Understanding Disagreement: An Annotation Study of Sentiment and Emotional Language in Environmental Communication

Christina Barz, Melanie Siegel, Daniel Hanss & Michael Wiegand

The 19th Linguistic Annotation Workshop: July, 31, 2025

h_da darmstadt university of applied sciences
member of EUT+ European University of Technology
universität wien

Context & Aim

- Limited research on **how emotional language mobilizes collective action**
- Our goal is to analyze the emotional language used in environmental communication
- **Emotional language** = expressions conveying affective states
- **We focus on strategic use in group communication, not individual feelings**

- Part of a larger project on **emotional language in environmental activism**
- Case study: **Extinction Rebellion (XR)** — tweets analyzed for sentiment & emotional language
- Found substantial **annotator disagreement** → insights for annotation & perspectivism in NLP

Research Questions & Contributions

Research Questions:

- **RQ1**: What factors contribute to variation & disagreement in annotations?
- **RQ2**: How can we learn from disagreement to improve future annotation?

Main Contributions:

- First annotated, public dataset on XR's emotional language
- Systematic analysis of annotator disagreement & perspectives
- Implications for perspectivism: multiple valid interpretations → no single "ground truth"

Annotation Process

- 2,199 English tweets from **Extinction Rebellion** (2022–2024)
- **Document-level annotation**: tweets can have multiple labels
- **3 independent annotators**

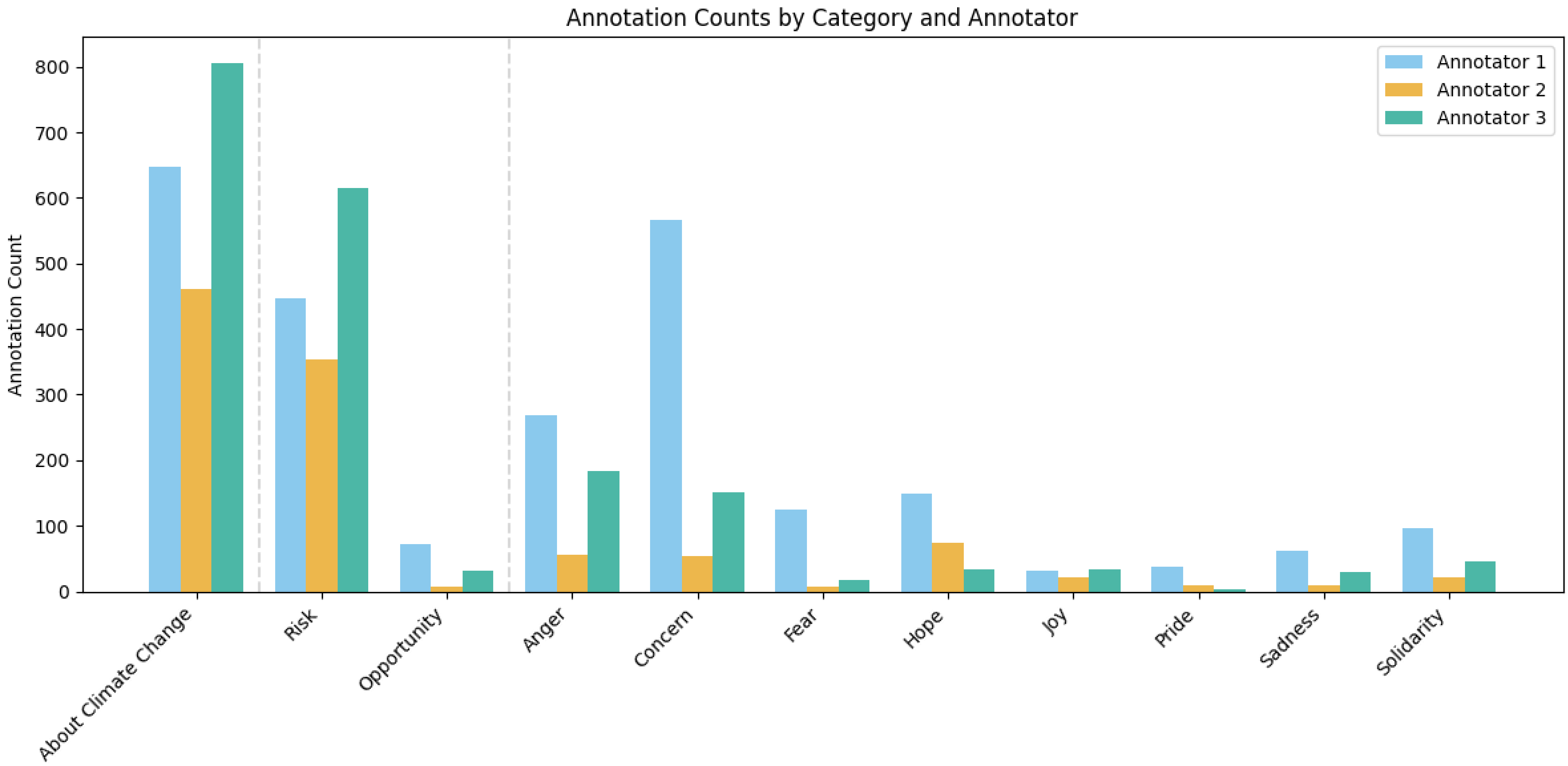
- Categories:
  - **Climate**: detection + sentiment (risk, opportunity, neutral)
  - **Emotions**: anger, concern, fear, hope, joy, pride, sadness, solidarity

- **Guidelines** with clear definitions & examples
- Small pilot (10 tweets) + individual **feedback sessions** after every 500 tweets

Understanding Annotator Disagreement

We analyzed **label frequency** per annotator *(2,199 tweets):*

- Large differences across annotators → e.g., *A2 more conservative*, A1 most frequent
- Emotions like **Pride, Joy & Sadness** rarely assigned



Annotation Counts by Category and Annotator
Annotation Count
0
100
200
300
400
500
600
700
800
Annotator 1
Annotator 2
Annotator 3
About Climate Change
Risk
Opportunity
Anger
Concern
Fear
Hope
Joy
Pride
Sadness
Solidarity

Understanding Annotator Disagreement

- Computed Fleiss' Kappa (overall) & Cohen's Kappa (pairwise) for all labels
- Compared agreement across categories & annotator pairs

	Climate Detection	Climate Sentiment	Anger	Concern	Fear	Hope	Joy	Pride	Sadness	Solidarity
Overall IAA: Fleiss'Kappa	0.4715	0.4571	0.2561	0.0978	0.0586	0.1781	0.4409	0.2293	0.1877	0.3647
1 vs 2: Cohen's Kappa	0.4766	0.4612	0.0754	0.0913	0.0385	0.2614	0.5127	0.2932	0.1367	0.2768
1 vs 3: Cohen's Kappa	0.6280	0.5583	0.4730	0.1930	0.1004	0.1092	0.4222	0.1400	0.1829	0.4204
2 vs 3: Cohen's Kappa	0.3173	0.3536	0.1242	0.1236	0.0754	0.1506	0.3927	0.3059	0.3033	0.4166

Understanding Annotator Disagreement

- Ran **Pointwise Mutual Information (PMI)** analysis for 3 frequent emotions
- Tested for **lexical bias** → do annotators rely on explicit emotion words?

ANGER			CONCERN			HOPE		
A1	**A2**	**A3**	**A1**	**A2**	**A3**	**A1**	**A2**	**A3**
murdering	tree	murdering	massively	corruption	warned	equitable	comments	**hope**
allow	hundred	**angry**	ongoing	threatening	massively	gather	expiration	touch
protested	immediate	denounce	escalating	reached	widely	preserve	helping	bit
false	helping	sleepwalking	allow	problems	horrific	joined	allowing	reasonable
lobbyists	training	address	twice	changing	suffer	motorway	degree	planning
sentence	lethal	hands	cultural	trust	deal	threats	faster	conference
murderous	claims	murderous	describes	result	ignore	achieve	linked	civilization
sleepwalking	politician	escalating	poorest	trees	positive	expiration	date	greed
polluting	camp	failure	tool	develop	propaganda	positive	ourselves	firm
exposing	release	behind	horrific	produce	further	voice	prevent	glass

Understanding Annotator Disagreement

- Applied BERTopic to analyze topic biases
- 17 topics identified (e.g., Climate Policy, Activist Action, Personal Responsibility)
- Emotion-Topic Link:
  Hope → strong variation: each annotator links Hope to different topics
  Anger & Concern → more overlap in topics across annotators

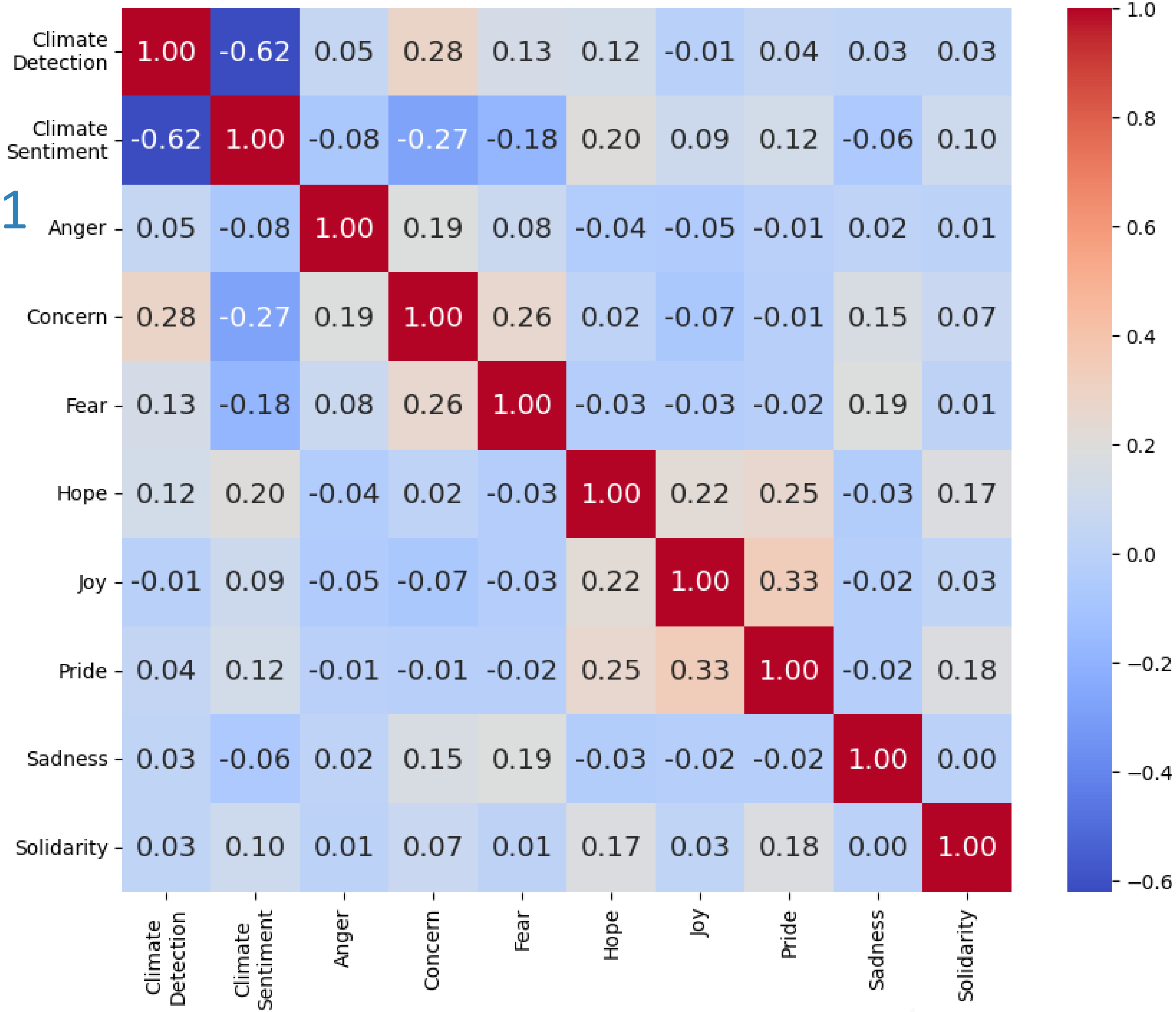
Understanding Annotator Disagreement

- Calculated **Spearman correlations** between all label pairs
- Analyzed co-labeling patterns per annotator → test how emotions cluster together

Key Results

- A1 & A2:
  - Higher correlations (up to 0.33) among positive emotions
  - Indicates tendency to co-label Hope, Joy, Pride, etc.
- A3:
  - Correlations near zero → clearer separation between emotion categories
  - Suggests stricter emotional differentiation

Correlations for A1



Qualitative Interviews

Emotions Mirror Labels: *Annotators often assign labels matching their own emotional reactions*

- A1 feels **concern** → labels *Concern* most
- A2 feels **anger** → labels *Anger* most
- A3 feels **fear** → labels *Fear* least

Group Imagery Shapes Perception: *Radical associations may trigger more emotional interpretations*

- A1: Thinks of **radical groups** (e.g., XR) → labels **more emotions**
- A2 & A3: Think of **moderate groups** (e.g., FFF, Greenpeace) → label **fewer emotions**

RQ2: How can we learn from disagreement to improve future annotation?

- **Annotator-specific metadata:** sociodemographics, stance, emotions, media exposure
- **Daily self-reports:** track mood, news events, or emotional shifts
- **Contextual variables:** natural disasters or political events can bias interpretation
- **Combine human annotation + LLMs to separate writer intention from reader perception**
- Supports **human-centered NLP** and **citizen science approaches**

Conclusion

- **Annotation disagreements in environmental communication are not flaws—they are meaningful**
- Environmental communication is **inherently subjective**; perspectives shape interpretation
- Following perspectivist NLP, disagreement reveals **diverse emotional and interpretative reactions** (Buechel & Hahn, 2022; Du et al., 2023; Ostarek et al., 2024)
- Rather than aiming for "ground truth," we treat variation as **data**, not noise (Cabitza et al., 2023; Uma et al., 2021; Rodríguez-Barroso et al., 2024; Valette, 2024)

References

Sven Buechel and Udo Hahn. 2022. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. *arXiv preprint arXiv:2205.01996*

Federico Cabitza, Andrea Campagner, and Valerio Basile. 2023. Toward a perspectivist turn in groundtruthing for predictive computing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 6860–6868.

Quanqi Du, Sofie Labat, Thomas Demeester, and Veronique Hoste. 2023. Unimodalities count as perspectives in multimodal emotion annotation. In *2nd Workshop on Perspectivist Approaches to NLP (NLPerspectives 2023), co-located with the 26th European Conference on Artificial Intelligence (ECAI 2023)*, volume 3494. CEUR-WS. org.

Markus Ostarek, Brent Simpson, Cathy Rogers, and James Ozden. 2024. Radical climate protests linked to increases in public support for moderate organizations. *Nature Sustainability*, pages 1–7.

Nuria Rodriguez-Barroso, Eugenio Martinez Camara, Jose Camacho Collados, M Victoria Luzon, and Francisco Herrera. 2024. Federated learning for exploiting annotators' disagreements in natural language processing. *Transactions of the Association for Computational Linguistics*, 12:630–648.

Alexandra N Uma, Tommaso Fornaciari, Dirk Hovy, Silviu Paun, Barbara Plank, and Massimo Poesio. 2021. Learning from disagreement: A survey. Journal of *Artificial Intelligence Research*, 72:1385–1470.

Mathieu Valette. 2024. What does perspectivism mean? An ethical and methodological countercriticism. In *Proceedings of the 3rd Workshop on Perspectivist Approaches to NLP (NLPerspectives) @LREC-COLING 2024,* pages 111–115, Torino, Italia. ELRA and ICCL.

Thank you for your attention!