriented language in their reflections, suggesting a measurable shift toward a more deliberate and organized writing strategy. A supplementary evaluation further demonstrates that current commercial AI tools are capable of replicating both baseline ratings and the treatment effects derived from faculty assessments, underscoring the potential for automating aspects of the review process. These findings provide evidence that integrating GPT can promote improvements in academic writing quality while maintaining the core intellectual contributions of the work. This study contributes to ongoing discussions on the judicious incorporation of AI in academic settings and offers practical insights for institutions seeking to modernize scholarly communication and support emerging researchers in achieving greater clarity and efficiency in their writing.

Panel 2a: Elections, scandals, multilingualism, and diversity

Automated Detection of Electoral Pledges in Political and Media Discourses: A Case Study of the 2022 French Presidential Election

R. Passerotti[1]

T2

Sciences Po Grenoble - Université Grenoble Alpes, PACTE

During electoral campaigns, candidates' pledges are communicated through various mediums, such as speeches, press conferences, interviews, or press releases. Political scientists consider these pledges central to the study of representative democracy; consequently, they need reliable methods to easily identify them across different sources and languages. However, the specific nature of pledge definitions makes this task both complex and time-consuming to perform manually. To address this challenge, researchers have sought to develop automated methods for identifying promises. But these models are highly dependent on training data and essentially trained on party platforms: they could not perform as well as expected on other mediums such as interviews or debates. Furthermore, prior research has highlighted the importance of language-specific features, yet not studies have been conducted on French data. This study seeks to fill these gaps by presenting a comparison of supervised automated text analysis methods to identify electoral pledges across various forms of political communication. It relies on an original annotated dataset for promises made by the 8 main candidates in the 2022 French presidential election, encompassing meeting speeches, press conferences, interviews, debates, and textual documents such as press releases. To tackle this challenge, we use a Random Forest algorithm with word embedding and active learning, and several BERT models. This approach paves the way for finer-grained analysis of party competition throughout the electoral cycle, using pledges as units of measurement, and sheds light on the effects of campaign dynamics on promises.

Investigating Political Scandals: Combining Conjoint Analysis and Audio-Based Experiments

S. Sangiovanni[1], **A. Ceron**[1], **G. M. Invernizzi**[2]

T2

[1] University of Milan
[2] Bocconi University

This study explores how political scandals influence voter evaluations using a dual experimental design that combines conjoint analysis with multi-voice text-to-speech (TTS) technology. While voters evaluate candidates based on policy positions and partisan affiliation, political scandals represent critical negative valence shocks that can shape perceptions of integrity and competence. This research investigates how voters prioritize scandals relative to other candidate attributes and examines whether shared values, such as co-partisanship and ideological alignment, moderate their impact on voter assessments. In the first experiment, we employ a fully randomized conjoint design to estimate the relative importance of scandals compared to party affiliation, policy stances, and competence displays. By differentiating between financial (e.g., corruption, bribery) and personal (e.g., sexual misconduct, drug abuse) scandals, the experiment sheds light on the conditions under which scandals are most consequential and how their effects interact with voters' ideological predispositions. The second experiment complements the conjoint design with an audio-based approach, leveraging multi-voice TTS technology to simulate realistic political debates. Debate scripts generated using a large language model are converted into speech, where one politician attacks their opponent based on a scandal or a valence dimension while the other redirects the discussion to policy proposals. By varying the tone of delivery and comparing calm and aggressive accusations, we test whether rhetorical style influences how voters process valence-related information. The combination of these two experiments not only provides causal leverage but also represents a methodological innovation in experimental design through audio-based methods that integrate multi-voice TTS technology. This approach deepens our understanding of how voters evaluate candidates involved in scandals and contributes to literature on negative campaigning and electoral behaviour, offering insights into how tone, delivery, and framing of political attacks influence candidate perceptions and voter decision-making.

Opposing autocratization across regimes

T. Gessler[1], **L. Kraft**[1], **D. Morgenstern**[1]

[1] European University Viadrina, Frankfurt (Oder)

T2

The manuscript introduces a novel approach for comparing opposition to incumbent-led autocratization across different regime types, building on methods used in the study of social movements such as protest event analysis and contentious episodes analysis. By focusing on the interactions between government authorities and challengers surrounding specific policy proposals that weaken democratic institutions, the method examines the composition of the opposition, the strategies adopted by opposition groups, and the outcomes of these interactions.

In analyzing specific 'contentious episodes of autocratization', the method sheds light on the factors influencing opposition strategies and their effectiveness in slowing or halting autocratization. The paper diverges from traditional studies of autocratization, which often focus solely on cases of democratic breakdown (Ginsburg and Huq 2018). Instead, it allows to compare opposition to autocratic policy proposals across countries with varying levels of democracy, thus avoiding common challenges in autocratization research, such as selection bias and the lack of cross-regime comparisons. By doing so, the manuscript provides a broader and more nuanced understanding of how opposition movements function in different political contexts.

The paper presents the methodological framework for measuring opposition to autocratization across regimes and provides initial findings from five autocratic policy proposals before the collection of more comprehensive data. The manuscript discusses methodological challenges of the approach, as well as issues of data collection in different types of political regimes. Substantively, it contributes to the study of autocratization by broadening the scope of analysis to include not just cases of democratic failure, but also the dynamics of opposition to autocratic proposals in diverse political contexts. It highlights the importance of understanding the strategic choices of opposition groups and the broader political conditions that influence their success in challenging autocratization.

Social Inequality and Super-diversity: The impact of minority-contact on political preferences. A field experiment with text-as-data insights

S. Morgenstern[1], **R. Bach**[2]

T2

[1] Universität Mannheim
[2] MZES

How does diversity-outgroup contact shape social policy preferences? Despite the large number of studies on the well-known contact hypothesis, the current literature lacks experimental field evidence and behavioural results. In this project, I investigate this relationship in an online environment, a place where polarisation is particularly extreme. In collaboration with an NGO, English-speaking participants are randomly assigned to Zoom conversations with a person from a perceived minority group that differs in religion, gender, sexuality, ethnicity, physical or mental health. The results show that diversity-outgroup contact increases political solidarity with perceived minorities. Mediation investigations based on novel quantitative and qualitative textual analyses indicate a mechanism via the concept of cognitive liberalisation and reduced perceptions of social inequality. In a world of increasing social inequalities and super-diversity, this research project aims to contribute with a solution-oriented approach to social change.

Panel 2b: Networks, AI, and conflict analysis

Leveraging LinkedIn data for career trajectory clustering of interest group representatives

E. Wolfes-Wenker[1]

T2

Ruhr University Bochum

Professional social media platforms such as LinkedIn have become essential tools for relationship building and self-promotion across diverse industries and sectors. While research has rarely considered LinkedIn as a platform related to political communication (see as an exception Kreiss Jasinksi, 2016), the widespread presence of political actors such as interest group organizations and their staff provides an opportunity to gain unique insights into the composition and structure of political elites (e.g., Trampusch, 2005 for parliamentary careers). While data science methods effectively support data processing and visualization, researchers must ensure construct validity when using these techniques (Dorsey, 2020, p. 252), and the scraping of data from LinkedIn and other platforms in the post-API era presents legal and ethical challenges (e.g., Fei, 2024; Mancosu Vegetti, 2020). My starting point is the German lobby register, which gives me a lobbying population with interest group representatives for each of the organizations listed. I link this data to semi-structured data on career paths (Carter et al., 2009) derived from publicly available LinkedIn profiles. After reflecting on my approach to scraping LinkedIn profiles, I discuss the computational analysis of career trajectory information and present a framework for analyzing and comparing career pathways across different interest group types. My study employs text mining to extract information from a partially standardized corpus of career data, using hierarchical clustering to explore the structure and diversity of career trajectories among interest group representatives. I propose solutions for standardizing and automating heterogeneous career data using NLP techniques, while also discussing the potential incorporation of sequence analysis to examine longitudinal career patterns and network analysis to reveal structural patterns in the mobility network of interest group representatives.

Parameter Stability in Semi-parametric Neural Networks

K. Seng[1], **J. Garritzmann**[2], **D. Broska**[3]

[1] Zeppelin University Friedrichshafen
[2] Goethe Universität Frankfurt
[3] Stanford University

T2

Semi-parametric neural networks offer a novel possibility to analyze social science data with either discrete or continuous dependent variables that may appear in a cross-sectional or panel data context. The big advantage of these models is that they rely less on parametric assumptions like linearity or additivity and that the artificial neural network offers more flexibility that results in a better fit. The parametric part of these models allows for parameter estimates and their respective uncertainty which also makes statistical inference possible. This theory-driven perspective is dominant in comparative research and could become more relevant also in machine learning as machine learning techniques advance over time (Grimmer et al. 2021). Overall, these models produce better out-of-sample predictions than conventional parametric models. On the other hand, these models need relatively large amounts of data in order to produce stable coefficient estimates for the parametric part. Previous studies show that typical social science applications with roughly 1000 observations and a dozen of features (control variables) in the non-parametric part of the model lead to instability even with moderate sizes of the neural network (e.g. 5 hidden layers with 128-64-32-16-8 nodes). Changing the initial conditions might already lead to substantially different parametric results while the out-of-sample predictions remain stable. This makes the empirical assessment of theories and their respective hypotheses problematic if not impossible. Our contribution is to assess the parameter stability of semi-parametric neural networks with simulated and real data examples for different sample sizes and numbers of features in different programming environments (R and Python) in order to explore the conditions under which these models can be applied so far. This is especially relevant for applied researchers who still largely rely on parametric models when it comes to explanatory or confirmatory analyses and who could benefit from these new methods.

Bridging the Gap Between User and Hardware: Research-Ready AI Infrastructure for the Social Sciences

J. Bernauer[1]

T2

University of Mannheim

The advent of Artificial Intelligence (AI) shapes current infrastructural decisions and developments throughout universities, departments and institutes. Complementing centralised solutions such as bwHPC in Baden-Württemberg, single institutions or even chairs have started to build their own highly specialized GPU clusters to enable shortcuts to cutting-edge research relying on machine learning, large language models and beyond. In short, a diverse multi-level infrastructure is taking shape, requiring careful coordination and steering to find the optimal balance between decentralized and pooled resources for research. The presentation introduces an ongoing initiative for research-ready high performance AI infrastructure at the MZES, which has been ignited with a server featuring dedicated hardware (GPU). A use case is presented, predicting populism in political text using large language models on the MZES GPU cluster. Crucially, we propose to provide more than just the hardware, and namely build a long-term strategy involving trained staff and guidelines on methodological, ethical and ecological aspects as well as issues of data protection. In addition, further resources are planned to preserve the experiences of researchers as well as data sets and pre-trained models. In short, the goal is to bridge the remaining gap between highly skilled researchers and the hardware. We seek feedback from the methodological community on these considerations.

Why Discriminate? The Logic of Targeting in Gaza's Aerial Occupation

J. Hellinger[1], **R. Bhavnani**[2]

T2

[1] Geneva Graduate Institute
[2] IHEID

Withdrawing its settlements and ground forces from Gaza in 2005, Israel continued to exercise its control by means of an aerial occupation. Over the past 19 years, this strategy produced alternating cycles of aerial bombardment. Despite their sheer magnitude, there is however remarkably little understanding of their variation and spatiality across multiple episodes. Our analysis combines spatial data on aerial attacks from all major military operations in Gaza, including 2008-2009, 2014, and 2023-2024 with geo-coded data on casualties and socio-economic indicators. We find that aerial violence intensified in scale, duration, lethality, and destruction, while simultaneously becoming more indiscriminate over time, in spite of advancing Israeli military technology. This leads us to discuss the changing nature of violence in Gaza in relation to common but insufficient theoretical explanations for resorting to selective or indiscriminate violence, suggesting that previously unaccounted rationales are at play in Gaza. Across episodes, we also identify a socio-economic targeting bias: selective casualties cluster in areas characterized by higher measures of wealth and education, while indicators of poverty show the opposite effect. Over time, however, this spatial pattern of selective violence also fades out. Additionally, bombardment density varies as a function of topography and urban morphology, with sites located at higher altitudes or in more densely built-up areas being targeted more frequently. Despite the repeated wide-spread destruction and tens-of-thousands civilian casualties, we conclude that the aerial occupation of Gaza has failed, leading to a reverse towards a hybrid of aerial and terrestrial occupation as of October 2023.

Panel 3a: Groups, identities, and voting behavior

What Advances the Causes of Disadvantaged Groups in Identity Politics? Experimental Evidence on Particularistic and Universalistic Arguments

M. Kollberg[1], **P. Dinesen**[2], **K. Sonderskov**[3]

[1] Humboldt-Universität zu Berlin
[2] University of Copenhagen
[3] Aarhus University

T1

A core element of contemporary "identity politics" concerns rectifying injustices experienced by disadvantaged groups. Achieving this goal requires advocates of these groups to convince others – especially members of dominant groups – of their position. However, we know little about which arguments are the most effective in advancing the causes of disadvantaged groups. In this paper, we examine how different types of arguments disseminated on social media – a common battleground for debates about identity politics – affect citizens' preferences for measures that advance such causes as well as their evaluation of the affected groups. We identify two distinct lines of argument frequently used in discussions about identity politics, and theorize their persuasive efficacy on a given identity political issue as well as their broader effects on affect toward the disadvantaged group in question. More specifically, we examine particularistic arguments, which present a given issue as the result of oppression of a disadvantaged group by a dominant group, and therefore calls for policies rectifying or accommodating this injustice, and universalistic arguments, which construes accommodating policies as a matter of "common humanity" in terms of universal values of equality of opportunity and dignity of the individual. We test these two lines of argument in a large survey experiment conducted on a nationally representative sample of adults in the UK (n 12,000). We embed the arguments in thousands of different fictitious social media profiles to increase ecological validity. In the experiment, we randomly assign the type of argument (particularistic vs. universalistic (and a control case without an argument)) across eight issues frequently discussed in the context of identity politics. Further, to examine the possibility that the effect of each argument is modified by its source, we also randomly vary the sender of the argument (member of the disadvantaged group in question vs. member of the dominant group). Our results will have important implications for understanding the disagreements over contentious issues related to identity politics.

The Social Foundations of Political Hostility: Disentangling Political and Social Identities through a Factorial Survey Experiment

N. Buntfuß[1]

T2

TU Chemnitz

Increased hostility between supporters of different parties is often attributed to a rising importance of political identities, but it may be more deeply rooted in social structures. This phenomenon - referred to as social sorting - is increasingly being discussed as a driver of political polarization. This study seeks to expand our understanding of the extent to which negative affect is actually political and to which extent this political hostility has a socio-structural underpinning. In order to discriminate between the relative effects of different political and, in reality, often correlated social characteristics, I conducted a factorial survey experiment randomizing partisanship, issue positions, social class, gender, and, region in Germany. By showing respondents more or less "sorted" profiles, this study examines whether profiles with fewer cross-cutting attributes provoke greater negative affect and, in turn, create social distance. The findings will deepen our understanding of how political and social identities interact to shape affective polarization.

Are LLMs up to complex annotation tasks? The limits of generative AI for detecting references to social groups in multilingual texts

L. Erhard[1], P. Kaminski[1], Z. O. Dunivin[2], S. Šarenkapa[3], E. Deiss-Helbig[4], I. Guinaudeau[1], T. Matthieß[5], R. H. Heiberger[1]

T1

[1] University of Stuttgart
[2] University of California, Davis
[3] SciencesPo
[4] University of Konstanz
[5] University of Göttingen

This study examines how large language models (LLMs) align with human coders in detecting references to social groups within complex textual data. Social groups are often ambiguous, fluid constructs influenced by social and linguistic contexts, making their identification challenging and requiring long and complicated qualitative coding instructions to guide expert annotation. While supervised text classification relies heavily on human annotation, LLMs have recently demonstrated promising capabilities for automating tasks like summarization or translation without the need for large quantities of human-annotated data. However, we know relatively little about how well LLMs can handle intricate, contested concepts such as social groups. We present a novel dataset of newspaper sentences from German and French sources annotated by domain experts for references to social groups. Employing multiple prompt configurations and comparing the resulting LLM-based annotations to human-coded benchmarks, we measure inter-annotator agreement (Krippendorff's α) and analyze descriptor matches using a measure of partial overlap, as group labels may span several tokens. Our findings show that GPT-4 can reliably identify sentences without social groups. However, while the model has moderate agreement with positive instances of group labels, performance suffers on increasingly complex or ambiguous instances, and the model over-identifies group references. Supporting our claim that the LLM struggles with increasing conceptual and linguistic ambiguity, LLM agreement with human coders increases as human-human inter-annotator agreement increases. Although improved prompt engineering and larger models enhance performance slightly, fundamental difficulties remain. These results indicate that LLMs can aid large-scale annotation by filtering out texts lacking social groups, reducing human workload. Nevertheless, researchers must remain cautious and sentences containing references to social groups should be either hand-coded by researchers or by models fine-tuned for their specific task.

Voter Overrepresentation and Vote Overreporting in Election Surveys: Patterns and Trends Across Eight Decades of Vote Validation Research

P. Selb[1], **J. Wiederspohn**[1]

T2

University of Konstanz

Election surveys grossly overestimate turnout. Two distinct phenomena contribute to this discrepancy: overrepresentation of actual voters and vote misreporting by actual nonvoters among survey respondents, both of which have the potential to severely bias inferences about political behavior. Vote validation studies have been conducted to disentangle and scrutinize these errors. We report results from a meta analysis of 78 validation studies from six countries, 1956-2023. Our focus is on broad trends and their implications for survey data quality, such as falling response rates, increased use of online access panels as sampling frames, and declining election turnout.

Panel 3b: All politics is local

Community-wide effects of a randomized citizen council

S. Ziaja[1], **A. May**[2]

T2

[1] GESIS - Leibniz Institute for the Social Sciences
[2] University of Mannheim

This study implements a real-world participation experiment and determines the treatment's effect on the political climate in a rural municipality in North Rhine-Westphalia, Germany. The treatment is a citizen council with randomly invited participants. The citizen council's topic is the future of the run down local event hall. The citizen council will act in an advisory capacity in this process and will be flanked by participatory tools such as "opinion machines" and social media exchange. The citizen council is hosted by the mayor and the municipal council and organized by Initiative Offene Gesellschaft, an association specializing in innovative participation processes. The citizen council will conclude in February with a presentation at the municipal council. To examine the citizen council's effects on the community, we conduct surveys among all adult residents of the municipality (N = 10,500), both before and after the implementation of the citizens' council. Our first (baseline) survey wave was completed in August 2024 and yielded a response rate of around 10 percent (N = 1,080). We expect to gain insights on the causal effects of the council on political attitudes by leveraging three levels of treatment, i.e., levels of exposure to the citizen council: (1) randomly assigned invitations to declare one's willingness to participate in the citizen council (N = 3,500), (2) family and friends of the 30 actual final participants of the citizen council (N 600), and (3) the entire population of the municipality compared with a synthetic control generated from two parallel waves of the German Longitudinal Election Study (GLES). Our second wave will start on 28 February and close on 6 April 2025. We will be able to present preliminary results at the DVPW methods meeting, based on the survey responses that will have been submitted at the time. Our presentation will focus on building a synthetic control with GLES data. The study has been preregistered.

The Effects of Participatory Democracy on Political Attitudes: Causal Evidence from Citizens' Assemblies

M. C. Duarte[1], **J. M. Fernandes**[2], **M. M. Pereira**[3]

[1] University of Lisbon
[2] CSIC
[3] LSE

T1

Contemporary democracies face growing contestation. The populist wave of the early XXI century feeds on citizens' frustrations with the status quo. Governments worldwide have tried to revert this trend by promoting participatory governance initiatives. We test the ability of citizen assemblies to foster democratic attitudes by leveraging the random selection of participants in Lisbon's citizen assemblies. We rely on data from Lisbon's Conselho de Cidadãos, held over two consecutive weekends in April 2024. Participants discussed five local welfare issues: housing, immigration, homelessness, health, and intergenerational solidarity. Lisbon City Hall sent 50,000 physical invitation letters to a representative sample of the population, inviting them to register for the event. A total of 1,243 Lisbon residents registered. A sample of 50 participants was randomly selected from this list through a lottery process. We surveyed participants the day before the 2024 Lisbon citizens' assembly and the week after the event. The same post-treatment survey was also fielded with non-participants who registered for the event (control group). We measured six outcomes: issue salience, issue knowledge, internal political efficacy, external political efficacy, green lantern theory, and political trust. We isolate the effects of participation in the assembly in two ways. First, a between-subjects design compares randomly selected participants with non-participants after the event (N = 533). Second, a within-subjects design allows us to compare outcomes among participants before and after the citizens' assembly (N = 94). Results suggest that participation in two full-day sessions of a citizen assembly increased external political efficacy and political knowledge and raised the salience of the issues discussed in the meeting. However, we find no evidence that these effects spill over to institutional trust or help citizens better recognize the challenges of the policymaking process. The study reveals how participatory governance initiatives can bolster the connection between citizens and the democratic process.

Challenging the Challengers? The Causal Implications of the Bündnis Sahra Wagenknecht on German Party Competition

S. Eschenwecker[1], **T. König**[1]

T1

University of Mannheim

Elections are ultimately a zero-sum game, where the entry of a new party inevitably leads to electoral losses for some established parties. This article examines the causal effect of a populist challenger's candidacy on party competition, analyzing whether it primarily mobilizes formerly abstaining voters, exacerbates the decline of mainstream parties, or attenuates the rise of other challenger parties. Leveraging a natural experiment coinciding with the 2024 European election, we study County Council elections in five German states, where the newly founded Bündnis Sahra Wagenknecht (BSW) contested in some counties but not others. Using covariate balancing propensity scores, we weight municipalities on key sociodemographic, economic, and political covariates, ensuring credible as-if randomization to estimate the average treatment effect of BSW candidacy on turnout and established party vote shares. Our robust findings, generalizable beyond the local context, reveal a significant reduction in vote shares for all established parties except the right-wing populist Alternative für Deutschland (AfD), with no evidence of increased voter mobilization due to BSW's entry. These results have important implications for party competition dynamics and the functioning of representative democracy in the presence of multiple populist challengers.

Irregular immigration and the reintroduction of national border controls: Evidence from Germany

K. Lindemann[1], **L. Schramm**[2]

[1] University of Mannheim
[2] Ludwig Maximilians University of Munich

T2

Recently, several member states of the Schengen free-travel area have introduced temporary controls at their national borders as a means to control and limit irregular immigration. Germany's decision, in the summer of 2024, to reinstate checks at all its borders has led other member states to follow suit, thereby threatening the future of free traveling inside the Schengen area. While political elites claim that such measures reflect public demands, it remains unclear how supportive citizens actually are. In this paper, we therefore investigate public opinion of border controls. We expect that citizens are more likely to support border controls if they experience minimal personal consequences. By contrast, citizens are likely more reluctant if significant individual-level consequences arise, such as travel delays or disruptions in their daily work commute. Collecting original data representative for the German population (N = 2,500), we first assess support for border controls and the EU more generally with a battery of questions. In a second step, we investigate the underlying mechanisms by focusing on the trade-offs that citizens are willing to accept for stricter immigration policies by conducting an original survey experiment. Our findings have implications not only for discrepancies between elite communication and public opinion, but also for immigration policies and the future of European integration.

Panel 4a: Images as data

AI-Generated Visuals: A Solution for Image Demand in Social Science Research

I. E. Acquarone[1], **A. C. Cinar**[2], **G. D. Landro**[3], **N. Zilberstein**[3]

T2

[1] Arizona State University
[2] Universidad Carlos III - Madrid
[3] Rice University

Human interactions with the world have predominantly become mediated by visuals, leading to a growing demand for images in social science research. We explore the potential of AI-generated visuals to meet this demand. Using real images of British parliament members—one of the world's largest national assemblies—and a stable diffusion model, we generate highly realistic synthetic representations of these legislators. Through two randomized control trials conducted online in the United Kingdom and the United States, we test the comparability of real and synthetic images in terms of perceptions of reality and relevant political attributes. Our results show that real and synthetic images are perceived similarly, with no significant differences in citizens' evaluations of legislators' attributes. This demonstrates the suitability and usefulness of AI-generated images for further research. We extend our work to various contexts and provide labeled visual data of potential legislators worldwide, offering a valuable resource to advance research in identity and politics within the social and experimental sciences.

Manipulating Protest Size Estimates with AI-generated Images

S. Scholz[1], **N. Weidmann**[1]

T1

University of Konstanz

The size of the crowd at large political protests is an extremely political number, since it is usually interpreted as the level of popular support for a regime or a certain political issue. This is particularly the case in autocracies, where protest is one of the few ways for citizens to express their political preferences. Not surprisingly, it is in the interest of political actors to manipulate perceptions of the crowd size, in order to boost support for, or to delegitimize, a particular political issue. In this paper, we study how perceptions of the crowd size can be manipulated with AI-modified protest images. This is one of the more subtle ways in which AI-generated visual content can be used, since it is very difficult to verify the authenticity of these images. We develop an experimental setup where participants are asked to rate the size of the protest crowd from a series of social media images. These images are taken from actual protest events. We study the effect of two experimental manipulations. First, we modify some of the images to show a larger (or smaller) crowd size. This is done by automatically recognizing protesters on protest images and inserting more people around them with segmentation and latent diffusion models. We also use the same models to remove parts of the recognized protesters from the protest images. Second, we include a warning in the experimental setup to alert participants about the possibility that images on social media can be subject to manipulation. We expect that while the manipulated images will lead to higher (or lower) crowd estimates as compared to the unmodified images, a warning about image manipulation will reduce the magnitude of this effect, since it makes observers more alert. Our results gauge the reception of AI-generated visual content by observers on social media, and assess how easy or difficult it is to affect popular perceptions with these images.

Structuring Quantitative Image Analysis with Object Prominence

C. Arnold[1], **A. Küpfer**[2]

[1] University of Birmingham
[2] TU Darmstadt

T1

When photographers and other editors of image material produce an image, they make a statement about what matters by situating some objects in the foreground and others in the background. While this prominence of objects is a key analytical category to qualitative scholars, recent quantitative approaches to automated image analysis have not yet made this important distinction but treat all areas of an image similarly. We suggest carefully considering objects' prominence as an essential step in analyzing images as data. Its modeling requires defining an object and operationalizing and measuring how much attention a human eye would pay. Our approach combines qualitative analyses with the scalability of quantitative approaches. Exemplifying object prominence with different implementations—object size and centeredness, the pixels' image depth, and salient image regions—we showcase the usefulness of our approach with two applications. First, we scale the ideology of eight US newspapers based on images. Second, we analyze the prominence of women in the campaign videos of the U.S. presidential races in 2016 and 2020. We hope that our article helps all keen to study image data in a conceptually meaningful way at scale.

Tactical Diversity Within Protest Events: An image-to-data analysis

A. Aísa[1], **B. González**[2], **S. Scholz**[1], **C. v. Soest**[3], **N. B. Weidmann**[1]

[1] University of Konstanz
[2] University of Konstanz ; University of Zurich
[3] German Institute for Global and Area Studies

T2

How can we identify and track violent tactics within protest events? Protests are usually classified as peaceful or violent based on the most salient form of participants' behaviour. However, we do not know much about what else happens within protest events that may eventually lead to the appearance of violent tactics. We present a spectrum of violent tactics that participants can use to show dissent against the government. We argue that while some protests display a consistent preference for peaceful or violent resistance from the outset, others show diversity in their behaviour, adding uncertainty to their overall outcome. Harnessing a new image-to-data methodology, we analyze 15,567 protest images across 10 countries between 2016 and 2021. In doing so, we provide two contributions to the field of organized dissent. First, we explain how within protest violence can be identified using the composition of particular objects. We do so by leveraging the information of over 8.1 million objects detected on the protest images using the Large Vocabulary Instance Segmentation (LVIS) dataset. Second, we extract and classify image-specific features on a diverse pool of participant behaviour over our entire sample, allowing us to recognize violent events using images. Overall, this new data collection effort and analysis help us gain insights into the inner workings of protest events.

Panel 4b: Survey experiments

Detecting heterogeneity in conjoint experiments on arms trade preferences in 5 major Western export countries: An application of two recent methods using survey data

F. Haggerty[1], **P. W. Thurner**[1], **L. Rudolph**[2]

T1

[1] LMU Munich
[2] University of Konstanz

Despite significant politicization and heated public debates in some major exporting countries, systematic cross-country research on public support for arms trade is currently lacking. We go beyond the usual single item approach and implement conjoint experiment in major Western export countries (France, Germany, Italy, UK, USA). This allows us to elicit differentiated compound attitudes for varying combinations of background conditions, which are derived from a political economy approach to arms transfers. More specifically, economic considerations are matched with various security-related and norm-related aspects. To investigate underlying causes of the heterogeneity of effects, we use recently proposed (machine learning) approaches from Goplerud, Imai and Pashley (2024) and Robinson and Duch (2024). The former proposes a Bayesian mixture of regularized logistic regressions to identify groups of units that show similar patterns of treatment effects. The later employs Bayesian additive regression trees (BART). We draw on an original population-representative survey with embedded experiments (ca. N = 10,000), fielded in June-August 2023. Despite predominant similarities in the fundamental calculus, with normative considerations trumping strategic aspects themselves followed by economic criteria we are able to detect remarkable differences between the five countries with respect to strategic aspects and principal considerations, and a major divide between Germany and Italy on one side and France, the UK and the US on the other. We ascribe this to varying strategic cultures of these countries with their different historical paths and their responsibilities as nuclear powers.

(What) is it worth to employ a serious game? An experimental analysis of information and efficacy in political opinion formation

W. El-Ajou[1], **G. Brückmann**[1], **I. Stadelmann-Steffen**[1]

T1

University of Bern

With more and more challenges to democracy, textual information treatments often employed by political scientists often do not alter politically relevant outcomes, leading scholars, funders and other stakeholders to turn to new, more. immersive treatments. Among these suggested new treatments are serious games. In this paper, our main goal is to test the extent to which serious games can influence individuals' opinion formation on complex issues, here, exemplified by the energy transition. Theoretically, we build on insights from opinion formation research, the knowledge deficit model, and the role of information to argue that a serious game, due to its immersive nature, may be more effective than more conventional information treatments in triggering effects on perceived efficacy and ultimately influencing domain-specific policy opinions. For this, we co-created a full-fledged serious game, developed in collaboration between social scientists, energy system modelers, and game designers, as an experimental treatment. We implement this experimental treatment in a large-scale Swiss population survey. As not everyone assigned to gaming will play, we estimate local average treatment effects for compliers, using an instrumental variable design combined with a pre-post design and block randomization to ensure high statistical power. Results inform other researchers if expensive serious games alter efficacy and policy acceptance in the population at large.

Validating Stated against Actual Baviour in Experiments – Comparing Vignette and Quasi-Experimental Evidence from Bangladesh

L. Rudolph[1]

T1

University of Konstanz

To what extent do replies to hypothetical survey-experimental scenarios replicate in real-world behavior? To this end, we study a population in rural Bangladesh along the banks of the Jamuna river, at risk of being affected by riverbank erosion and floods during the monsoon season in July-September each year. Leveraging a large-scale panel survey with 1600 household heads, we compare the same respondents' on two dimensons: First, replies to questions on hypothetical movement behavior if affected by or at risk of scenarios of natural disaster affectedness in June 2021 (pre-monsoon); second, probabilities of movement intentions and actual moves 6 months later (January 2022), after part of the sample has actually, quasi-experimentally, been affected by such events. Given the evidence base for the actual external validity of responses to survey-experiments is slim, this study contributes a crucial comparison from a rarely studied developing-country context on the extent to which survey-experimental estimates can replicate actual behavior.

When statistical truth serum is not enough – Lessons from a failed list experiment to measure corruption in Ukraine

C. Sajons[1], **B. Müller**[1], **M. Samoiliuk**[2]

T1

[1] University of Mannheim
[2] Centre for Economic Strategy, Kyiv

List experiments are designed to elicit truthful responses to sensitive topics by providing a great degree of anonymity to the respondents. However, they rely on the assumption that respondents do not react to the survey design. In this paper, we report the results of a list experiment that aimed to measure the extent of corruption and bribery in the application process for an international support program for small and medium enterprises (SMEs) in Ukraine. The experiment fails a critical validity check and produces unrealistic results. Closer inspection of the data and information from a follow-up survey among participants indicates that a significant share of program beneficiaries may have intentionally misstated their experiences in the application process. The results suggest that applying list experiments to detect illegal behavior with possible adverse consequences for the participants may not be equally effective as in cases of attitudes, experiences, and behaviors that may be socially unaccepted, embarrassing, stigmatized, or even traumatic, but do not provoke legal repercussions.

Friday, March 28

Panel 5a: Measuring positions and salience from text

Favor Exchange between Government and Media Owners

B. Zafer[1]

T1

University of Rochester

Scholars have extensively investigated the interest of incumbents in controlling media and the crucial role of financial independence in safeguarding press freedom. Nonetheless, a gap in the literature persists in providing evidence on how governments employ ostensibly legal methods to supply financial resources to connected media in exchange for favorable coverage. This paper leverages cutting-edge NLP and causal estimation methods to identify market-based exchanges as two-way favors between the government and media owners, exploiting the acquisition of Dogan Media Group, the largest media company in Turkey, by a pro-government conglomerate in 2018. First, I use the triple-differences (DDD) regression fixed-effects estimator to identify state advertising allocations as politically motivated favors distributed to the connected outlets. Using private advertising as a "placebo stratum" in a DDD design, I rule out conventional economic and market-based explanations for the choice of advertising platforms and identify the effect of political incentives. I show that the 2018 acquisition led to a significant increase in state advertising in newspapers previously owned by Dogan Media despite a simultaneous decline in their circulation and private advertising. Second, I analyze the coverage in these newspapers leveraging a large data set of online news retrieved from the CommonCrawl repository. I utilize HTML sections and fine-tune a Bidirectional Encoder Representations from Transformers (BERT) model to select my corpus and construct a measure of pro-government bias in the coverage of domestic economic news using GPT. My analysis shows that newspapers sold to the pro-government conglomerate positively bias their coverage of the state of the Turkish economy. My results establish that state advertising in Turkey is politically driven and incompatible with economic incentives, and the connected newspapers cater to the government, which can be at the expense of circulation.

From Weather to Words – The Effect of Extreme Weather on Political Elite Communication

J. Haak[1], **S. Morét**[1], **L. Schwarz**[2]

[1] University of Hamburg
[2] Johannes Gutenberg-Universität Mainz

T1

How does exposure to abnormal and extreme weather affect political behavior of political elites? While previous studies suggest that weather exposure affects climate attitudes among the public, little is known about how political elites react to such events. We address this gap by investigating how individual party politicians adjust their attention to climate change issues in their communication in response to weather exposure. We argue that candidates have an incentive to strategically address the issue of climate change following abnormal and extreme weather in their constituency, as these events are locally confined and can heighten the salience of climate change for voters. Social media platforms are crucial in this context, allowing politicians to react in real-time and bypass the delays typical of traditional press releases. Our analysis builds on over 250.000 tweets from more than 550 candidates or MPs of the German Bundestag throughout 2021 and follows three steps: First, we use a finetuned state-of-the art large language model (XLM-RoBERTa) to filter climate change-related statements. Then we employ a novel stepwise zero-shot classification approach using GPT4 models guided by a custom codebook to identify mentions of actions to address climate change, along with their sentiment. Finally, we link these Twitter data with geo-referenced weather data. Thus, we can show whether candidates have strategically exploited weather events to advance their (anti-)environmental agenda.

Improving Cross-Domain Text Classification When Measuring Multi-dimensional Party Positions

H. H. V. Nguyen[1]

T1

Deutsches Institut für Erwachsenenbildung

Measuring party positions from texts often involves simplifying politics into low-dimensional spaces, which may fail to capture the complexities of party competition in elections and policymaking. The most viable approach to measuring high-dimensional party positions is grouping documents by policy dimensions before scaling, yet this approach is challenging due to annotation scarcity and performance drops in cross-domain classification. This article proposes a cross-domain classification pipeline combining a series of domain adaptation techniques with pseudo-labelling to enhance text classification in cross-domain settings. The proposed pipeline demonstrates significant performance gains across various policy topic classification tasks in political science, especially when the baseline model exhibits strong reliability. An ablation study highlights the components of the pipeline that are most critical for effective cross-domain classification. Implementing this pipeline to measure multidimensional positions of German political parties from manifestos and parliamentary speeches reveals the intricate dynamics of electoral competition and policymaking which are often overlooked by traditional low-dimensional approaches.

Applying Large Langue Models for the Measure of Populism in Election Manifestos

L. Siefken[1]

T1

Universität Mannheim

In recent years, populist parties have witnessed a surge in popularity within Western democracies. A valid measure that can capture Populism's multi-dimensionality, latency, and non-compensatory nature in a fully automated way is yet to be proposed. This thesis aims to evaluate the applicability of Large Language Models (LLMs) for the automated text analysis of populist rhetoric. More specifically, I employ Google's Gemini AI to classify the paragraphs of German election manifestos from five national-level elections. Investigating the classification reveals that Gemini sufficiently overlaps with the work done by a human coder and significantly outperforms dictionary analysis. LLMs are revealed to be a promising fully automated method for the measure of ideational populism in text data. When used in an analysis, Gemini's coding comes to the same conclusions as the human-coded scores, showing that conveying a populist message correlates with the use of simplistic language.

Panel 5b: Parliamentary debate

Only Hot Air? National Parliamentary Discourse and Its Effect on Public Opinion on the European Union

M. C. Duarte[1], **R. Kittel**[2]

T1

[1] University of Lisbon
[2] Freie Universität Berlin WZB

The European Union (EU) is nowadays a hot issue in national politics. As a result of years of intense crises, such as the Eurocrisis, the refugee crisis or, more recently, the war in Ukraine, the EU and the process of European decision-making have become objects of contestation in national parliaments. Existing studies have unveiled that mainstream political parties adopt pro-EU views, whereas extremist political parties defend anti-EU stances. How national parliamentary discourse on the EU affects public opinion, however, remains unknown. Often, individuals have very little knowledge about the EU. Evidence demonstrates that EU politics are perceived as more complex and technical than national politics. Parliaments, the biggest arenas for deliberation and communication, can contribute to the dissemination of information and arguments that foster individuals to formulate their political preferences. Thus, the national parliamentary discourse can, arguably, influence public opinion on the EU. We hypothesise that when radical parties talk significantly more and in a negative tone about the EU, public opinion towards the EU becomes more negative. However, we expect radical right voters to exert a bigger influence over the EU attitudes of diverse types of voters. By looking at parliamentary speech data and individual-level data from eight EU countries between 2000 and 2021, we analyse how public opinion on the EU is shaped by the national parliamentary discourse. Methodologically, we first apply multilingual manifestoberta models to identify EU-related debates. Then we run a transformer model on these pre-selected debates to detect the emotions employed in EU-related discourse. Later we match these data points with public opinion data on the EU.

Distinct Rhetoric? Predicting Voting Defection from Parliamentary Speech Using Language Models and Machine Learning

V. Kunz[1]

T1

GESIS - Leibniz Institute for the Social Sciences

Do legislators discuss votes on the parliamentary floor differently when they defect from their party principals than when they are loyal? Plenary debates are generally considered a central arena for legislators to exchange arguments, appeal to voters, and justify their positions. Adopting an approach suggested by Peterson and Spirling (2018) that uses machine learning model performance as a substantive quantity of interest, I assess the distinctiveness of parliamentary speechmaking. More specifically, I compare how well different combinations of text representations (e.g. bag-of-words, embeddings) and machine learning algorithms (e.g. regularized regression, transformers) are able to predict whether a plenary debate contribution comes from a defecting or loyal legislator. When the algorithm performs well at distinguishing these speeches, this suggests that legislators focus on different aspects or use different rhetoric in response to their voting decision when taking the plenary floor. Empirically, I draw on an original data set of roll-call votes and associated multilingual debate contributions from the 7th, 8th, and 9th European Parliament (2009-2023). Results from these supervised classification tasks suggest that parliamentary speech differs depending on whether legislators defect from their party principals or not. Substantially, the findings contribute to our understanding of how legislators manage intra-party conflict, with implications for political representation. Methodologically, they contribute to evaluating the capacity of different text representations and models to generate valid measurements in a multilingual context.

Measuring 'Democratic Health' in Political Text: Evidence from 150 years of German Parliamentary Debates

T. Widmann[1]

T1

Aarhus University

Politicians and parties' behavior has significant implications for the health of democracies. We can thus learn about the state of democratic health by analyzing how elites talk about democracy and its subcomponents. This paper introduces a novel approach, using 150 years of German parliamentary debates and word embedding methods to measure and track how political language reflects democratic principles. I construct a continuous, text-based measure of democratic health and compare it with expert-led indicators, notably the V-Dem dataset. The results show that shifts in democratic discourse correspond with key historical events, including authoritarian breakdowns and democratization periods. Unlike expert surveys, which are often static and delayed, this method can capture real-time changes in elite rhetoric. Importantly, the most recent years reveal a worrisome trend of democratic backsliding. This underscores the potential of text-as-data techniques for providing timely, dynamic, and granular insights into evolving democratic norms.

Linking Textual Data - Potentials and Pitfalls of using Named Entity Linking and Record Linkage in Political Science Research

C. Leonhardt[1], **A. Blätte**[1]

T1

University of Duisburg-Essen

This paper presents how the approach of Named Entity Linking (NEL) can be used to robustly link textual data with structured data such as administrative data. The potential for substantive political science research and methodological pitfalls are our specific concern. Based on the automated identification, disambiguation and linking of mentions of locations and organizations in parliamentary speech-making, the discourse on the production and consumption of coal in the German Bundestag will be analyzed. With NEL, references to entities identified in unstructured textual data can be linked to structured (administrative) data via unique identifiers to gain insights into the characteristics of the entities referred to. This can, for example, shed light on the question of which regional-level characteristics (e.g., economic factors) might contribute to specific locations being mentioned in particular contexts (e.g., topic or sentiment). In our use case, we focus on parliamentary debates on the coal phase-out (2018-2024). We describe variations in reference-making in different political contexts. This allows for a discussion of domain-specific challenges of the approach while addressing a substantively relevant research question. Unlocking the potential of NEL for political science research follows the intuition that linking unstructured data (e.g., textual data) and structured data (e.g., administrative data or survey data) offers possibilities which are currently utilized only sporadically in substantive research. This seems particularly relevant as quantitative analyses of textual data have become common in the field. If these approaches can be combined with findings made possible by the analysis of established forms of structured data with its rich methodological tradition in social science research, new research questions can be addressed. We argue that NEL can be an important method for realizing this robustly. Its success, however, depends not only on its technical performance, but also on its usability in domain-specific contexts. Our R package dbpedia developed in the context of the measure “Linking Textual Data” of KonsortSWD is thus part of our contribution.

Panel 6: Studying public opinion

AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers

A. Wuttke[1], **M. Aßenmacher**[1], **C. Klamm**[2], **M. M. Lang**[3], **Q. Würschinger**[1], **F. Kreuter**[1]

T1

[1] LMU Munich
[2] University of Mannheim
[3] University of Oxford

Traditional methods for eliciting people's opinions face a trade-off between depth and scale: structured surveys enable large-scale data collection but limit respondents' ability to express unanticipated thoughts in their own words, while conversational interviews provide deeper insights but are resource-intensive. This study explores the potential of replacing human interviewers with large language models (LLMs) to conduct scalable conversational interviews. Our goal is to assess the performance of AI Conversational Interviewing and to identify opportunities for improvement in a controlled environment. We conducted a small-scale, in-depth study with university students who were randomly assigned to be interviewed by either AI or human interviewers, both employing identical questionnaires on political topics. Various quantitative and qualitative measures assessed interviewer adherence to guidelines, response quality, participant engagement, and overall interview efficacy. The findings indicate the viability of AI Conversational Interviewing in producing quality data comparable to traditional methods, with the added benefit of scalability. Based on our experiences, we present specific recommendations for effective implementation.

Assessing Policy-Driven Shifts in Public Opinion Across Nations: The Challenges of Staggered Adoption, Measurement Error, and Limited Sample Size

P. Koc[1], **M. Steenbergen**[2]

T1

[1] GESIS
[2] University of Zurich

This paper proposes a method to investigate how public opinion responds to policy changes. Existing approaches have often relied on two-way fixed effects or hierarchical models with the Mundlak device. However, these methods can yield biased estimates when the timing of treatment varies across units. In response, new difference-in-differences estimators have been introduced to address the problem of staggered adoption. While these new approaches are flexible, they have not tackled two additional challenges commonly encountered when studying policy effects on public opinion across nations: first, that public opinion is not directly observed and must be inferred from noisy measurements; second, that cohorts defined by the policy introduction year can be extremely small, sometimes consisting of a single country. To address these issues, we propose a Bayesian hierarchical model that explicitly accounts for known measurement error and partially mitigates small-sample concerns. We detail the model's underlying assumptions and limitations, and then present a small simulation study demonstrating its performance. Finally, we apply our approach to the case of same-sex couples' rights in Europe, illustrating how it can improve inference about the relationship between policy changes and shifts in public opinion.

Shifting Ground: The 2023 Earthquake and Electoral Accountability in Turkey

K. Bogatyrev[1]

T1

Bocconi University

Do crises and external shocks undermine public support for authoritarian governments? This paper examines whether natural disasters trigger a public backlash against authoritarian incumbents, focusing on the February 2023 earthquakes in Turkey. The earthquakes caused unprecedented human and material losses, sparking public outrage over the Turkish government's perceived failure to prepare and respond effectively. With national elections held three months later and local elections a year later, this context provides a unique opportunity to test how natural disasters affect public accountability in an authoritarian setting. Electoral outcomes across 45,000 Turkish localities were analyzed, combining georeferenced earthquake intensity data with official electoral results. Using causal panel data approaches, pre-registered hypotheses about the earthquake's effect on incumbent vote share and turnout were tested. Additional analyses merged electoral and seismological data with damage estimates and socioeconomic statistics. Findings show that voters hold authoritarian incumbents accountable for natural disasters, though the effects are complex and heterogeneous. In the 2023 national elections, the earthquake reduced AKP support in affected regions. By contrast, the 2024 local elections showed a weaker punishment effect. Turnout exhibited a parabolic response: moderately hit areas saw higher mobilization, while severely affected localities experienced demobilization, possibly due to displacement. This study contributes to research on public accountability in non-democracies and the electoral consequences of crises. It demonstrates that even in competitive authoritarian regimes, natural disasters can expose government failures and trigger backlash. The magnitude and duration of this effect depend on disaster intensity, government responses, and the political landscape. These findings highlight the need to account for effect heterogeneities in disaster studies and caution against generalizing across disaster severities.

The Tests of Time and Fluid Concepts: An Empirical Validation of Identity Measures Across Regime Shifts

M. Shiraef[1]

T1

University of Mannheim

What is the most durable measure of identity over an individual's lifetime? This paper presents a novel methodological approach to investigating identity dynamics within ethnic minority families across regime-level shifts. Drawing on extensive fieldwork and historical research in southern Albania, I examine identity decisions over critical periods, including state formation (1900-1913), two world wars (1914-1945), communist regime installation (1946-1989), the fall of communism (1990-1991), and post-communist transition (1992-2004). By combining detailed state censuses, cemetery inscriptions, and personal testimonies—and comparing for consistencies—the study evaluates identity measures' durability and variability across political transitions. The results reveal that first names on graves align consistently with state-level census data, serving as a robust indicator of chosen identity, regardless of regime type or time period of the record, while religious affiliations and script use demonstrate greater variability. Methodologically, the paper links family-level identity decisions to state-level records, offering a unique perspective on identity persistence under one of the world's most tightly controlled authoritarian regimes. This approach contributes to the field of political science by introducing cemetery data as an innovative and underutilized resource for examining identity dynamics over time. By bridging qualitative testimonies with computational text analysis, this research advances mixed-methods applications and highlights the empirical potential of unconventional data sources for understanding state-family interactions and identity persistence. These findings provide a replicable framework for studying identity in diverse political contexts and suggest broader implications for the practicalities of combining computational social science, historical analysis, and ethnographic methods.

List of Participants

Acquarone, Iris E.	Presenting author	Arizona State University
Aísa, Alejandro	Presenting author	University of Konstanz
Arnold, Christian	Presenting author	University of Birmingham
Bach, Ruben	Discussant	University of Mannheim
Bernauer, Julian	Presenting author	University of Mannheim
Bogatyrev, Konstantin	Presenting author	Bocconi University
Braun, Laura	Presenting author	University of the Bundeswehr
Brückmann, Gracia	Presenting author	University of Bern
Buntfuß, Nelly	Presenting author	TU Chemnitz
Cohen, Denis	Section speaker & local organizer	University of Mannheim
Duarte, Mariana Carmo	Presenting author	University of Lisbon
Eschenwecker, Stefan	Presenting author	University of Mannheim
Gessler, Theresa	Section speaker & co-organizer	Europa-Universität Viadrina Frankfurt (Oder)
Haak, Jessica	Presenting author	University of Hamburg
Haggerty, Fabian	Presenting author	LMU Munich
Hellinger, Joshua	Presenting author	Geneva Graduate Institute
Kaminski, Patrick	Presenting author	University of Stuttgart
Kittel, Rebecca	Presenting author	Freie Universität Berlin & WZB
Koc, Piotr	Presenting author	GESIS - Leibniz Institute for the Social Sciences
Kollberg, Markus	Presenting author	Humboldt-Universität zu Berlin
Kunz, Verena	Presenting author	GESIS - Leibniz Institute for the Social Sciences
Leininger, Arndt	Section Speaker & co-organizer	TU Chemnitz
Leonhardt, Christoph	Presenting author	University of Duisburg-Essen
Licht, Hauke	Presenting author	University of Innsbruck
Lindemann, Korinna	Presenting author	University of Mannheim
Morét, Sophia	Presenting author	University of Hamburg
Morgenstern, Sandra	Presenting author	University of Mannheim
Nguyen, Hung H. V.	Presenting author	Deutsches Institut für Erwachsenenbildung

Passerotti, Rémi	Presenting author	Sciences Po Grenoble - Université Grenoble Alpes, PACTE
Rajski, Hannah	Local organizer	University of Mannheim
Ratkovic, Marc	Presenting author	University of Mannheim
Rittmann, Oliver	Local organizer	University of Mannheim
Rudolph, Lukas	Presenting author	University of Konstanz
Sajons, Christoph	Presenting author	University of Mannheim
Sangiovanni, Stefano	Presenting author	University of Milan
Scholz, Stefan	Presenting author	University of Konstanz
Selb, Peter	Presenting author	University of Konstanz
Seng, Kilian	Presenting author	Zeppelin University Friedrichshafen
Sen, Indira	Presenting author	University of Mannheim
Shiraef, Mary	Presenting author	University of Mannheim
Siefken, Leon	Presenting author	University of Mannheim
Weidmann, Nils	Presenting author	University of Konstanz
Wenz, Alexander	Discussant	University of Mannheim
Widmann, Tobias	Presenting author	Aarhus University
Wolfes-Wenker, Erik	Presenting author	Ruhr University Bochum
Wuttke, Alexander	Presenting author	LMU Munich
Zafer, Bahar	Presenting author	University of Rochester
Ziaja, Sebastian	Presenting author	GESIS - Leibniz Institute for the Social Sciences

Locations and Useful Information

Please click on the highlighted links below to view the locations and routes on Google Maps.

Hotel and conference venue

- The **conference hotel** is InterCity Hotel Mannheim Mitte, Ecke L13, Schlossgartenstraße 1, 68161 Mannheim. It is located directly next to Mannheim Hbf.

- **Presentations** and the **DVPW Section Members' Meeting** will be held at the Mannheim Centre for European Social Research (MZES), A5, 6, Bauteil A, 68159 Mannheim. The venue is a 20 minute walk, or a 7 minute bus ride from Mannheim Hbf and the conference hotel.

Meals

- The **pre-conference dinner** on Wednesday night will be held at HiKoo Noodle Kitchen, M2 3, 68161 Mannheim. It is a 10 minute walk from Mannheim Hbf and the conference hotel. The menu can be viewed here.

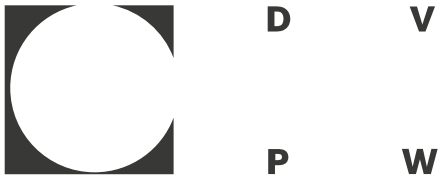
- The **conference lunch** on Thursday will be provided **on-site** at the conference venue by our caterer Kleines Café. Vegetarian and vegan options will be provided.

- The **conference dinner** on Thursday night will be held at St. James, Hafenstraße 25-27, 68159 Mannheim. It is a 15 minute walk from the conference venue, and a 30 minute walk or 15 minutes bus ride from/to the conference hotel. Participants can choose between an omnivore, a vegetarian, and a vegan variant of a three-course menu. We will collect your pre-orders in advance.

- The **closing lunch** on Friday will be held at Vegana Vietnamese Restaurant, P6 14, 68161 Mannheim. It is a 20 minute walk from the conference venue, and a 15 minute walk or a 10 minute tram ride to Mannheim Hbf and the conference hotel. The menu can be viewed here. We will collect your pre-orders in advance.

Acknowledgments

As organizers, we gratefully acknowledge financial support by the MZES. We thank the MZES Executive Board, composed of Sabine Carey, Marc Helbling, and Thomas Bräuninger, for accepting our conference proposal and for the generous funding under the MZES Conference Program.

Moreover, we gratefully acknowledge additional financial support by the Lorenz-von-Stein Foundation and by the DVPW.

Lastly, we also thank the MZES for administrative support. We are particularly grateful to Susanne Frey, Nicola Gerngroß, Constanze Nickel, Christine Stegmann, and Philipp Heldmann for their support in organizing this conference, to Nick Hollermeier and Jan Dillhöfer for public relations work and photography on conference day, and to our research assistants, Anna Aigner, Lisa-Marie Zikesch, Marietta Sujin Hüttemann, and Till Kolloge, for their help during the conference.