
HOW THE ONLINE CROWD AND EXPERTS COLLABORATE TO CREATE INNOVATIVE DATA

DR. ALEXANDER HORN

SOCIAL SCIENCE RESEARCH CENTER BERLIN (WZB)/AARHUS UNIVERSITY

AHORN@PS.AU.DK

BASIS FOR PRESENTATION

Free version:

Interview on occasion of the DJØF Publishing Prize 2018 for “Crowdcoding of Parties’ Equality Concepts”, Djøfbladet 10/18 p.48-51; free online under djoefbladet.dk [**Tusindvis af internetbrugere skal samle data til forskere**]

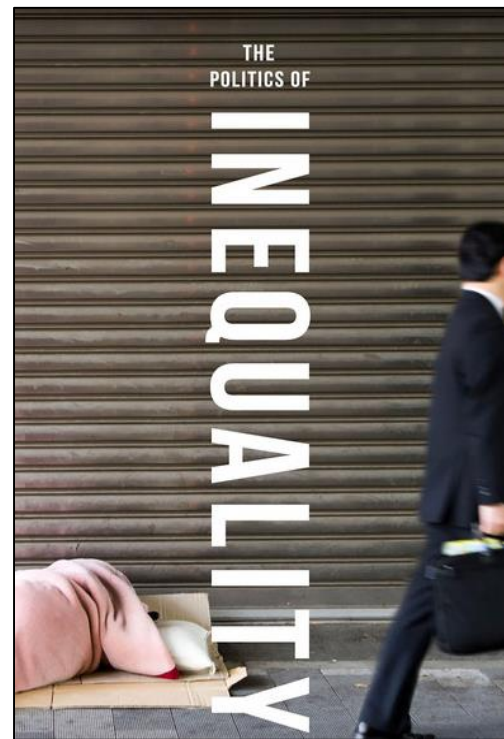
Pay-walled:

Can the Online-Crowd match Real Expert Judgments? How Task Complexity and Coder Location Affect the Validity of Crowd-Coded Data.
European Journal of Political Research. Online first 15.4.2018.

EXAMPLE OF A BIG TOPIC WHERE WE LACK DATA: INEQUALITY

One of *the* if not the challenge of the 21st century
Despite relevance, huge imbalances in research

- ▶ Know a lot about trends in economic inequality by now
- ▶ Studies on equality preferences of voters “uncountable”
- ▶ **Don't know what parties have to say about (in-)equality**
- ▶ Surprising? ...

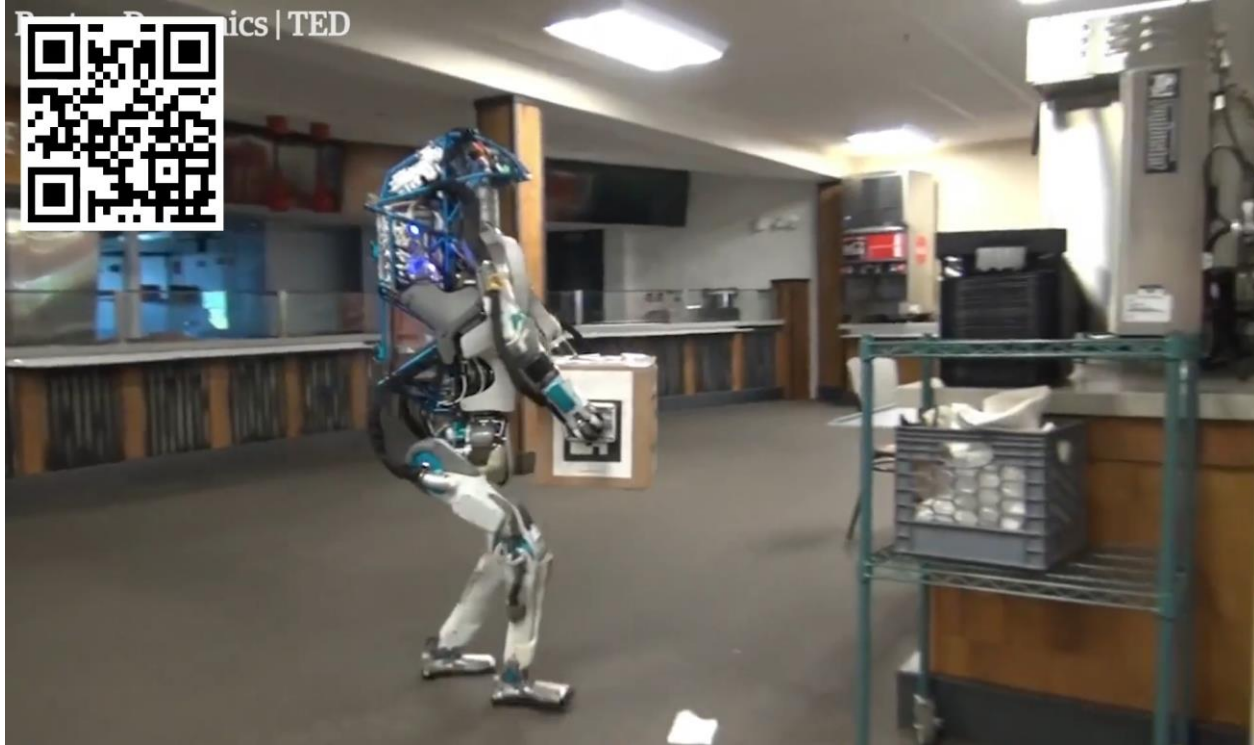


I'M THE GUY WHO DOES HIS JOB

YOU MUST BE THE OTHER GUY

BASED ON AUTOMATED CONTENT ANALYSIS ...

... MISSION IMPOSSIBLE (HOPKINS AND KING 2010, BENOIT 2016)



Source: Boston Robotics

SOLUTION? ONLINE CROWDCODING OF POLITICAL TEXTS FROM PARTIES. CROWD CODING IS A DATA COLLECTION TECHNIQUE ...

- Combine context knowledge/validity of experts with efficiency of crowd
- online categorization of large amounts of data that **combines repeated and independent judgments for the same unit** by paid non-expert coders
- coders must pass an **entry quiz** and are **screened with test questions** to deselect underperformers/spammers throughout actual coding process
- this way we get **a trust score for each coder** → have uncertainty estimate
- has its **roots** in idea that **multiple independent judgments** by a diverse crowd of contributors can (**when aggregated**) **match expert** judgements

ADVANTAGES/POTENTIAL

1. Transparency: extend the replication standard to data collection process
2. Affordability: prospect of scalability (limited based on expert judgment)
3. Content validity: possibility to move beyond the “off-the-shelf” data sets
→ improve limited content validity of canonical data

EXPERT-CROWD MATCH AT 3 LEVELS

L1: Individual decision by coder	= one statement coded one time
L2: Coding decision for each unit	= aggregated at statement level
L3: Level or results	= at party level, relative emphasis

But what matters?

- L1: Miscodings do not matter when most others code the same statement correct
- L2: May not matter if sufficiently high match (assuming no systematic error/bias)
- L3: Expert-crowd match at level of party priorities should lead to same conclusions

look at L2 and L3 (L2 simple measure of % match, L3 visual inspection + correlations)

SIMPLE AND COMPLEX TASK

Complex:

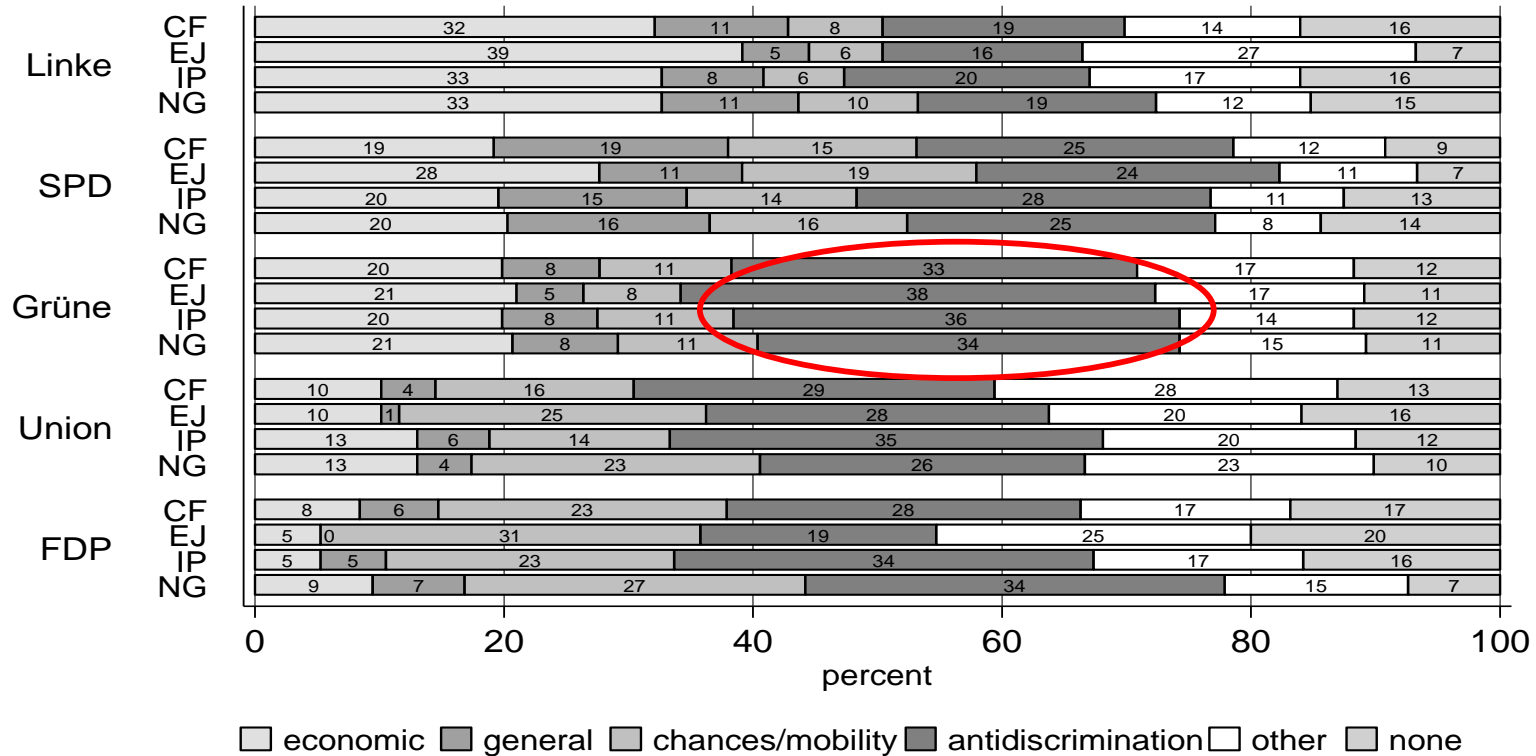
- ▶ Does the following statement from a party program exert a positive reference to equality, social justice, and/or equal treatment of all people? **If YES, which of the categories 1 to 5 fits best?** If NO, please select category 6. Please read all of the 6 category descriptions carefully! If you are of the opinion that more than one category is relevant, please decide which category is emphasized most clearly.

economic general chances/mobility antidiscrimination other none

Simple:

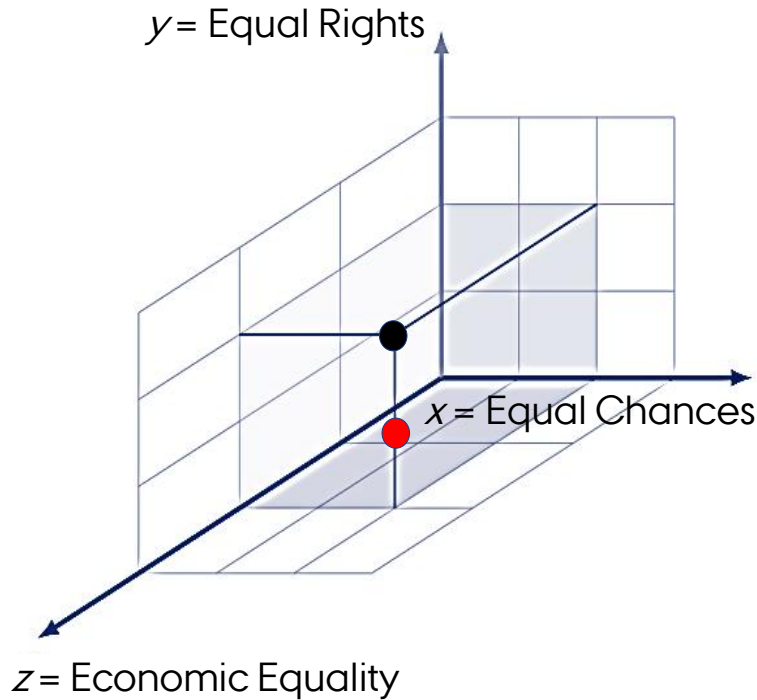
- ▶ Does the following statement from a party program exert a **positive reference** to equality, and/or social justice, and/or equal treatment of all people? If YES, please select category 1 /yes? If NO, please select category 2 /no.

EQUALITY CONCEPTS AT PARTY LEVEL [BELOW: ONLY COMPLEX TASK] (HERE: GER)



CF = based on crowd coding via CrowdFlower without german IP restriction; EJ = Expert Judgments; IP = crowd coding via CrowdFlower only with coders that have a German IP; NG = crowd coding via CrowdFlower of coders with non-German IPs. Value labels are rounded values.

WHAT DO I DO WITH THIS?



We can compare what parties say on equality across countries, parties, time ...

E.g.: party remains economically egalitarian, pro-mobility, but de-emphasizes equal rights:

- SD 1980 (2/2/2)
- SD 2015 (2/1/2)



TAKE AWAYS

- Results obtained from crowd coding can match results of real experts!
- Especially if we use test questions and IP address restrictions for complex jobs
- Correlations between expert and crowd results can be up to 94% for complex job
- For more simple tasks (often less interesting), neither of this matters that much
- If the questionnaire is made publicly available, others can replicate the data we collected
- So let us choose questions by real relevance, not convenience ("for what do I have data")

Thank you for your attention!

ALEXANDER HORN

AARHUS AND WZB

AHORN@PS.AU.DK