Faster Peace via Inclusivity: An Efficient Paradigm to Understand Populations in Conflict Zones

Jordan Bilich*, Michael Varga*, Daanish Masood, Andrew Konya***

*Remesh, ** United Nations

Motivation

The impact of violent conflict:

- Since 2011, conflicts worldwide have killed up to 100,000 people a year and caused 3-15x more deaths indirectly
- By 2030, it is projected that over half the world's poor will be living in countries affected by high levels of violence.
- Migration, malnutrition, destroyed infrastructure, and distressed environments due to conflict lead to poor health, increased infant mortality, and decreases in the quality of childhood education.
- Conflict disproportionately impacts those with lower socio-economic status, increasing economic inequality
- Overall economic losses due to conflict have doubled over the last decade to an estimated $1trillion per year
- $27b spent annually on peacebuilding efforts

Peacebuilding and Inclusivity

Article 33 of United Nations Charter

"the parties to any dispute, the continuance of which is likely to endanger the maintenance of international peace and security, shall, first of all, seek a solution by negotiation, enquiry, mediation, conciliation, arbitration, judicial settlement"

UN practice shows that for a mediation and dialogue process to be successful, **inclusivity is vital.**

Inclusivity: *(United Nations definition)*

The extent and manner in which the views and needs of conflict parties and other stakeholders are represented and integrated into the process and outcome of a [conflict] mediation effort

Inclusivity Challenges

- Positions of stakeholder populations tend to shift
- Mediators grapple with various tensions between inclusivity and efficiency.
- Existing methods for inclusivity manifest **tradeoff between conversational agility and statistical reliability.**
- ML often viewed as too risky for high-stakes decisions due to lack of result trustworthiness

Real-time Synchronous Large-scale Dialogue Process

Each **minute-scale cycle** of dialogue process:

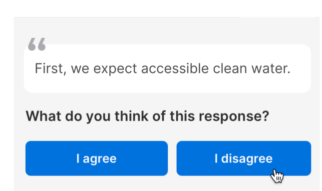
	Dialogue moderators	Participating population
1.	*Send open-ended question*	
2.		*Respond in natural language* *Vote on other's responses*
3.	**Compute results**	
4.	*Review results* *Decide what to ask next*	

Technical Challenges

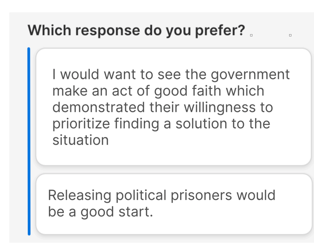
1. Min time scale = sparse data = need prediction model
2. High stakes = need confidence in result
3. Total compute time must be on the order seconds

Model

Voting exercises from phase 2:



Agreement
We denote the even that participant i agrees with response j as a_{ij} and disagrees as d_{ij}



Pair choice
We denote the even that participant i prefers response j over responses k as c_{ijk}

We denote $X = \{i, j \mid x_{ij}\}$ and the utility of response j to person i as m_{ij} . We then write the likelihood as:

$$p(A, D, C|, M, B) = \prod_{i,j \in A} \sigma(m_{ij} + b_i) \prod_{i,j \in D} (1 - \sigma(m_{ij} + b_i)) \prod_{i,j,k \in C} \sigma(m_{ij} - m_{jk})$$

To take advantage of the low-rank nature of the utility matrix we leverage matrix completion with a nuclear norm constraint on M. This is equivalent to an L1 norm on a matrix comprised of M's singular values. We apply a uniform prior over the nuclear norm ball of radius τ . Setting the bias B to zero gives the posterior:

$$p(M|A, D, C) = \frac{1}{Z} p(A, D, C|M) \, \mathbf{1}_{||M||* < \tau}$$

Where:

$$Z = \int p(A, D, C|M) \, \mathbf{1}_{||M||* < \tau} dM$$

$$||M|| * = tr(\Sigma) \;\; given \; SVD \; of \; M = V \Sigma V^*$$

And predicted fraction of participants agreeing with response j is:

$$\alpha_j(M) = \sum_i \sigma(m_{ij})$$

Data collection

- Data collected in low risk environment
- 111 participants from Mechanical Turk
- Data collected over 4 minutes per question

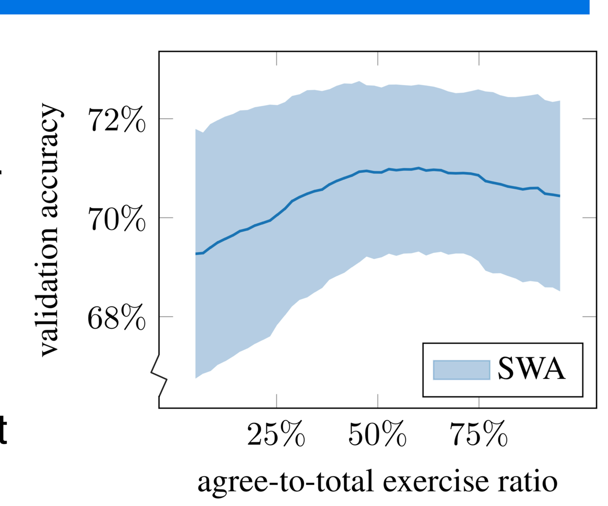
Question	Participants	Responses	Agree/Disagree	Paired comparisons
What is your favorite thing to do in your free time?	111	108	2030	1951
What motivates you the most in life, and why?	108	154	1489	1367
What will be the most important political issue in 5 years?	95	136	1345	1262
What could Amazon do to improve your experience on Mechanical Turk?	101	147	1283	1232

Average data collected per question:

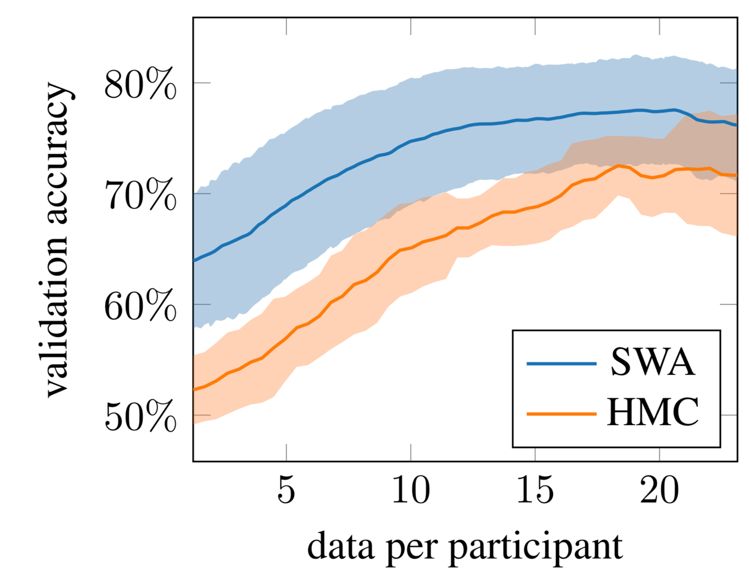
- 136 responses, 1.5k agreement votes, 1.5k pair choice votes

What is ideal mix of voting exercises?

- Validation accuracy in predicting individual agreement votes peaks near a 50:50 mix of agreement vs pair choices exercises holding total number of exercises fixed.
- Result is surprising but effect is relatively small

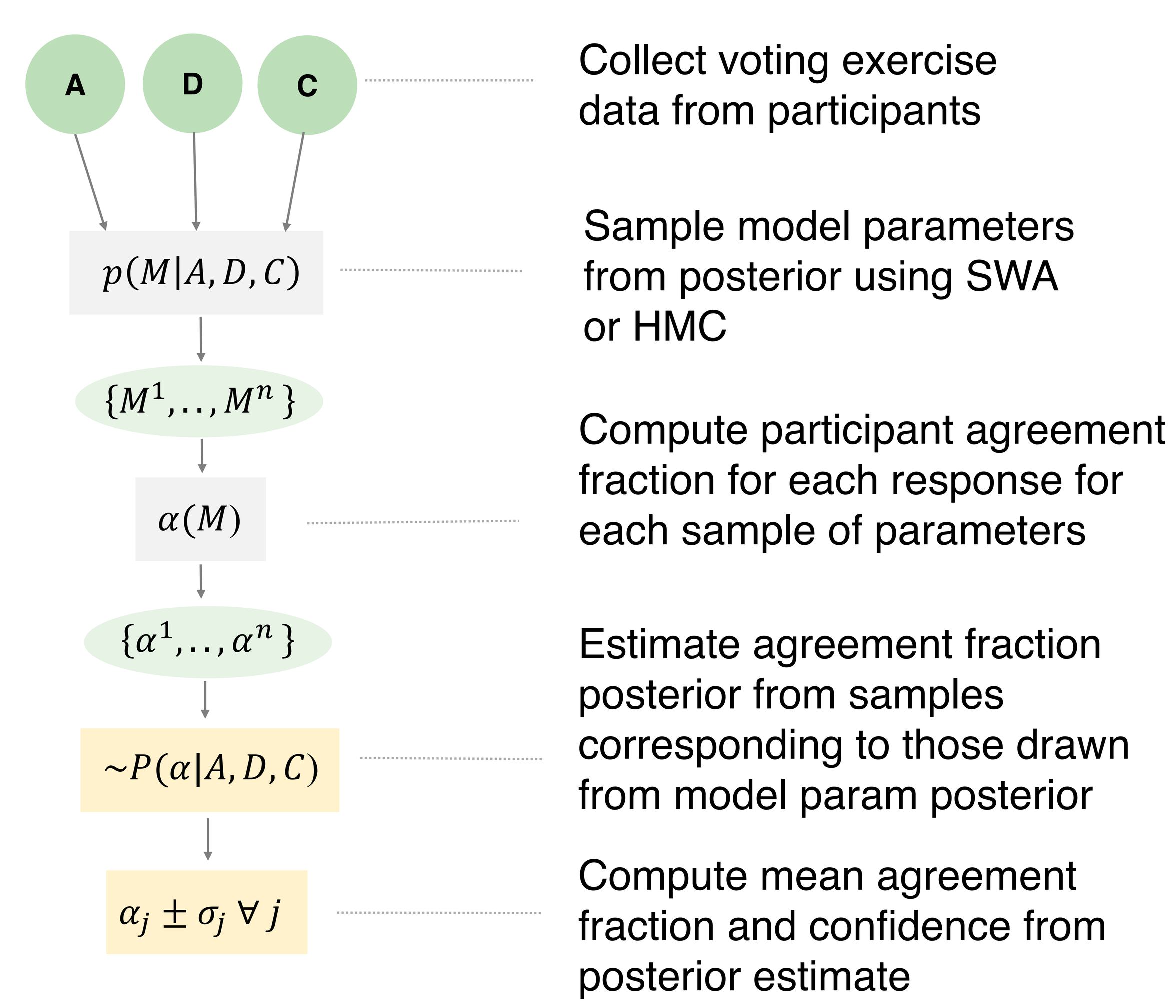


How many data points per response are needed?

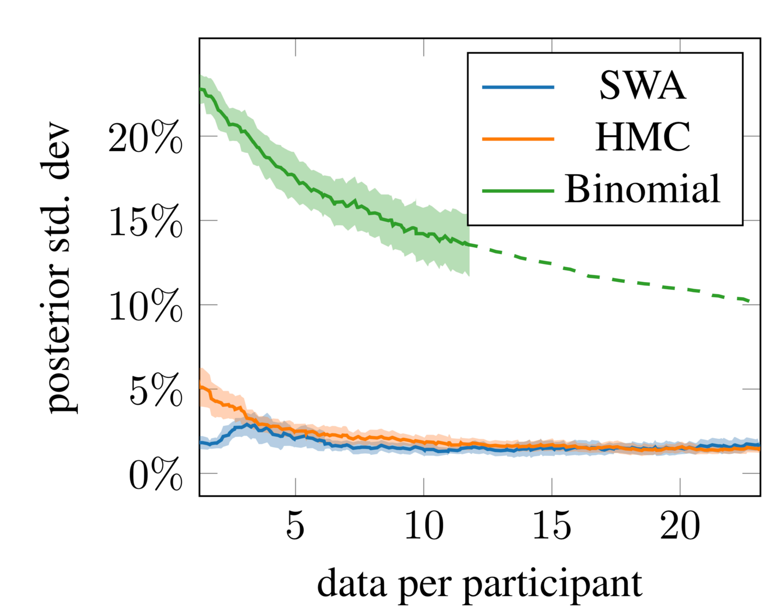


- Optimal parameters from SWA outperform HMC
- Accuracy saturates ~ 15 data points (exercises) per response
- With ~15 data points for a responses the remaining ~85 agreement data points are predicted with 70-80% accuracy

Confidence estimation



Results: Confidence



- Model yields higher confidence than binomial baseline (standard vote counting) for both HMC and SWA
- At 15 data points per person $\sigma = 1.5\%$ with HMC and SWA estimates differing by ~0.2% on average

Results: Compute time

	Mean runtime (s)		
DPP	SWA	HMC	MAE
2.5-5	9.44	356.31	9.63×10^{-3}
5-7.5	12.33	567.74	7.10×10^{-3}
7.5-10	11.10	833.41	5.95×10^{-3}
10-12.5	10.45	1186.67	4.77×10^{-3}
12.5-15	11.28	1216.39	3.18×10^{-3}
15-17.5	10.53	1407.59	1.90×10^{-3}
17.5-20	9.82	1809.92	2.03×10^{-3}
20-22.5	9.71	1881.24	1.96×10^{-3}

- At 15 data points per person HMC has a runtime of 23 minutes which is outside the scope of acceptability
- In contrast, SWA takes only 10 seconds – a 100x speedup.

Conclusions

- At 15 data points per person the model achieves confidence of $\sigma = 1.5\%$ in predicting the fraction of all participants which agree with each response.
- Using SWA, confidence can be estimated in ~10s.
- Using model presented and SWA, each dialogue cycle can take place in a few minutes and many cycles can take place over a one hour dialogue.

Risks & Policies

Risk 1: Non-representative data → Inaccurate results

Causes: biased questions, disengaged participants, non-representative population due to sample or malicious actors

Policies: require (a) dialogue moderators be trained in asking unbiased questions, (b) appropriate population sampling and participant validation scheme be applied (c) randomized human verification of data quality be regularly conducted

Risk 2: Bad prediction results → Inaccurate results

Causes: bad model, programming errors

Policies: require model performance verification take place on all new production deployments

Risk 3: Result misinterpretation→ Inaccurate conclusions

Causes: lack of proper context -- cultural, experiential, etc -- , confidence in results is miscalibrated.

Policies: require (a) relevant context be identified and then integrated into interpretation of results, (b) all ML-based results include estimates of confidence.

Key References

- United Nations Department of Political and Peacebuilding Affairs and Centre for Humanitarian Dialogue (2019). **Digital Technologies and Mediation in Armed Conflict.**
- Candes, E.J. (2010). **Matrix Completion With Noise**. *Proceedings of the IEEE*, 98. 925 - 936. 10.1109/JPROC.2009.2035722.
- Maddox, W.. (2019). **A Simple Baseline for Bayesian Uncertainty in Deep Learning**. *ArXiv*, abs/1902.02476.

contact: **andrew@remesh.org**