Universiteit van Amsterdam

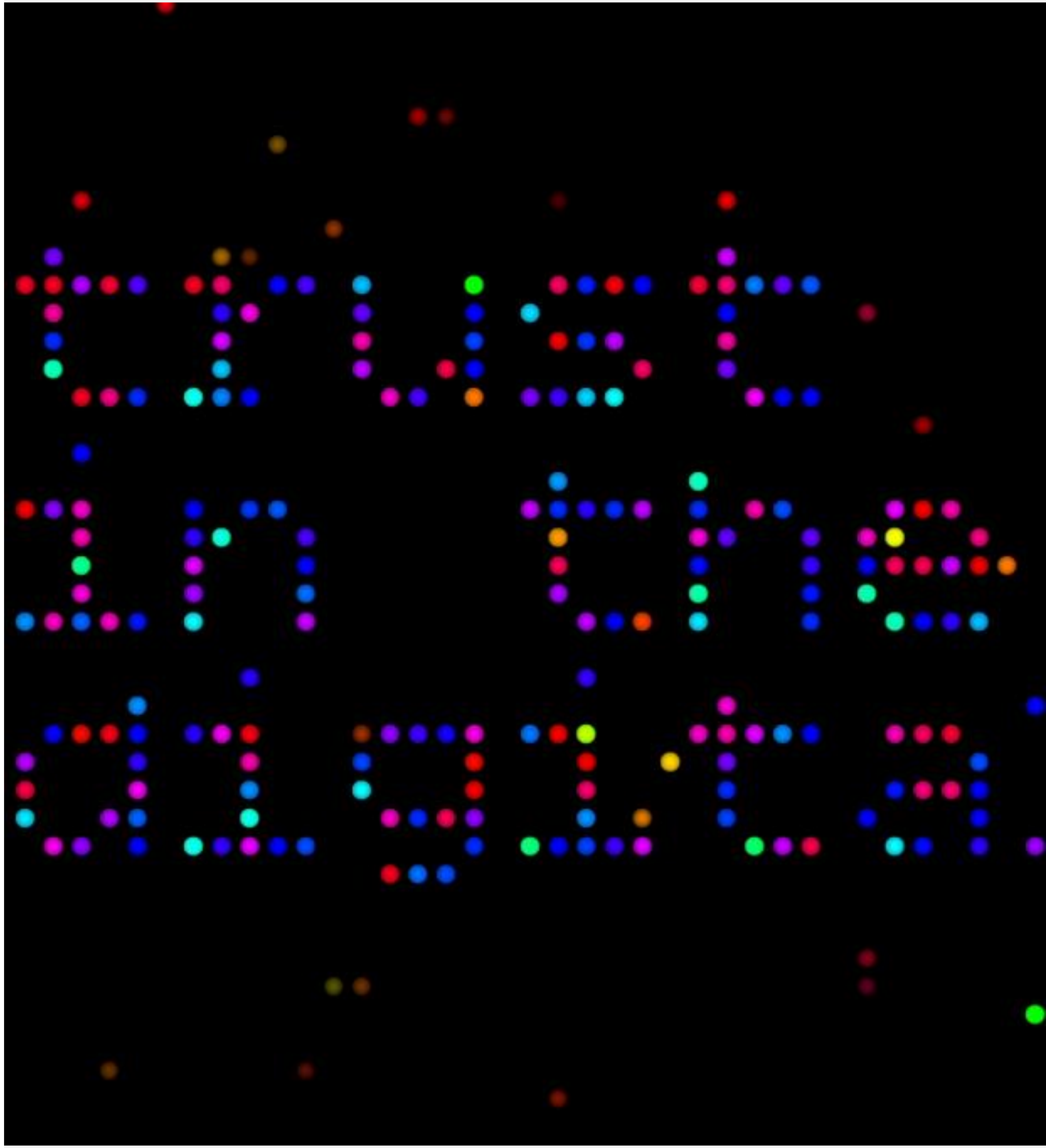
Competence Centre on Behavioural Insights Seminar

Trust in the digital society

Research Priority Area
University of Amsterdam

Balazs Bodo, Jan Engelmann, Theo Araujo,
Monika Simon, Tomasz Zurek

3 oktober 2024



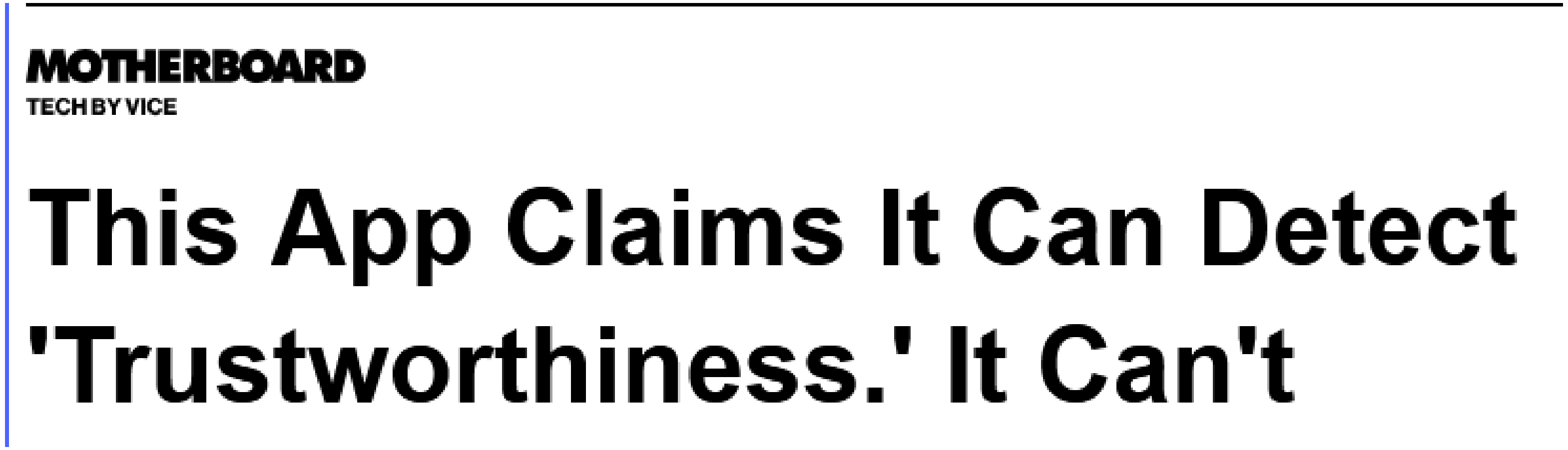
Interdisciplinary research collaboration

- Amsterdam Law School – Institute for Information Law
- Faculty of Humanities – Department of Media Studies
- Faculty of Social and Behavioral Sciences - Amsterdam School for Communication Research
- Faculty of Economics and Business - Amsterdam Neuroeconomics Lab
- Faculty of Science – Informatics Institute

https://digitaltrust.uva.nl/

Amsterdam Trust Platform

- Coordinate and platform academic research on trust related issues
  - Qualitative research
  - Experiments
  - Surveys
  - Big data analyses/ Computational Methods/Data Donation
  - Agent-based modeling
- Interface with industry on trust and safety efforts
  - Regulatory compliance (DSA/DMA, Data Act, AI Act)
  - Content moderation and filtering guidelines
- Interface with policymakers on trustworthiness enhancing policies
  - Regulation is not the only source of trustworthy tech, especially if regulator is untrusted or untrustworthy
- Contribute to maintaining and improving (well placed) trust in society
  - Build institutional frameworks of strategically managed distrust



MOTHERBOARD
TECH BY VICE

This App Claims It Can Detect 'Trustworthiness.' It Can't

Major research problems

1. New, technological forms of societal trust production emerge: dating apps, e-commerce platforms, social media, AI, mobile apps, search engines, etc.
2. This generates changes in (inter)personal trust relations
3. Shifts dynamics of societal trust relations: trust in public institutions, communal trust relations, private trust producers (medical, legal financial professions and institutions)
4. Raises questions about the trustworthiness safeguards and guarantees of technological trust producers

Current research lines

- New theories of trust in the digital society
- Trust in and by social media platforms
- Narratives of trust and distrust on social media
- Cognitive determinants of trust
- Automatic vs. deliberative trust - (political) ingroup and outgroup bias, and affective polarization
- Effects of social media usage on emotions and (interpersonal) trust – a multi-country experiment
- How do we conceptualize and measure trust, mistrust, and distrust in communication research and political science - a systematic review in the context of digital society and technology
- Trust calibration on (generative) artificial intelligence and automated communication
- Agent-based modeling of societal trust dynamics

New theories of risk and trust

- Bodó, B. (2021). Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators. *New Media & Society*, *23*(9), 2668-2690. https://doi.org/10.1177/1461444820939922
- Bodó, Balázs, The Commodification of Trust (May 11, 2021). Blockchain & Society Policy Research Lab Research Nodes 2021/1, Amsterdam Law School Research Paper No. 2021-22, Institute for Information Law Research Paper No. 2021-01, http://dx.doi.org/10.2139/ssrn.3843707
- Bodó, B., & Janssen, H. (2022). Maintaining trust in a technologized public sector. *Policy & Society*, *41*(3), 414–429. https://doi.org/10.1093/polsoc/puac019
- Bodó, B., & Weigl, L. (in press). The frameworks of trust and trustlessness around algorithmic control technologies: A lost sense of community. In J. Goossens, & E. Keymolen (Eds.), *Public Governance and Emerging Technologies: Values, Trust, and Compliance by Design*
- Bodó B, Weigl L, Araujo, T (under review): Governance by Trust Mediators in the Digital Society: A Literature Review and Research Agenda

Trust in / by platforms

Quantitative:

Bodó, B, Bene, M. and Boda, Zs. (under review): Standing Naked in the Storm– European Citizens' Trust in Social Media, Users, Information. *http://dx.doi.org/10.2139/ssrn.4368419*

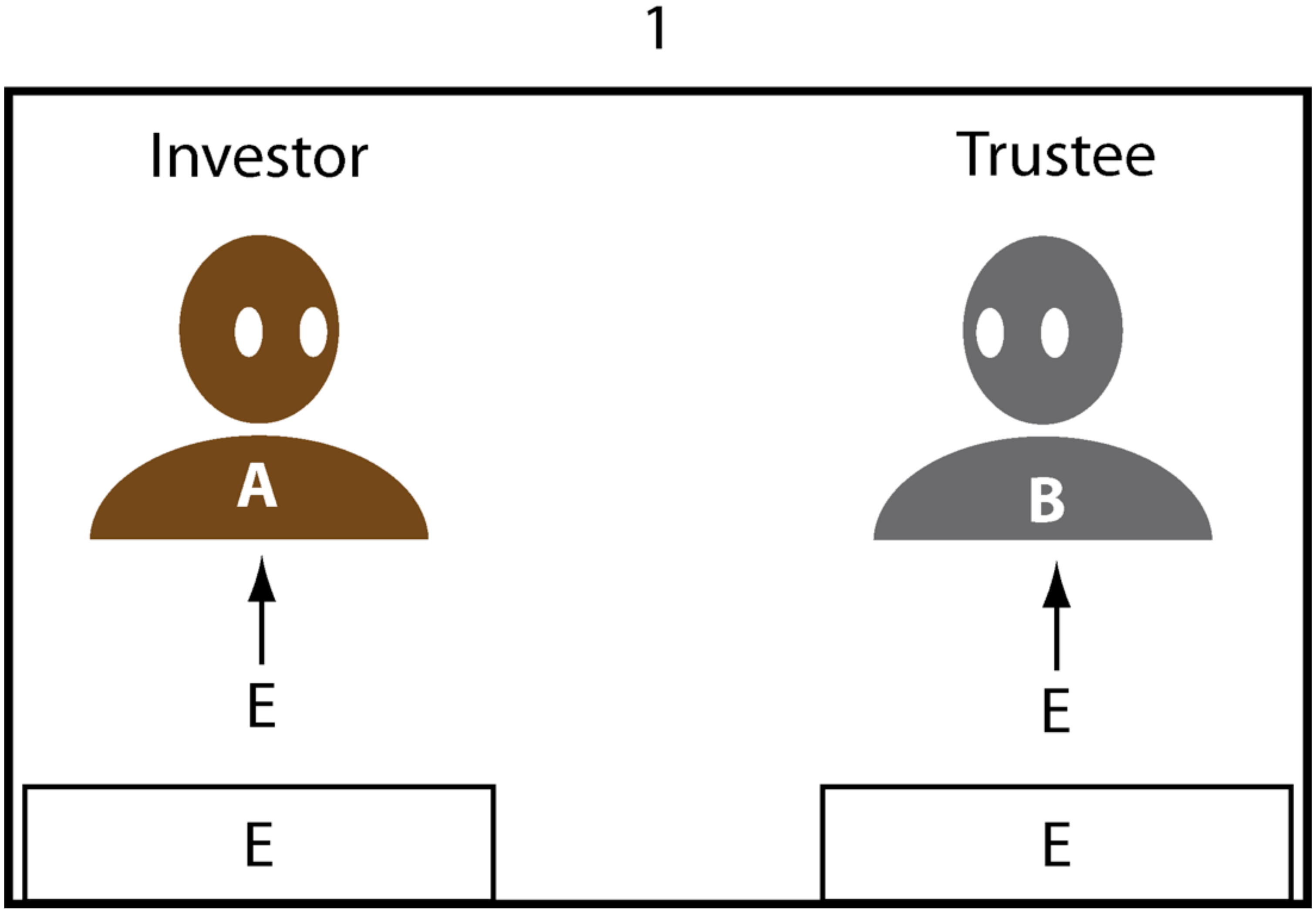
Users' trust in and on the platform depends on their self-confidence to detect and protect from harm, their faith in Meta protecting them, but hardly in regulators' efforts.

Qualitative:

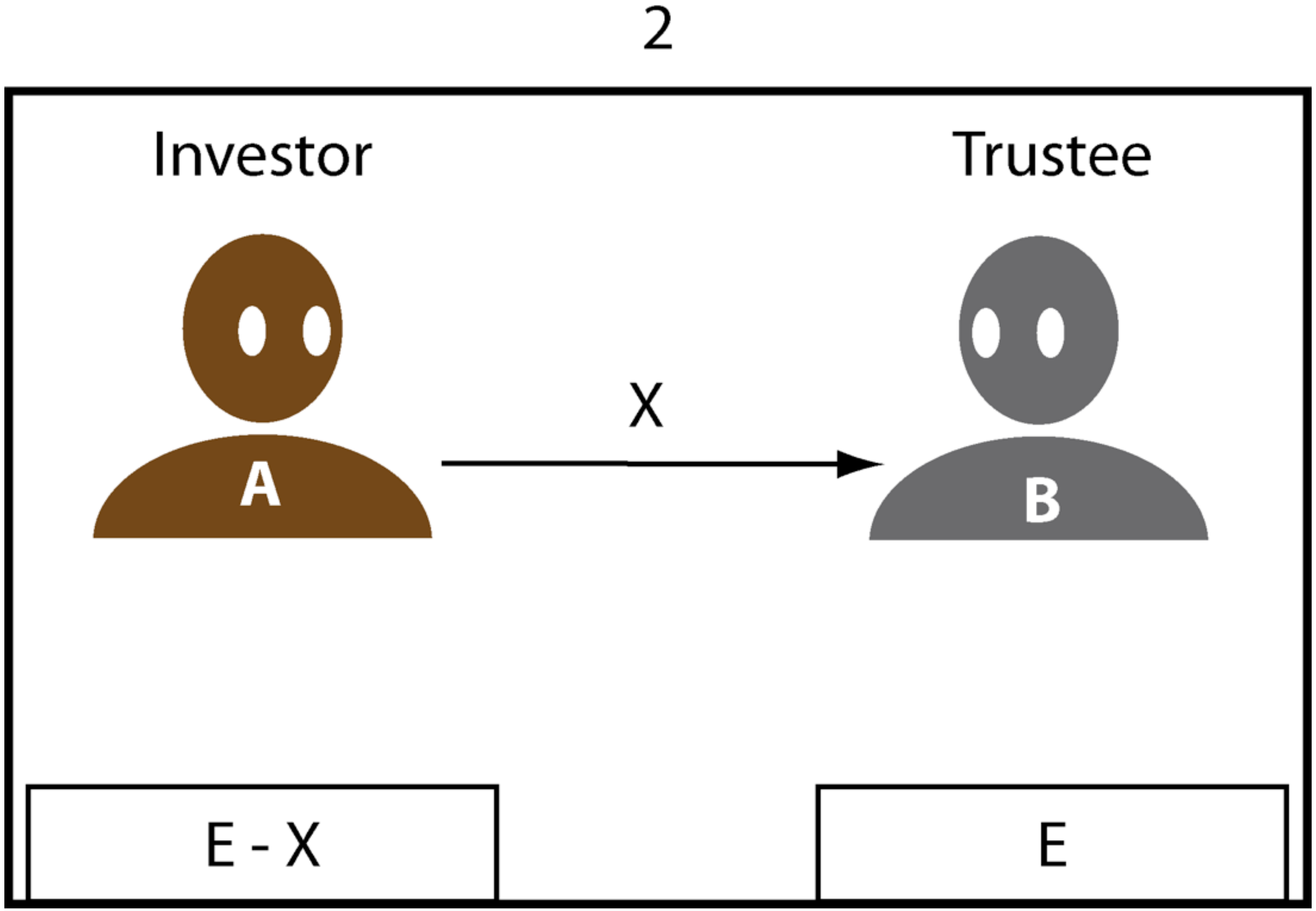
Weigl, L, Bodo, B (work in progress): ***Regulating 'Trust and Safety' Under the Digital Services Act***

Most platforms have set up internal "trust and safety" teams to manage the risks and potential harms. Our research focuses on the work of these "trust and safety" efforts, especially on how they define their tasks, how they are resourced, how they set their priorities, how they define trust and safety, and how they navigate and reconcile the competing pressures they are under.

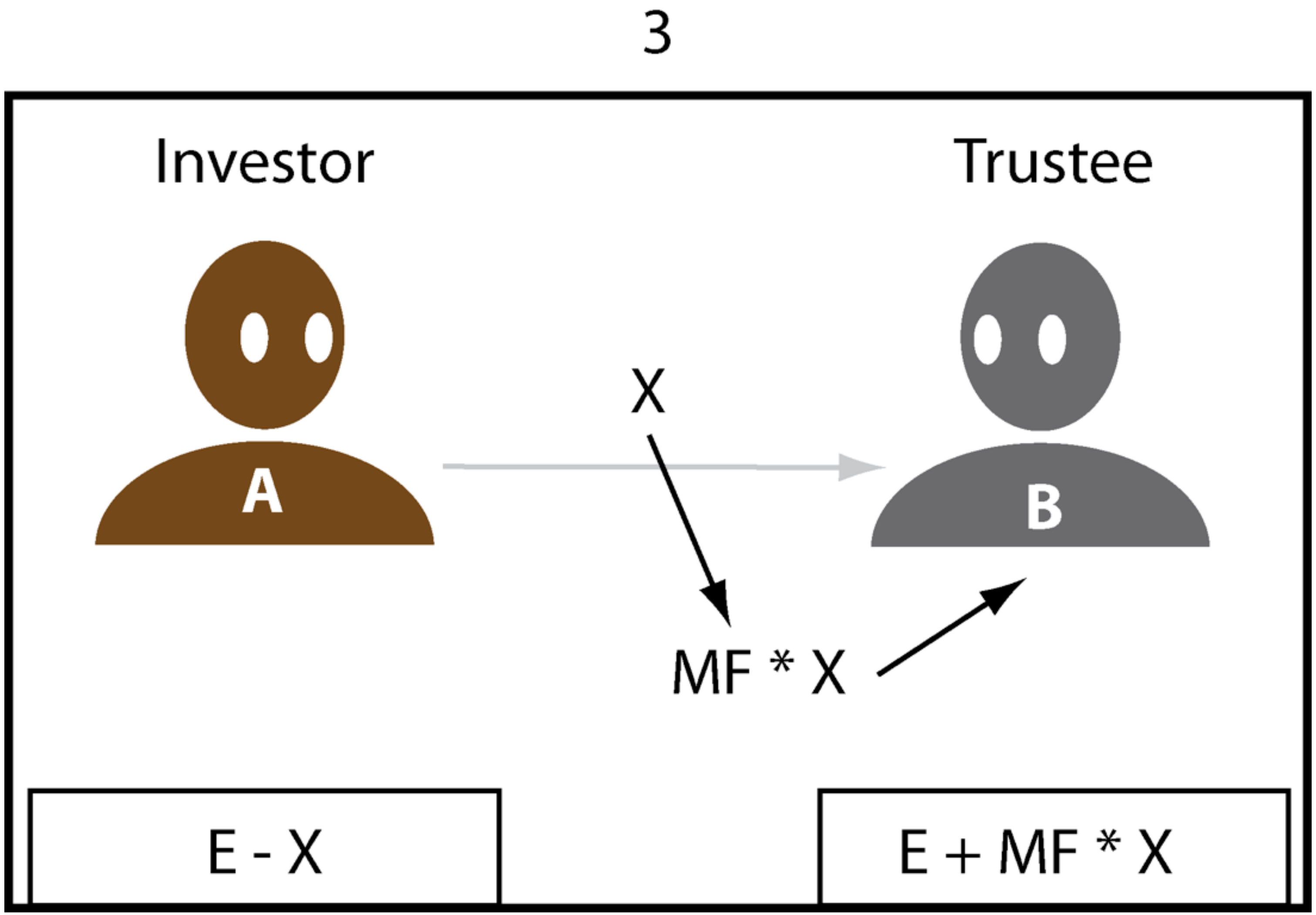
Operationalizing Trust



Operationalizing Trust



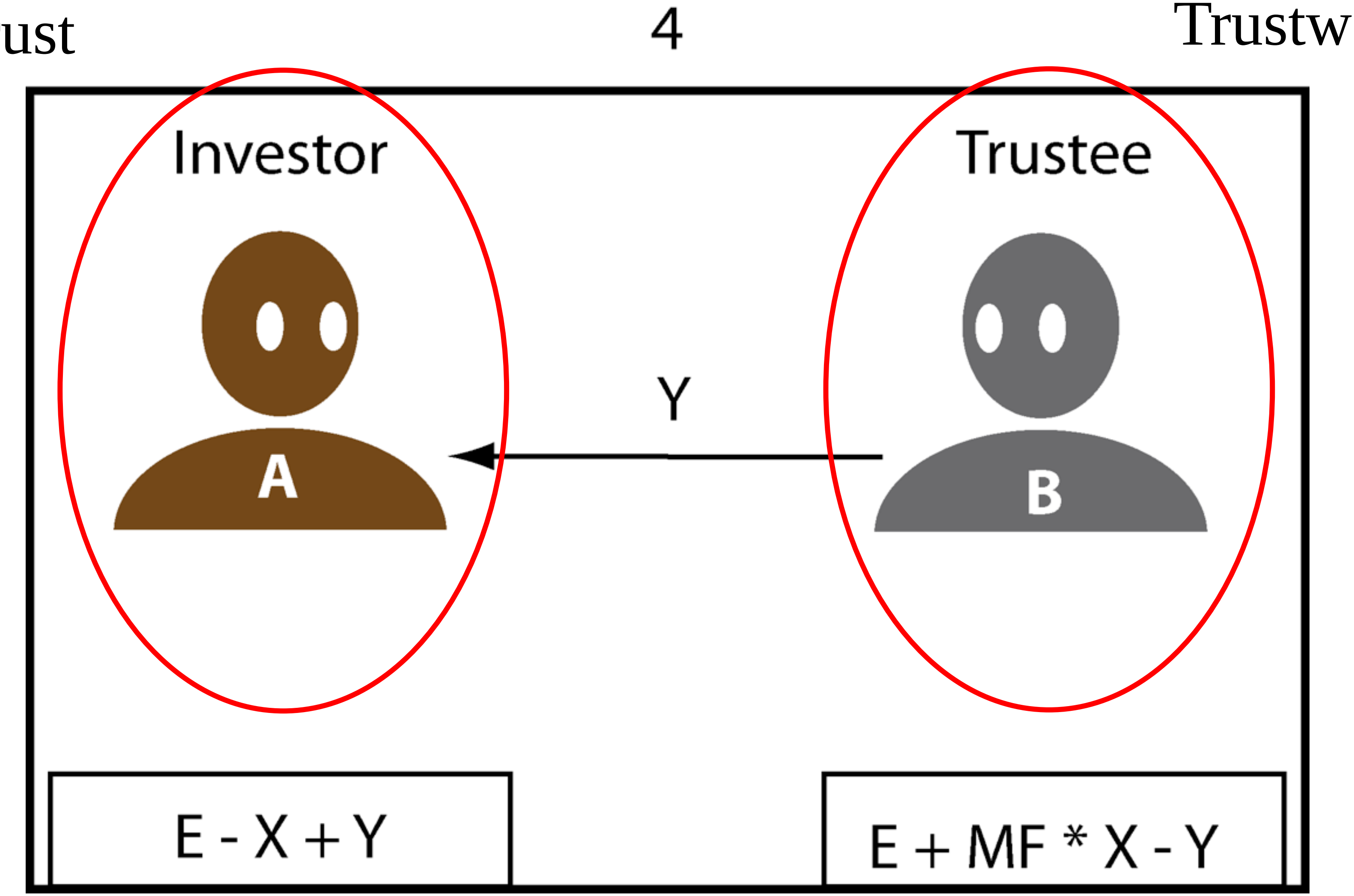
Operationalizing Trust



Operationalizing Trust

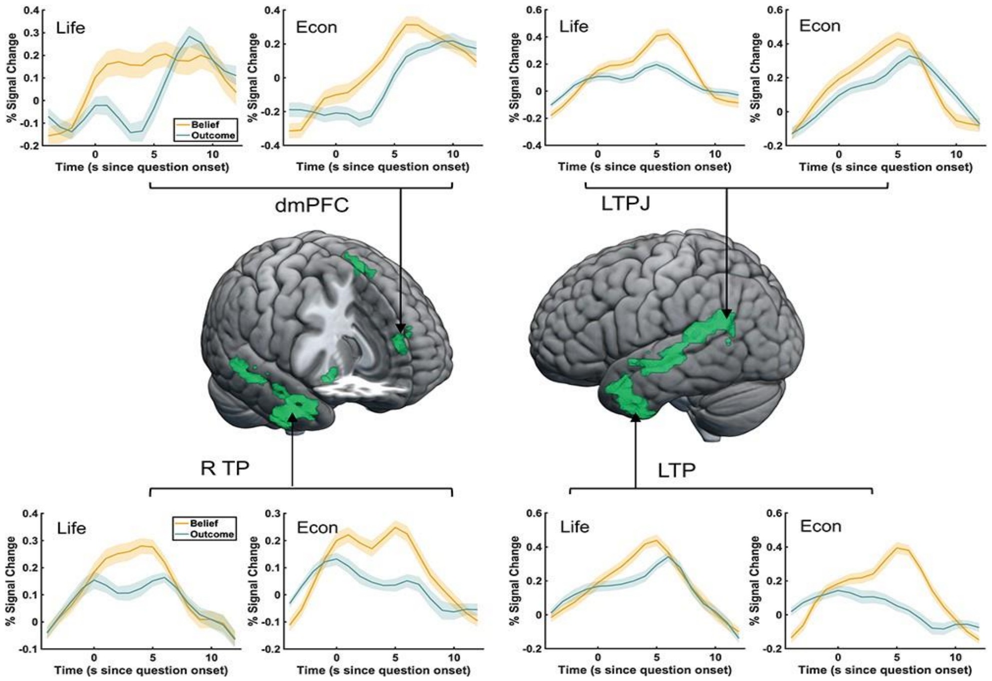
Trust

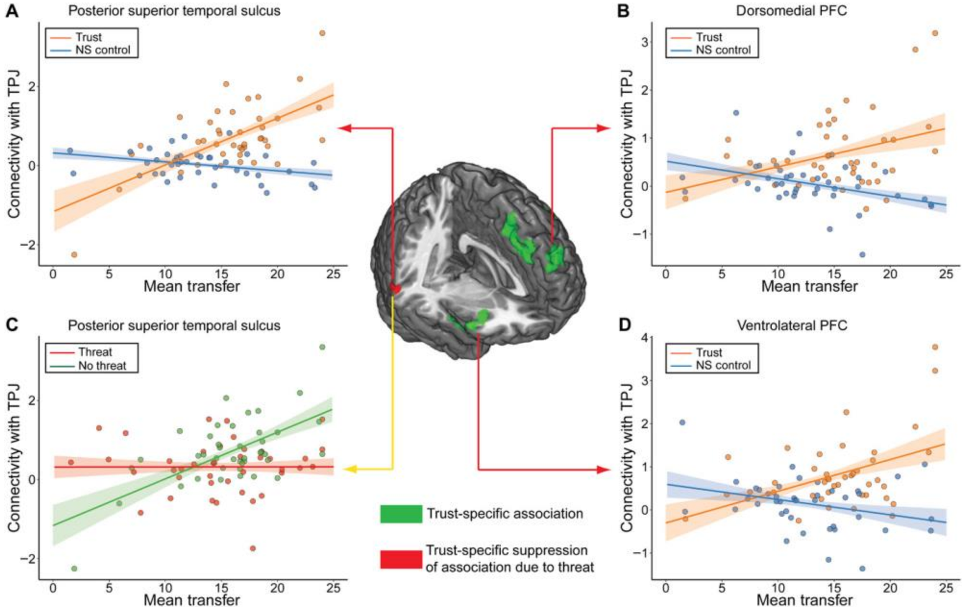
Trustworthiness



Determinants of trust (previous research)

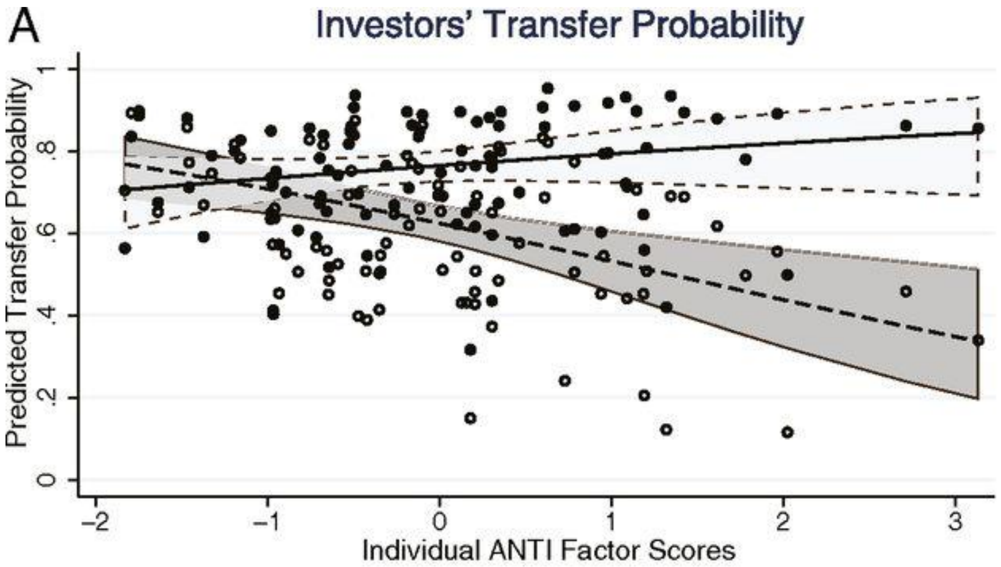
- Neural correlates of trust decisions involves theory of mind
  - Engelmann et al, 2019; Chang et al., 2023
- Affective components of trust
  - Engelmann et al., 2019; Chang et al., 2024
- Personality and Trust
  - Engelmann et al., 2019b
- First impressions and trust
  - Jaeger et al., 2022

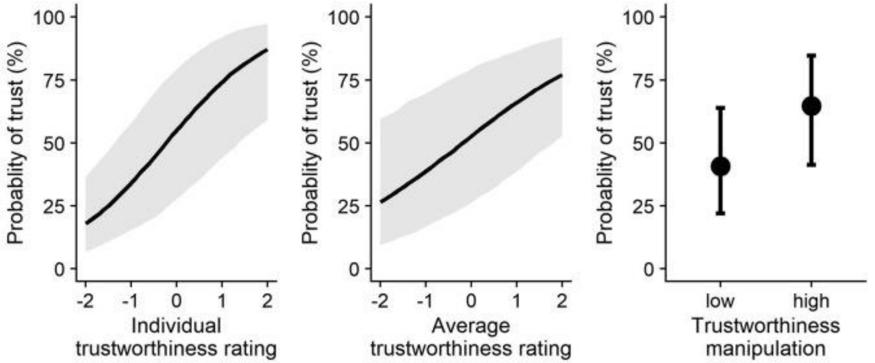


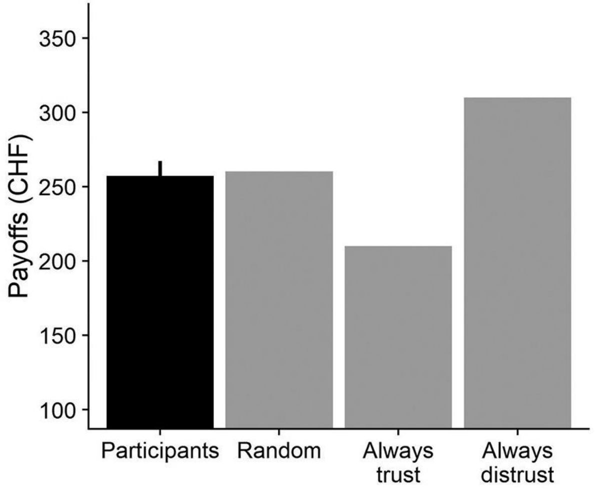


Determinants of trust (previous research)

- Neural correlates of trust decisions involves theory of mind
  - Engelmann et al, 2019a; Chang et al., 2023
- Affective components of trust
  - Engelmann et al., 2019a; Chang et al., 2024
- Antisocial Personality and Trust
  - Engelmann et al., 2019b
- First impressions and trust
  - Jaeger et al., 2022

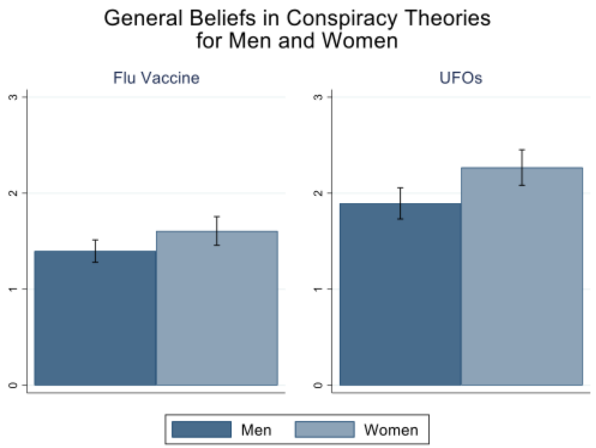


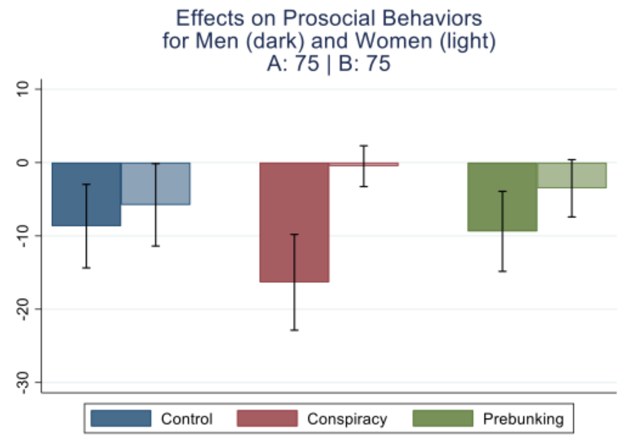


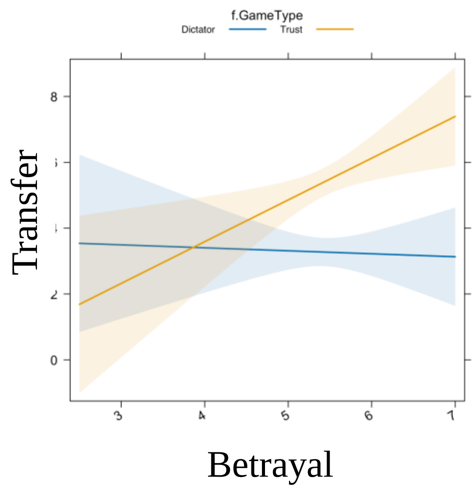


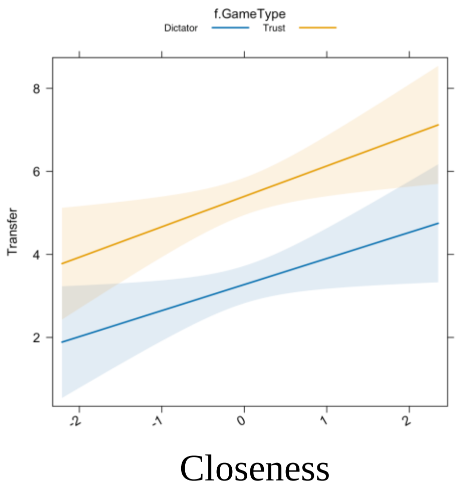
Trust in the context of social media (ongoing research)

- How does exposure to conspiracy theories affect generalized trust?
- Development of the Betrayal Reactivity Questionnaire (BRQ): The role of reactive betrayal and social closeness in trust.
- Relationship between social media use, affect and trust.
- Online gambling in teenagers and young adults in the Netherlands.


General Beliefs in Conspiracy Theories for Men and Women


Effects on Prosocial Behaviors for Men (dark) and Women (light)
A: 75 | B: 75





References

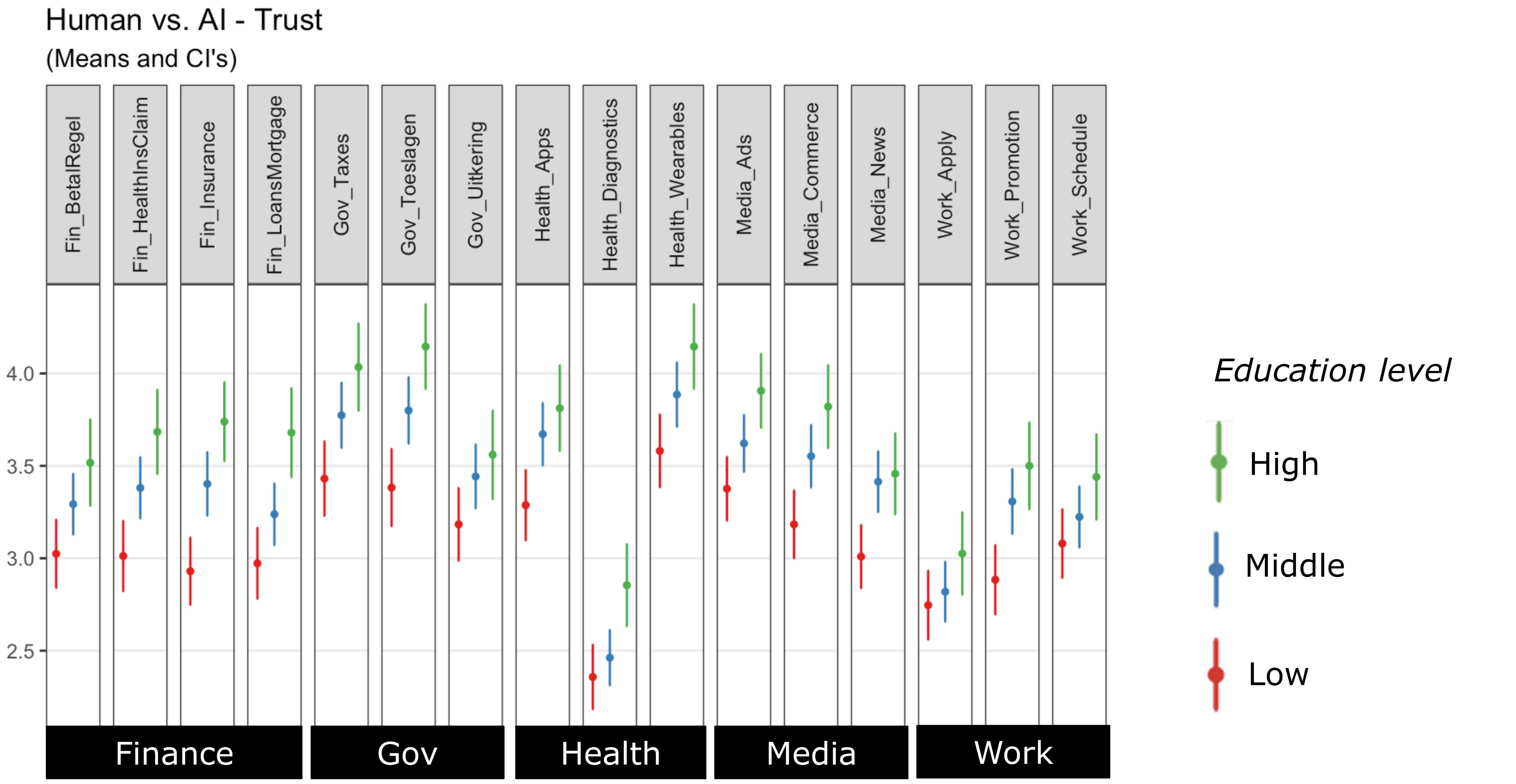
- Chang, LA & Engelmann, JB (2024) The impact of incidental anxiety on the neural signature of mentalizing. Imaging Neuroscience 2, 1-23
- Chang, LA, Warns, L, Armaos, K, de Sousa, AQM, Paauwe, F, Scholz, C, Engelmann, JB (2023). Mentalizing in economic games is associated with enhanced activation and connectivity in left temporoparietal junction. Social Cognitive and Affective Neuroscience, nsad023
- Jaeger, B, Oud, B, Williams, T, Krumhuber, EG, Fehr, E, Engelmann, JB (2022). Can people detect the trustworthiness of strangers based on their facial appearance? Evolution and Human Behavior 43 (4): 296-303
- Farolfi, F, Chang, L., Engelmann, JB (2021) Trust and Contextual Emotions. In Krueger, F (ed.) The Neurobiology of Trust, Oxford University Press
- Engelmann, JB, Schmid, B, de Dreu, C, Chumbley, J & Fehr, E (2019) On the psychology and economics of antisocial personality. Proceedings of the National Academy of Sciences 116 (26): 12781 – 12786
- Engelmann, JB, Meyer, F, Ruff, CC, Fehr, E (2019) The neural circuitry of emotion-induced distortions of trust. Science Advances 5(3) eaau3413
- Engelmann, JB and Fehr, E (2017) The neurobiology of trust: the important role of emotions. In Van Lange, PAM, Rockenbach, B & Yamagishi, T (Eds.) Trust in Social Dilemmas (Series in Human Cooperation), Oxford University Press.
- All papers can be found at www.neuro-economics.net

Trust in Artificial Intelligence

Research focus: Our interactions with an increasingly automated communication environment

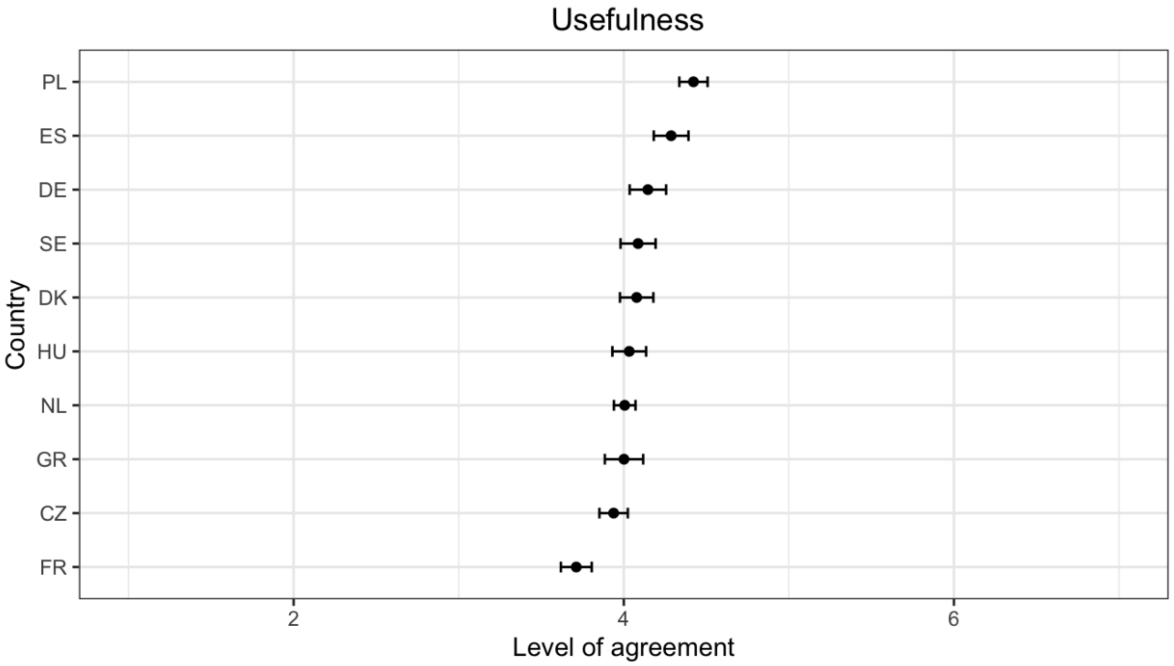
Causes, **C**ontents, **C**onsequences, and **C**ounterstrategies to empower individuals for trust calibration on (Generative) AI

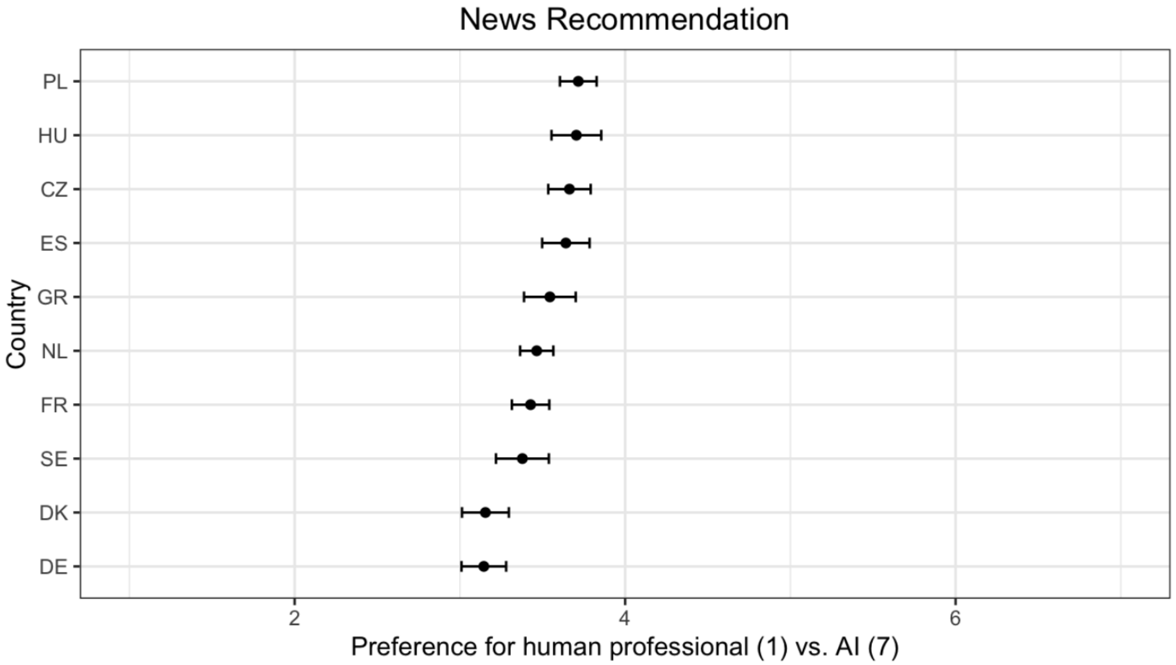
Trust in Artificial Intelligence



Araujo, ter Hoeven & de Vreese (working paper) | Representative sample of the Dutch population, N = 981

Perceptions about AI



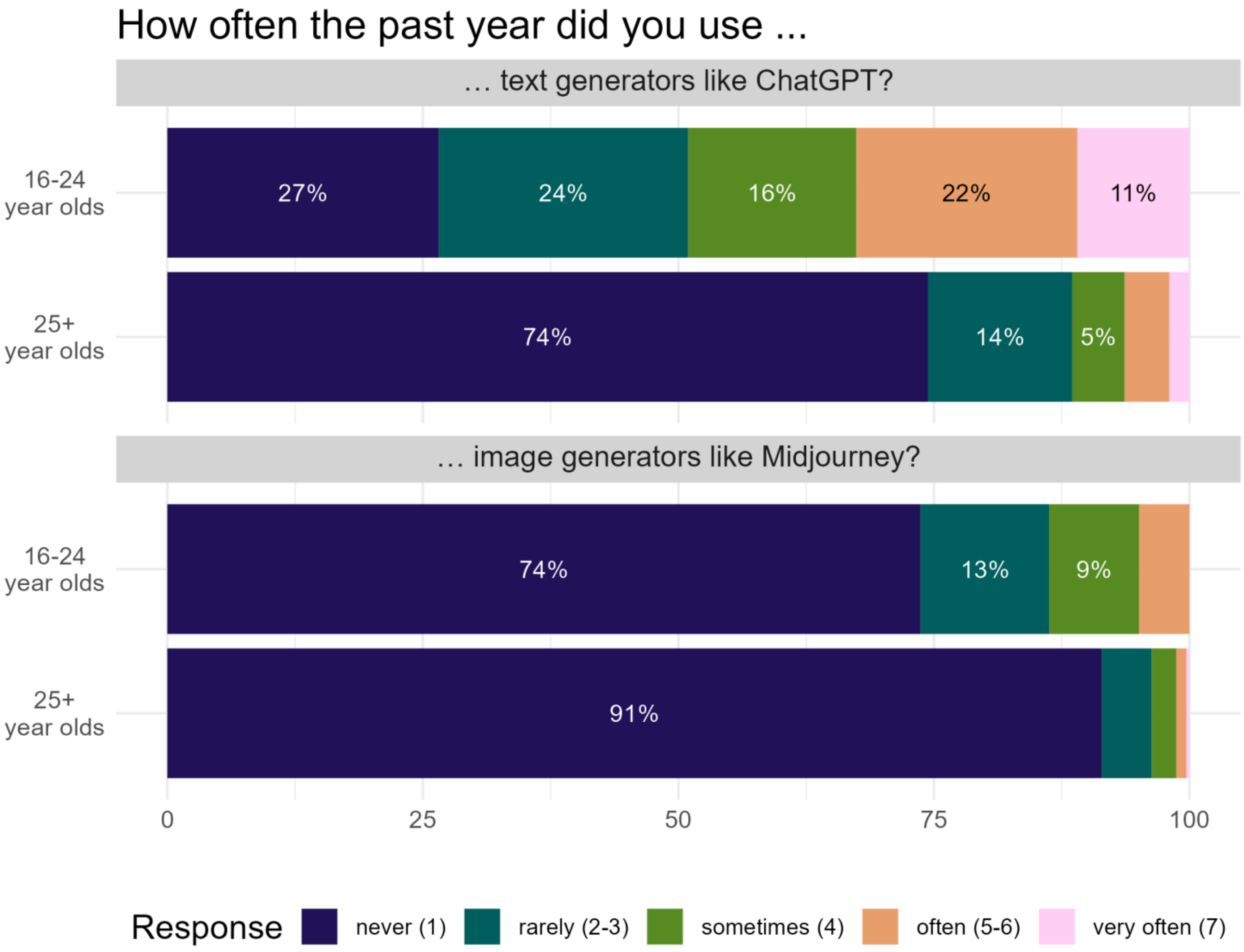


Araujo, T., Brosius, A., Goldberg, A. C., Möller, J., & Vreese, C. de. (2023). Humans vs. AI: The Role of Trust, Political Attitudes, and Individual Characteristics on Perceptions About Automated Decision Making Across Europe. *International Journal of Communication*, *17*(0).

Usage of Generative AI

Amsterdam School of Communication Research

AlgoSoc /



Ministerie van Onderwijs, Cultuur en Wetenschap

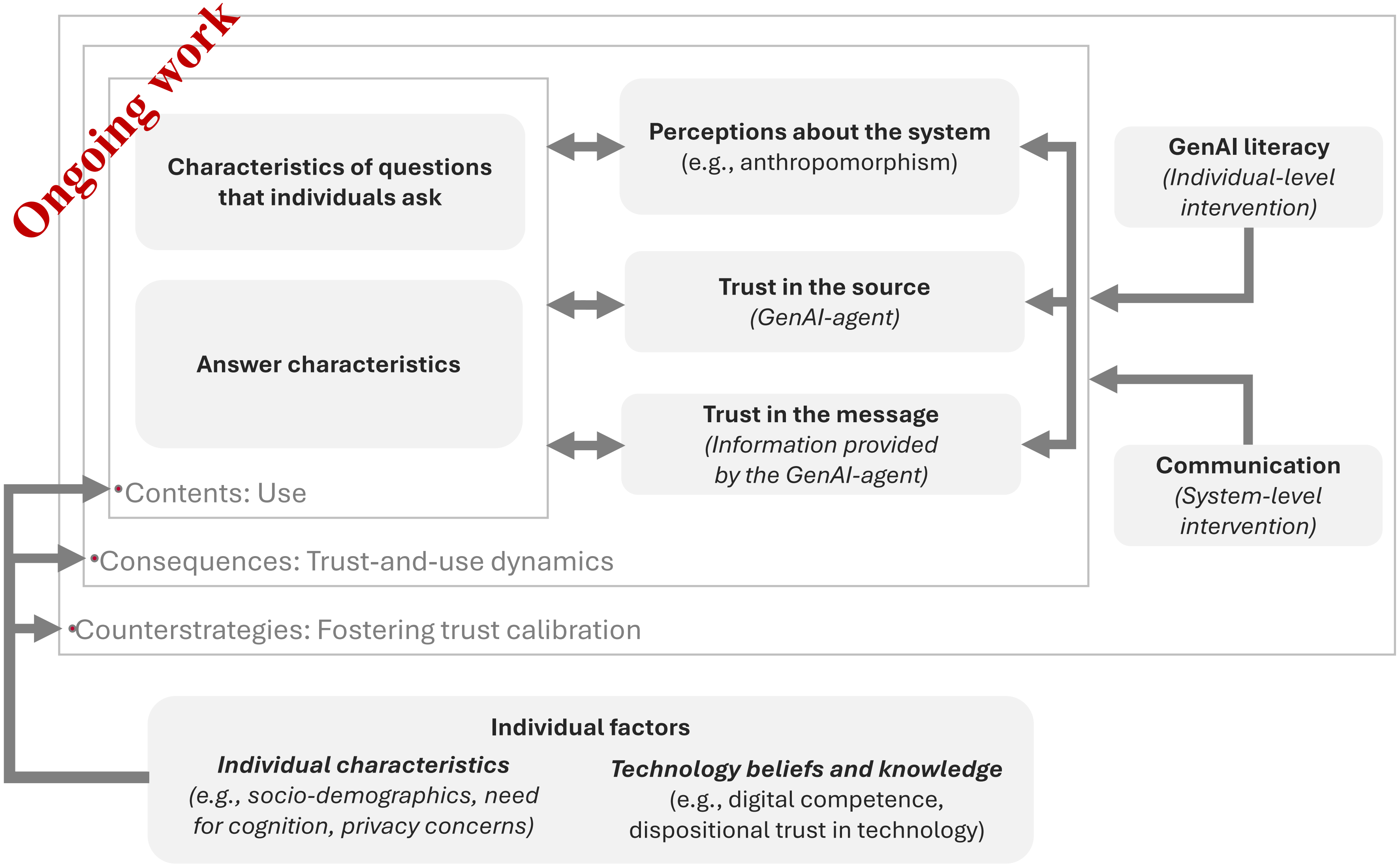
AlgoSoc is funded by the Dutch Ministry of Education, Culture and Science (OCW)
+ Gravitation
+ 024.005.017

Source: AlgoSoc.org

de León, E., Votta, F., Araujo, T., & de Vreese, C.H. (2024). Mind the A(I)ge gap? Emerging generational fault lines in public opinion on Artificial Intelligence. *Public Report.*

Trust in Generative AI

Trust calibration on GenAI-agents: an integrative model



Ongoing work

Characteristics of questions that individuals ask

Answer characteristics

Perceptions about the system
(e.g., anthropomorphism)

Trust in the source
(GenAI-agent)

Trust in the message
(Information provided by the GenAI-agent)

GenAI literacy
(Individual-level intervention)

Communication
(System-level intervention)

- Contents: Use
- Consequences: Trust-and-use dynamics
- Counterstrategies: Fostering trust calibration

Individual factors

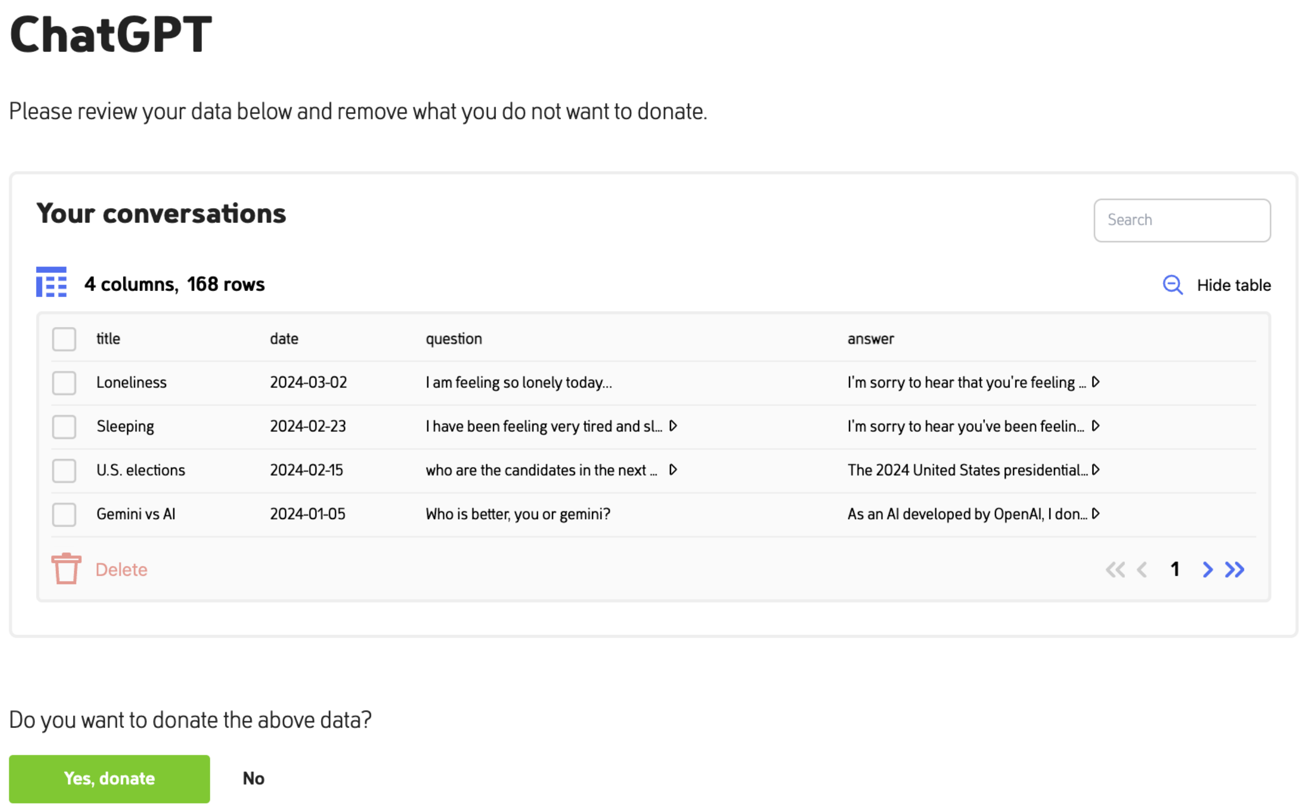
Individual characteristics
(e.g., socio-demographics, need for cognition, privacy concerns)

Technology beliefs and knowledge
(e.g., digital competence, dispositional trust in technology)

Trust in Generative AI

A citizen science approach: Data donation

Ongoing work



ChatGPT

Please review your data below and remove what you do not want to donate.

Your conversations

4 columns, 168 rows

□	title	date	question	answer
□	Loneliness	2024-03-02	I am feeling so lonely today...	I'm sorry to hear that you're feeling ...
□	Sleeping	2024-02-23	I have been feeling very tired and sl...	I'm sorry to hear you've been feelin...
□	U.S. elections	2024-02-15	who are the candidates in the next ...	The 2024 United States presidential...
□	Gemini vs AI	2024-01-05	Who is better, you or gemini?	As an AI developed by OpenAI, I don...

Do you want to donate the above data?

Pilot study

($N_{donations}$ = 52)

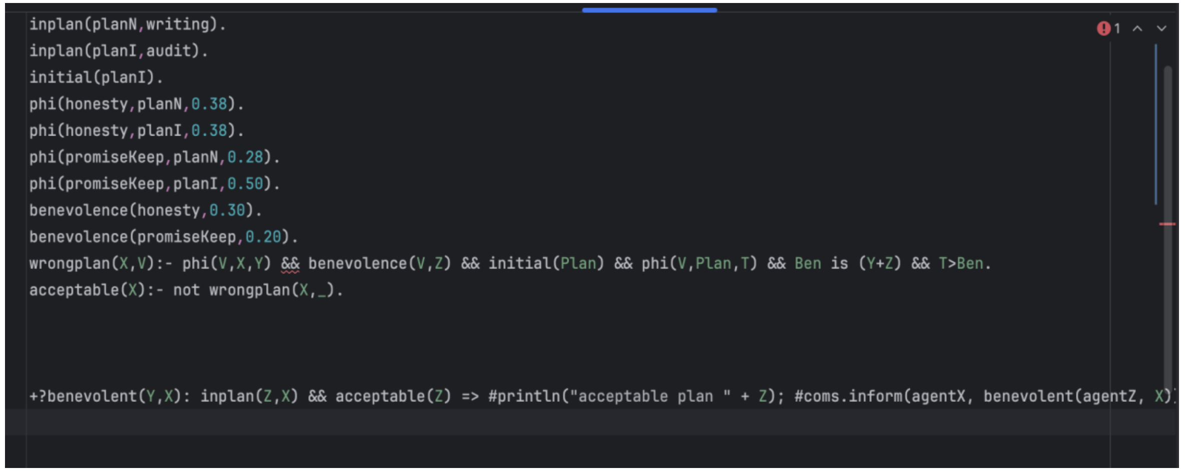
Average donation:

- **131** questions to ChatGPT
- **226** days (first vs. last chat)

Note: Small sample, large standard deviations, already excluding one outlier

Agent-based modeling of societal trust dynamics

- Can we represent trust and trustworthiness as computational phenomenon?
- Can we reproduce existing social science experiments in the world of artificial agents?
- Decomposition of trustworthiness evaluation into the evaluation of agent's Competence, Benevolence, and Integrity
- Model of trust dynamics: How experience and contact with other agents changes trust relations
- The model of complex socio-technical environment in simulated Multi-Agent system
- Verification of existing trust theories in simulated environment of artificial agents



```
inplan(planN,writing).
inplan(planI,audit).
initial(planI).
phi(honesty,planN,0.38).
phi(honesty,planI,0.38).
phi(promiseKeep,planN,0.28).
phi(promiseKeep,planI,0.50).
benevolence(honesty,0.30).
benevolence(promiseKeep,0.20).
wrongplan(X,V):- phi(V,X,Y) && benevolence(V,Z) && initial(Plan) && phi(V,Plan,T) && Ben is (Y+Z) && T>Ben.
acceptable(X):- not wrongplan(X,_).



+?benevolent(Y,X): inplan(Z,X) && acceptable(Z) => #println("acceptable plan " + Z); #coms.inform(agentX, benevolent(agentZ, X)
```

Points of cooperation

- *We are open to empirical research collaborations:*
  - Share survey questions, resources
  - (social) media analysis (using big data methods)
  - Experiments
- *We are open to policy work:*
  - White papers, reports, recommendations,
  - Stakeholder consultations
- *We are open to knowledge exchange/dissemination*
  - Policy dialogues
  - Expert contributions

Thank you for your attention