



NormBank: A Knowledge Bank of Situational Social Norms

Caleb Ziems, Jane A. Yu, Yi-Chia Wang, Alon Y. Halevy, Diyi Yang

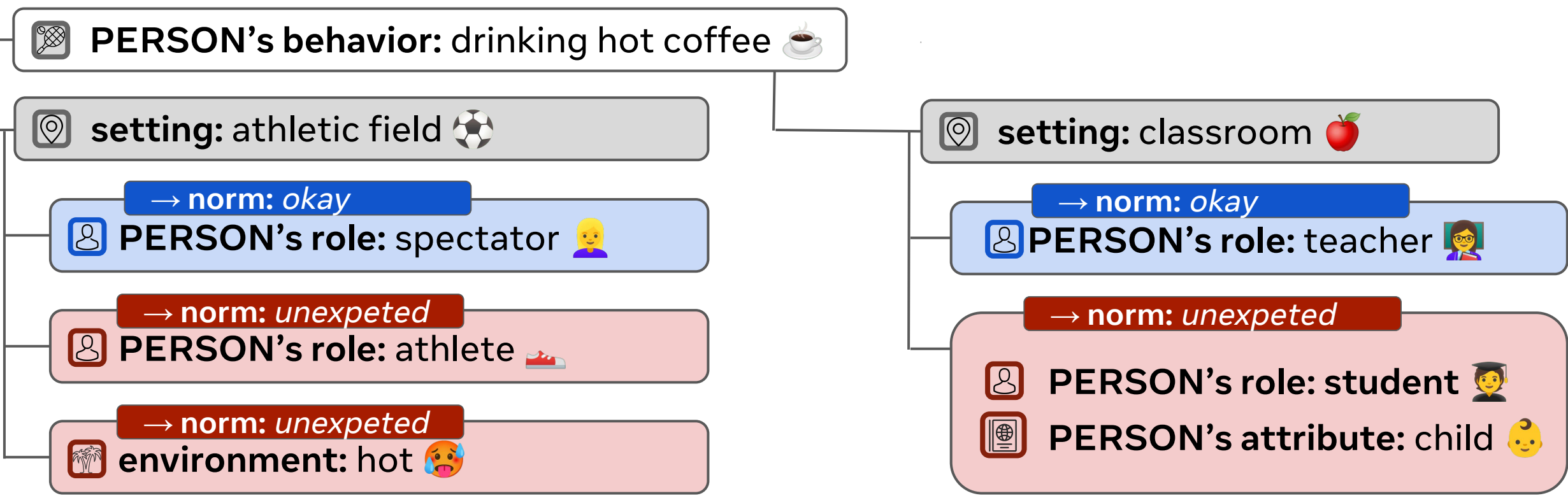
Stanford University, Meta AI Research

cziems@stanford.edu



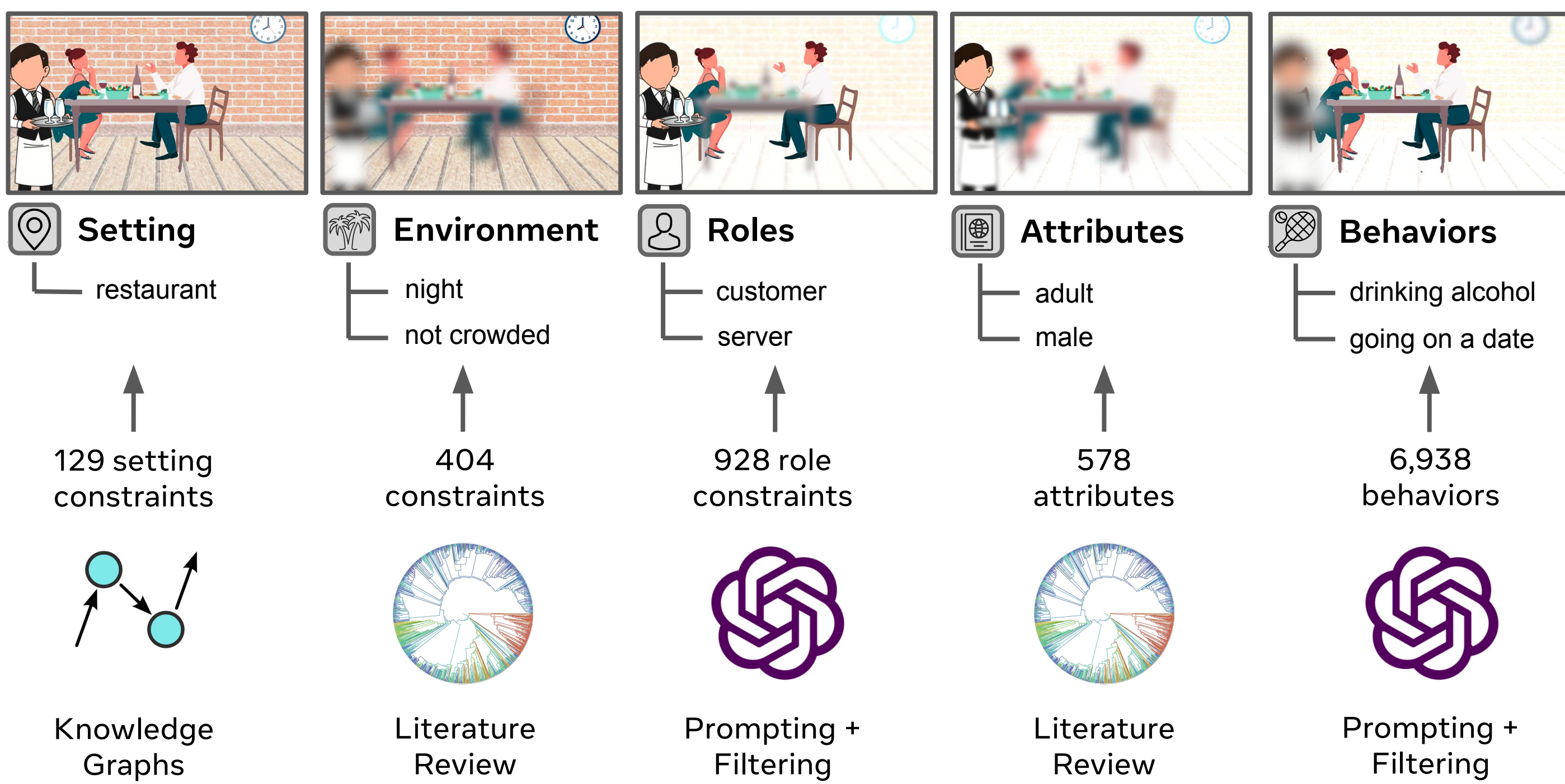
1. Introduction: Normative Reasoning

- **Motivation:** Assistive and collaborative AI need to understand flexible (defeasible) social norms
 - largely missing from pre-training data
 - current models struggle to update inferences with new information beyond the most prototypical cases



- **Contributions:**
 - **SCENE** — a taxonomy for constraining norms within their sociocultural frame
 - **NormBank** — a knowledge bank of 155k situational norms
 - neural knowledge-completion models
 - empirical evidence for the utility of NormBank on downstream tasks

2. SCENE: A Dramaturgical Taxonomy

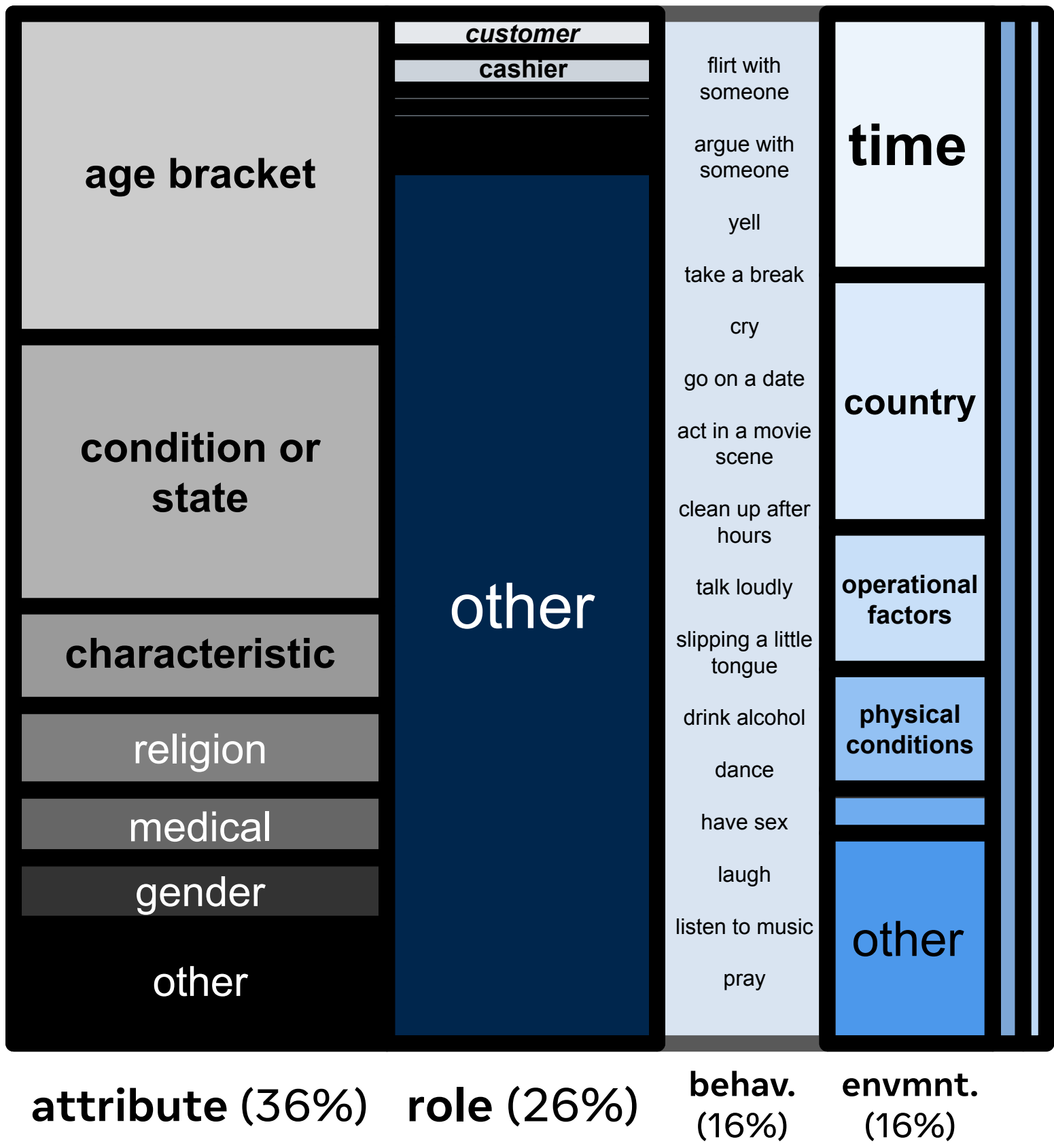


3. Annotation: Building NormBank

- **HIT:** start w/ behaviors and ask for *contexts* under which the behavior is:
 - norm: expected ✓
 - norm: okay ☁
 - norm: unexpected ✗
 - highly creative task
 - scaffolded by SCENE
 - rich contrast sets; fewer spurious artefacts
- **Quality Control:**
 - **qualification test:** six questions on taxonomy definitions
 - **staging round:** small-scale task (not feeding into the final dataset) where workers get personalized feedback
 - **programmatic filtering; random audits**

4. Summary Statistics

Distinct.....Constraints: 408,407
Taxonomic.....Constraints: 93.6%
Pre-Pop.....Constraints: 69.2%
Distinct.....Norms: 155,423
Social.....Situations: 70,215
Distinct.....Behaviors: 6,938



5. Experiments: Automatic Expansion

Classification:

given an unseen behavior + constraint set, predict the norm (expected, okay, unexpected)

Model	Precision	Recall	F1 Score	Accuracy
RoBERTa	73.3%	71.4%	72.1%	75.4%

Generation:

given an unseen behavior + norm label, generate constraints that make the norm label true

Model	Sensible Norm	Correct Norm	Normative Constraint	Helpful Constraint
GPT-3	95.0%	61.1%	91.8%	87.8%
BART	100.0%	46.0%	94.3%	94.3%
Human	82.5%	55.0%	72.9%	81.7%
Max	100.0%	100.0%	100.0%	100.0%

6. Experiments: Downstream Transfer

(via sequential fine-tuning)

	Base	+ NormBank
ANECDOTES	68.3%	68.7% ★
DILEMMAS	64.3%	71.1% ★
CosmosQA	59.8%	61.2%
Social IQa	59.9%	64.2%

7. Limitations + Ethical Considerations

- NormBank is descriptive rather than prescriptive (NOT to give you advice)
- Annotator Demographics: US English-Speakers
- Future work can evaluate transfer utility on additional downstream tasks
- SCENE's role and attribute constraints will allow more targeted bias mitigation efforts