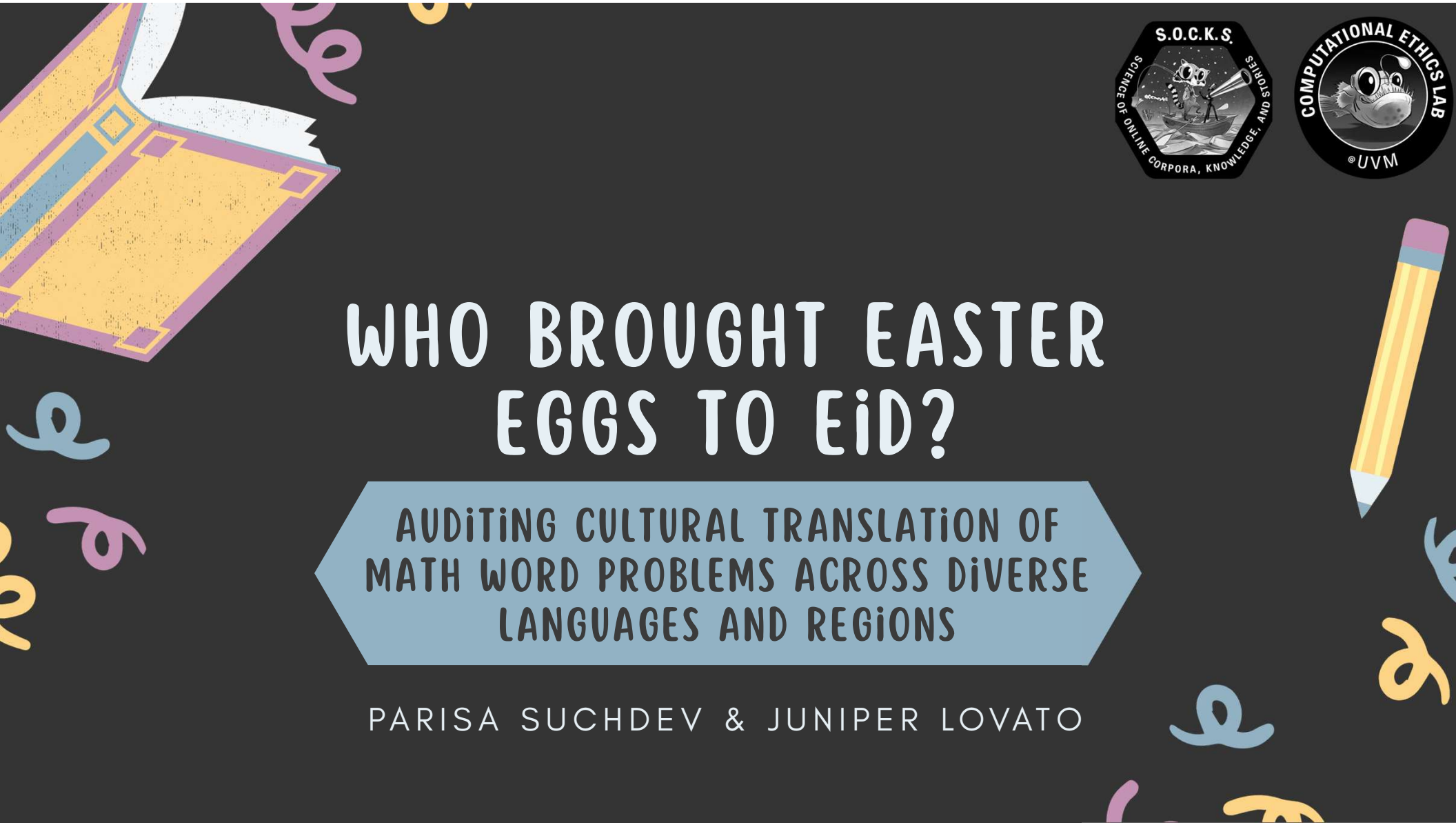


WHO BROUGHT EASTER EGGS TO EID?

AUDITING CULTURAL TRANSLATION OF MATH WORD PROBLEMS ACROSS DIVERSE LANGUAGES AND REGIONS

PARISA SUCHDEV & JUNIPER LOVATO

THE MARKET MATHEMATICIANS

Recife, Brazil, 1980s:

- Child street vendors (age 9+) calculating prices and making change in markets

On the street: 98% accuracy

- Real transactions, familiar context, everyday mathematics

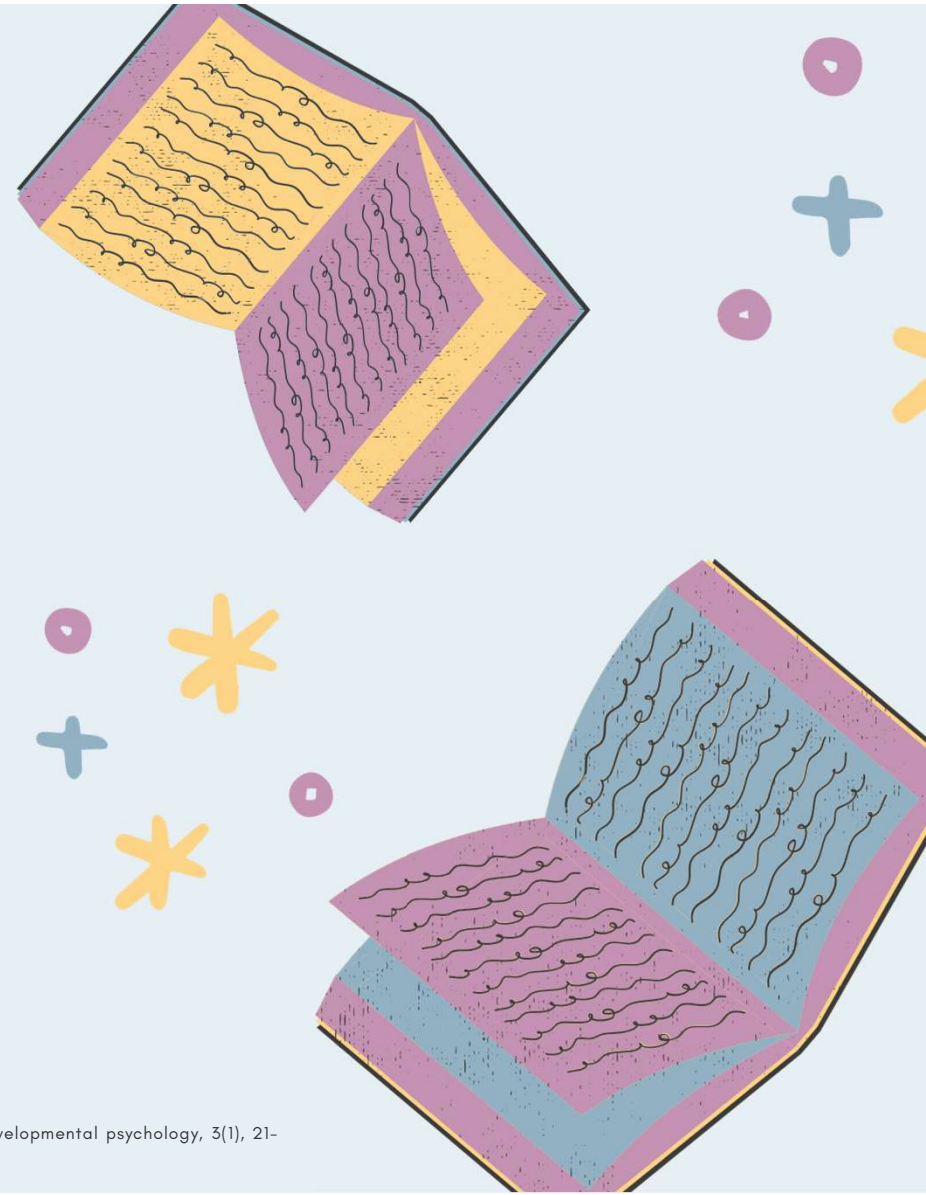
Same problems on paper in school: 37% accuracy

- Same math, different context

What changed?

- Not their mathematical ability, but the languages, names, context, social situation, and cultural meaning

Mathematical competence isn't just in the mind.
It's the mind situated within a cultural context



Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. British journal of developmental psychology, 3(1), 21-29.

WHAT DO MEAN BY CULTURAL ADAPTATION OF MATH WORD PROBLEMS?

ENGLISH MATH WORD PROBLEM

"**Sarah** bought 3 **apples** at the **farmer's market** for **$**2 each..."

→ **Cultural Translation** →

ADAPTED FOR ITALIAN

"**Sofia** ha comprato 3 **mele** al **mercato** per **€**2 ciascuna..."

THE NAME CHANGED. SARAH BECAME SOFIA. THE CURRENCY CHANGED. DOLLARS BECAME EUROS.

THE QUESTION IS: IS THAT ENOUGH?

CULTURE TREE FRAMEWORK

In culturally responsive teaching, cultural adaptation is NOT one thing, it operates on different levels

SURFACE CULTURE (VISIBLE, EASY TO CHANGE)

- Food, dress, music, holidays, celebrations
- The decorative elements you can see and touch

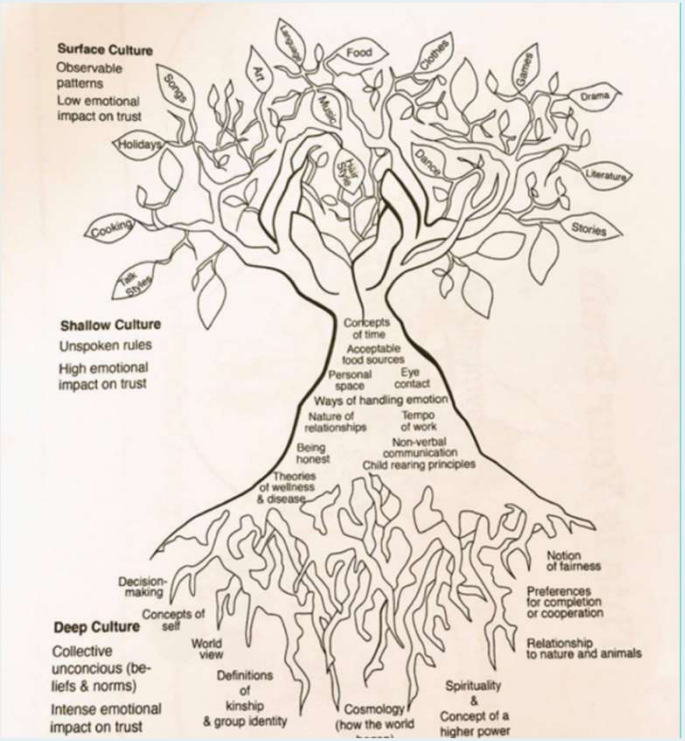
SHALLOW CULTURE (BELOW THE SURFACE)

- Courtesy, conversational patterns, personal space
- Social norms and communication styles

DEEP CULTURE (INVISIBLE ROOTS)

- Values, worldview, beliefs about fairness, time, learning
- How people make sense of the world

Deep culture grounds identity and shapes how students process information.



Hammond, Z. (2014). Culturally responsive teaching and the brain: Promoting authentic engagement and rigor among culturally and linguistically diverse students. Corwin Press.

HOW TEACHERS DO CULTURAL ADAPTATION

IT'S SKILLED PEDAGOGICAL WORK

TEACHERS DRAW ON DEEP KNOWLEDGE:

- Students' lived experiences, family routines, community practices
- What matters for mathematical understanding vs. surface decoration
- Ground instruction in students' funds of knowledge

THIS IS CULTURALLY RESPONSIVE TEACHING

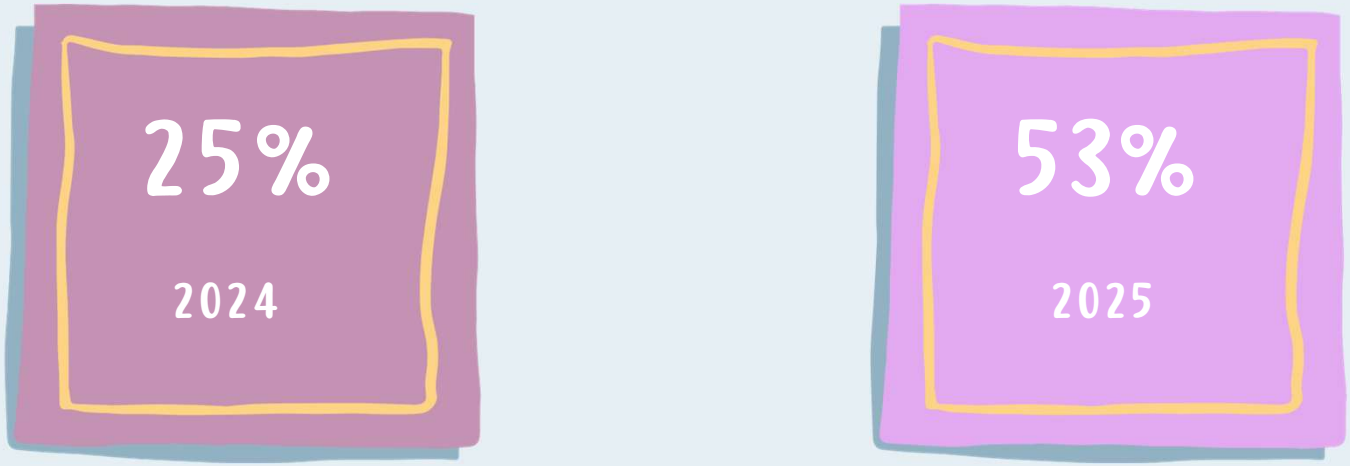
- Not just swapping names and currencies (surface culture)
- Understanding students' worldviews, learning approaches (deep culture)
- Making the math accessible without making it culturally incoherent

Moll, L., Amanti, C., Neff, D., & Gonzalez, N. (2006). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. In Funds of knowledge (pp. 71-87). Routledge

Gay, G. (2018). Culturally responsive teaching: Theory, research, and practice. teachers college press.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. American educational research journal, 32(3), 465-491..

WHAT iS THE PROBLEM: LLMS
WELL.. THEY ARE ALSO A SOLUTiON TO LiMiTED RESOURCES



of English language arts, math, and science teachers reported using AI for instructional planning or teaching

Common use: Translating and culturally adapting word problems for multilingual classrooms

Doss, C. J., Bozick, R., Schwartz, H. L., Chu, L., Rainey, L., Woo, A., ... & Dukes, J. (2025). AI use in schools is quickly increasing but guidance lags behind: Findings from the RAND survey panels. RAND.

RESEARCH QUESTIONS

Teachers do this with expertise and care. Now LLMs are increasingly doing this work. How do LLMs make these cultural choices?

01 **RQ1. Cross-model consistency:**
Do different LLMs produce the same cultural outputs?
→ **Model choice is a cultural decision, not just technical**

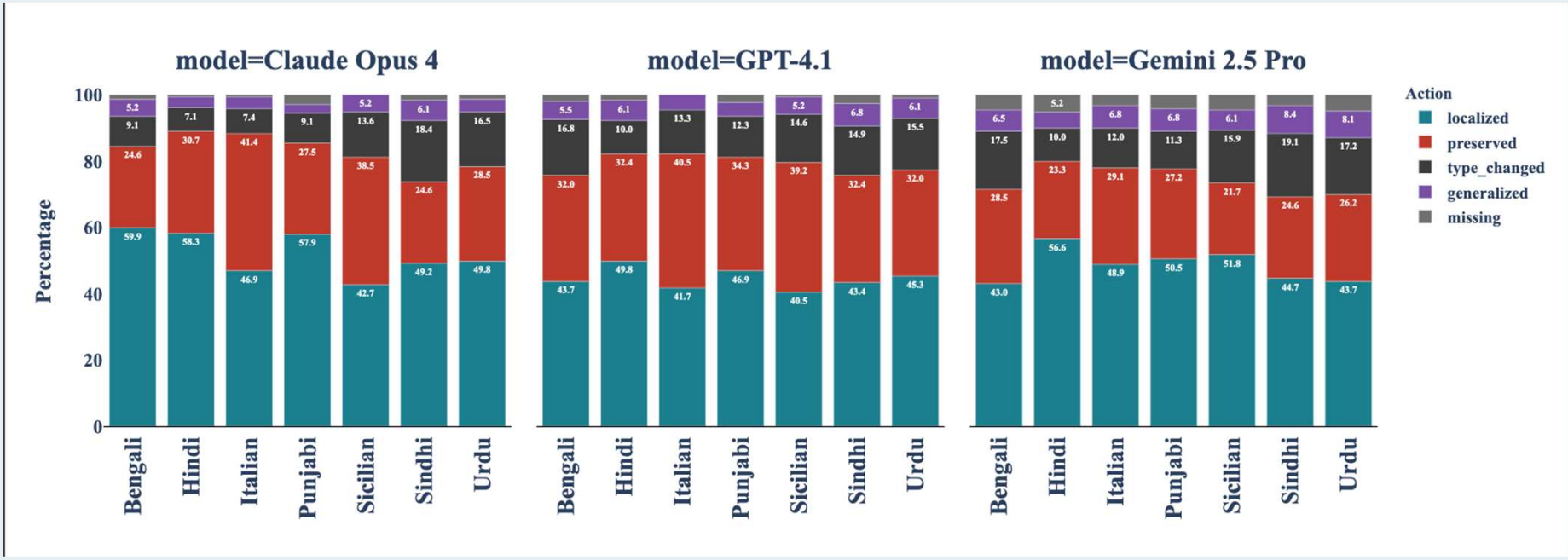
02 **RQ2. Cultural diversity:**
Does adaptation preserve or compress cultural variety?
→ **Localization at scale may paradoxically reduce diversity**

03 **RQ3. Cultural salience:**
Which entities do LLMs treat as culturally important?
→ **Do model priorities match teacher judgment?**

STUDY DESIGN

1. **Source**
   a. 60 English math problems from GSM-8K
   b. 66 entities identified (names, foods, places, currencies, etc.)
2. **Translation Task**
   a. Asked **3 frontier LLMs** to translate and culturally adapt each problem
   b. Models: Claude Opus 4, GPT-4.1, Gemini 2.5 Pro
   c. Target languages:
      a. India: Bengali, Hindi, Punjabi
      b. Pakistan: Urdu, Sindhi
      c. Italy: Italian, Sicilian
3. **Analysis**
   a. Tracked what happened to each entity across translations
   b. Five action types: preserved, localized, generalized, type_changed, missing

ACTiON DiSTRiBUTiON



RQ1: DO MODELS AGREE ON CULTURAL OUTPUTS?

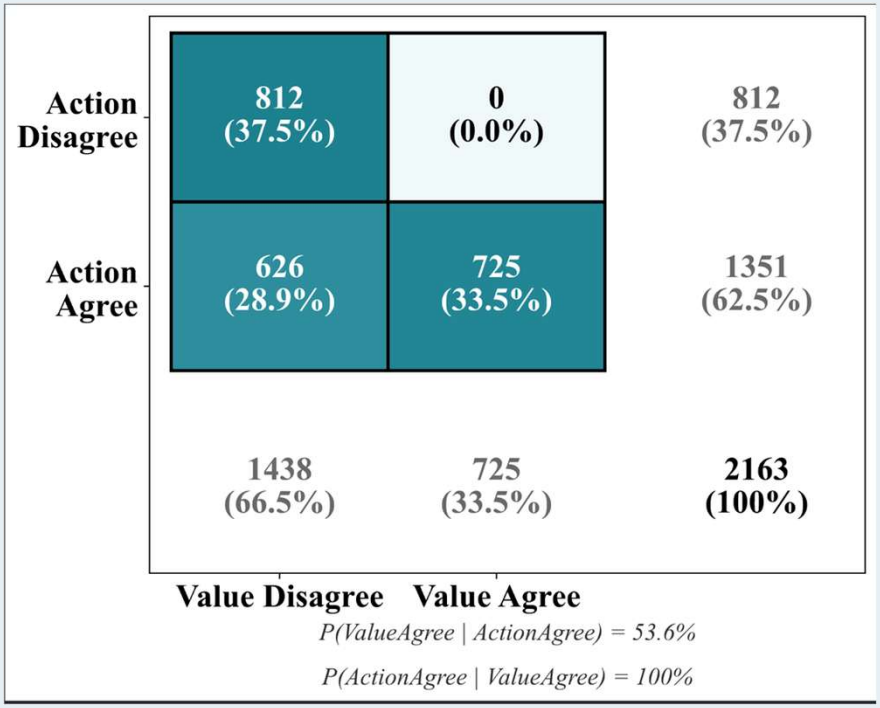
Level 1: Action Agreement (62.5%)

- Do all 3 models make the same type of decision?
- Example: All chose "localize" vs. one preserved, two localized
- Agreement rate: 62.5%

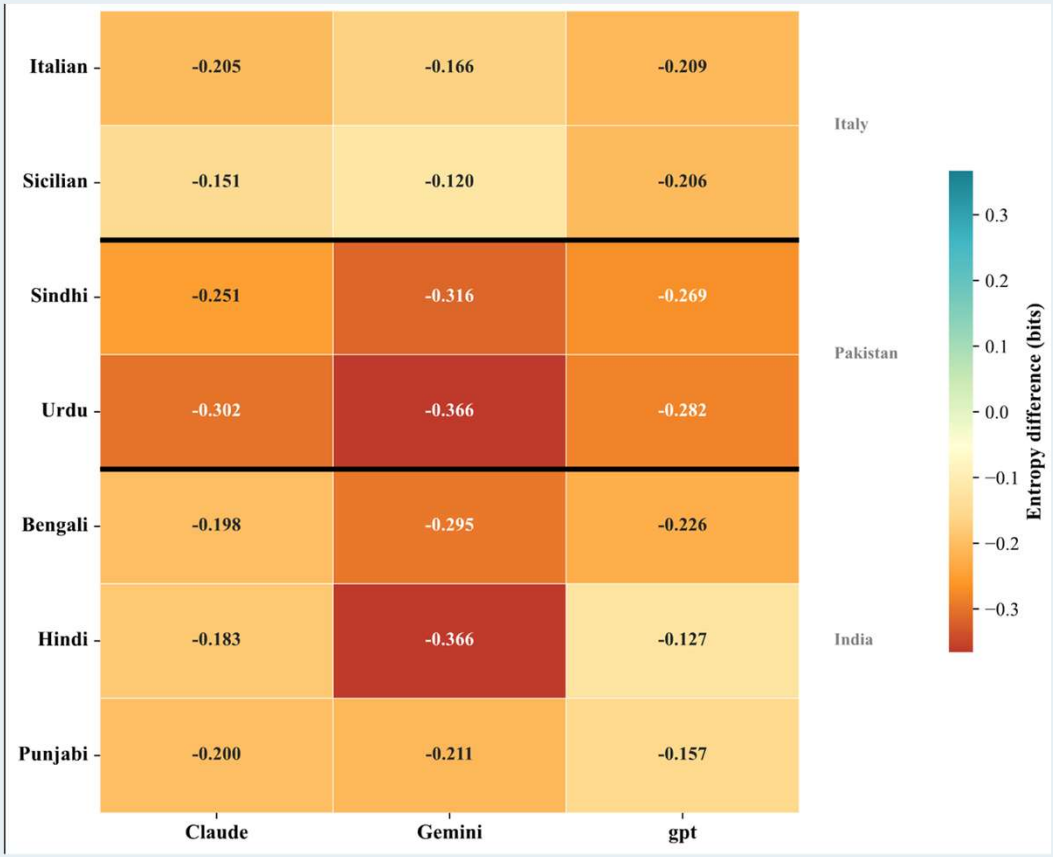
Level 2: Full Agreement - Action AND Value (33.5%)

- Do all 3 models take the same action AND produce the same output?
- Example: All localized "Sarah" → All chose "Sofia"
- Agreement rate: 33.5%

Model choice determines which cultural world students encounter



RQ2: DOES ADAPTATION COMPRESSES DIVERSITY?



How we measured cultural variety:

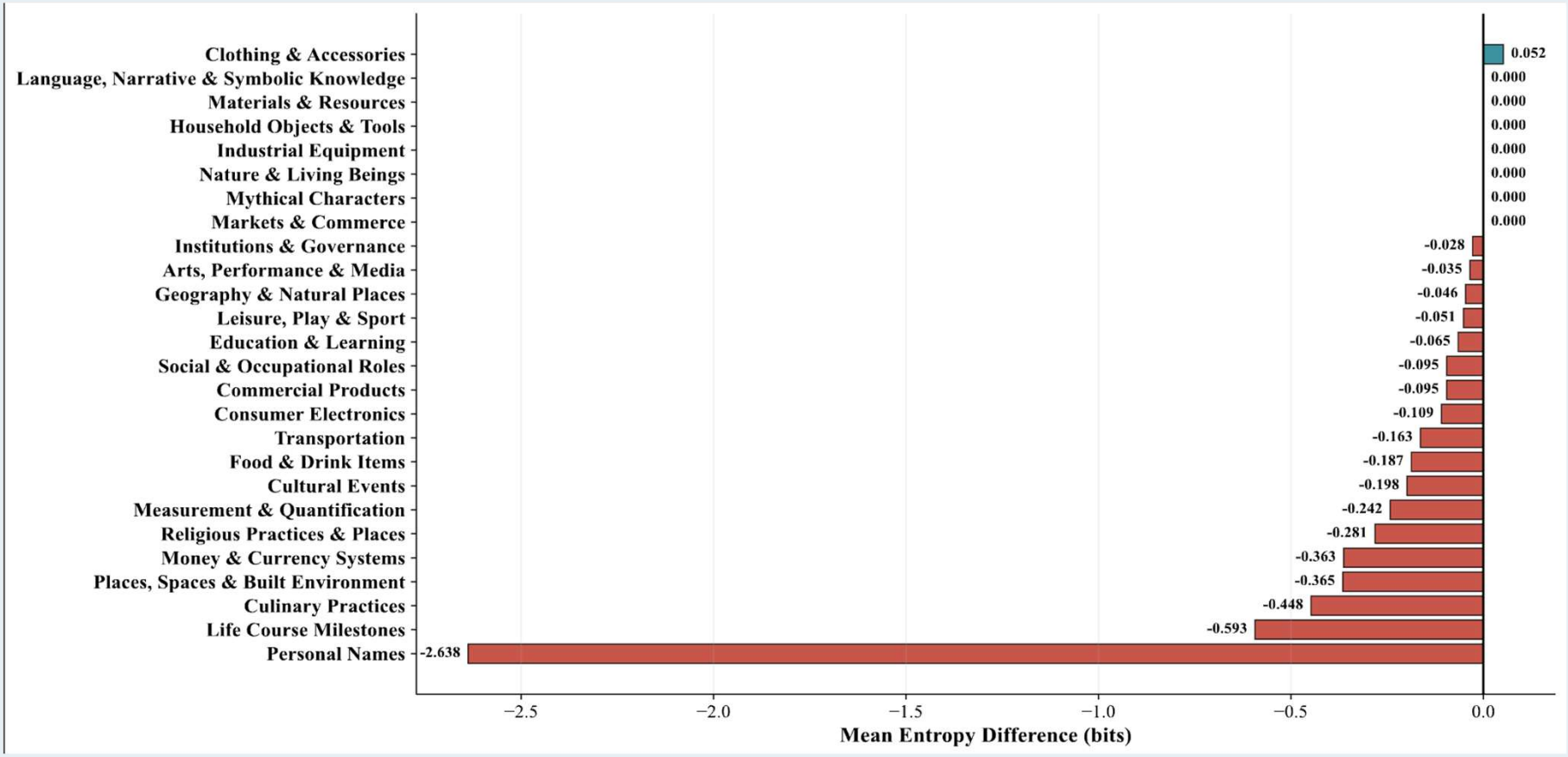
Entropy = measure of diversity/surprise

- High entropy: Many different entity types appear
- Low entropy: Same entities appear repeatedly

We compared:

- Source (English): How much variety exists in original problems?
- Adapted (7 languages × 3 models): How much variety in each adaptation?
- **Entropy difference:** Adapted entropy - Source entropy

RQ2: DOES ADAPTATION COMPRESSES DIVERSITY?



RQ3: WHICH ENTITIES DO LLMS TREAT AS CULTURALLY IMPORTANT?

Models CHANGE:

- ✓ Personal names (91.6%)
- ✓ Currency (94.3%)
- ✓ Food items (78.6%)

Models PRESERVE:

- ✗ Grade levels (100%)
- ✗ Roles: student, teacher (98%)
- ✗ Seasonal references (76%)

RQ3: THE EID HUNT PROBLEM

English source:

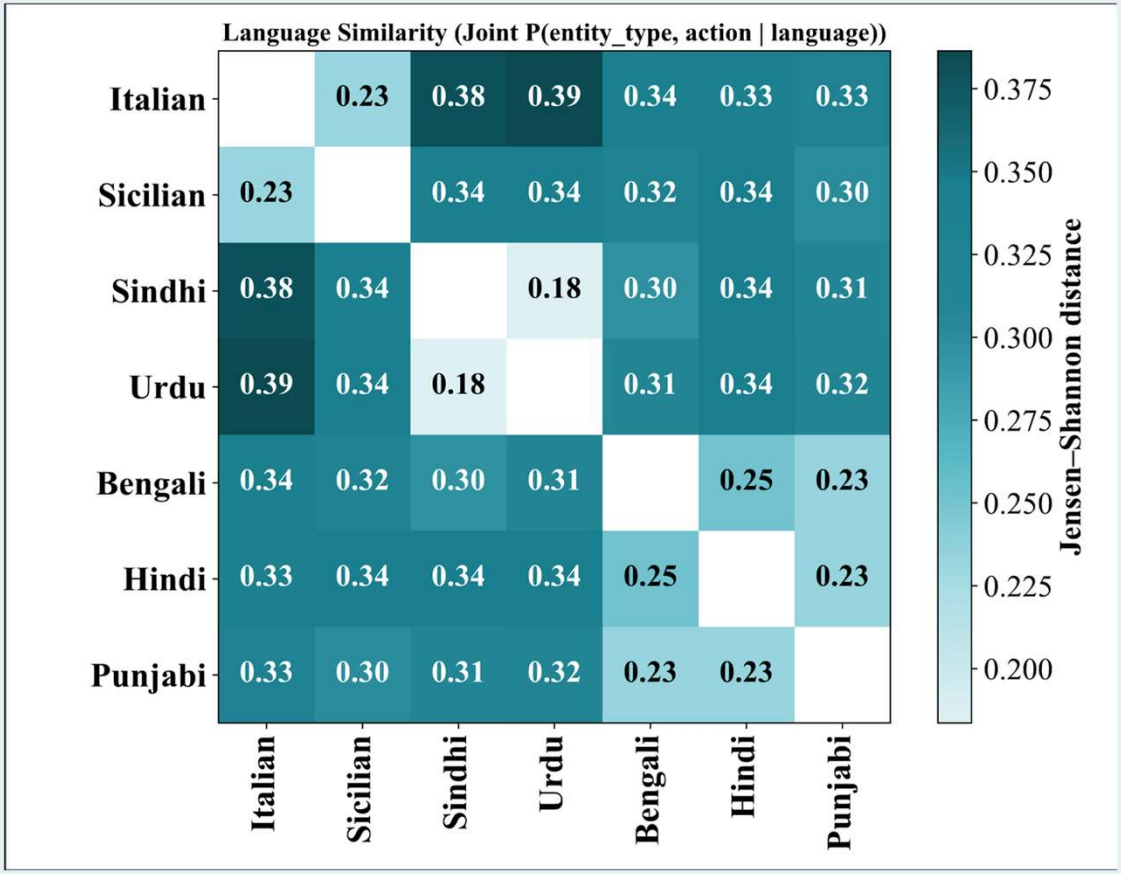
"Students organized an Easter egg hunt..."

Urdu/Sindhi adaptations:

"Students organized an egg search competition on Eid..."

✓ Holiday swapped (Easter → Eid)
✗ **Activity doesn't match** (egg hunts aren't an Eid tradition)
= Cross-cultural contamination

RQ3: REGIONAL CLUSTERING



How we measured cross-linguistic similarity:
Jensen-Shannon Distance (JSD) = how different are adaptation patterns?

- Compares: Which entity types get which actions in each language
- Range: 0 (identical patterns) to 1 (completely different)
- Lower values = more similar adaptation behavior

RQ3: REGIONAL MISSATTRIBUTION

PROMPT: "Teaching students in Bengali in India"

CURRENCY USED:

76.2% Bangladeshi taka *(Hindi: 100% Indian rupees)*

CULTURAL REFERENCE:

"Amar Sonar Bangla" (Bangladesh national anthem)

Training data: Bengali = Bangladesh

Explicit prompts couldn't override learned associations

TO CONCLUDE

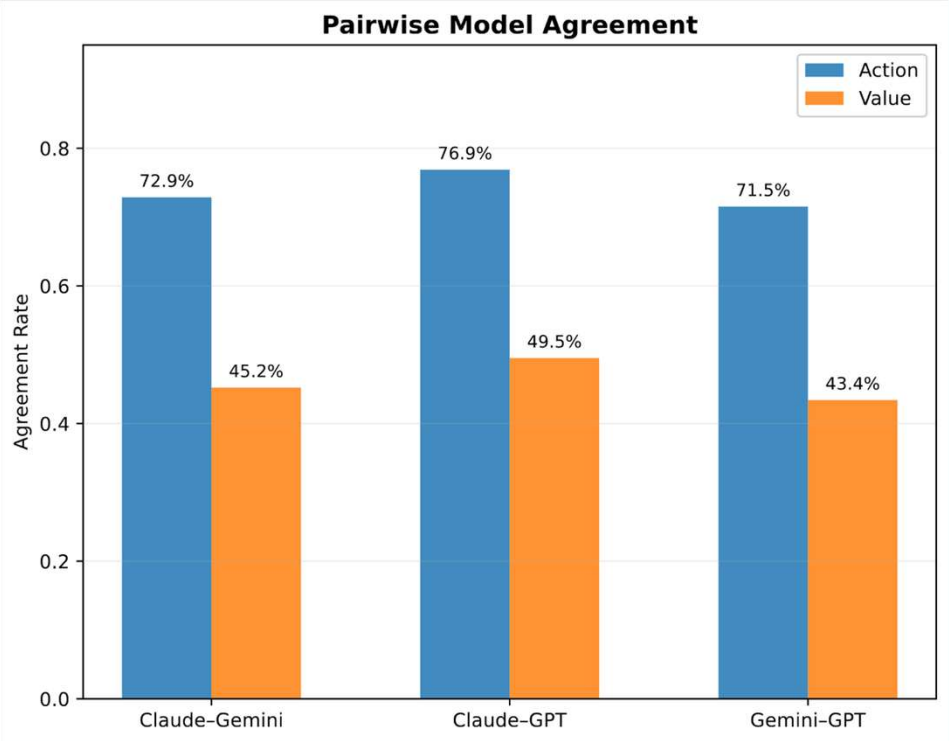
- Given this prompt, parameters, models in this study:

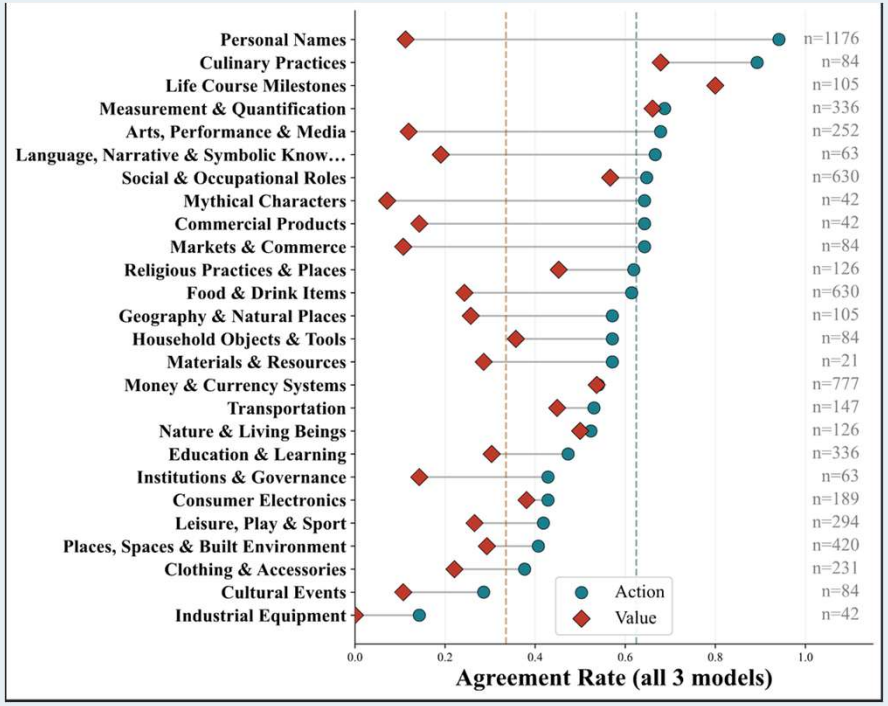
- **Are models stable?** No — models disagree 67% of the time, they choose different cultural worlds (They are not designed to produced the same results)
- **Are they diverse?** No — adaptation compresses variety across 7 languages, 3 models, 60 math word problems (homogeneity affects)
- **Are they culturally grounded?** No — they work at surface level, cross-cultural and cross-regional contamination (empathy can't be coded)

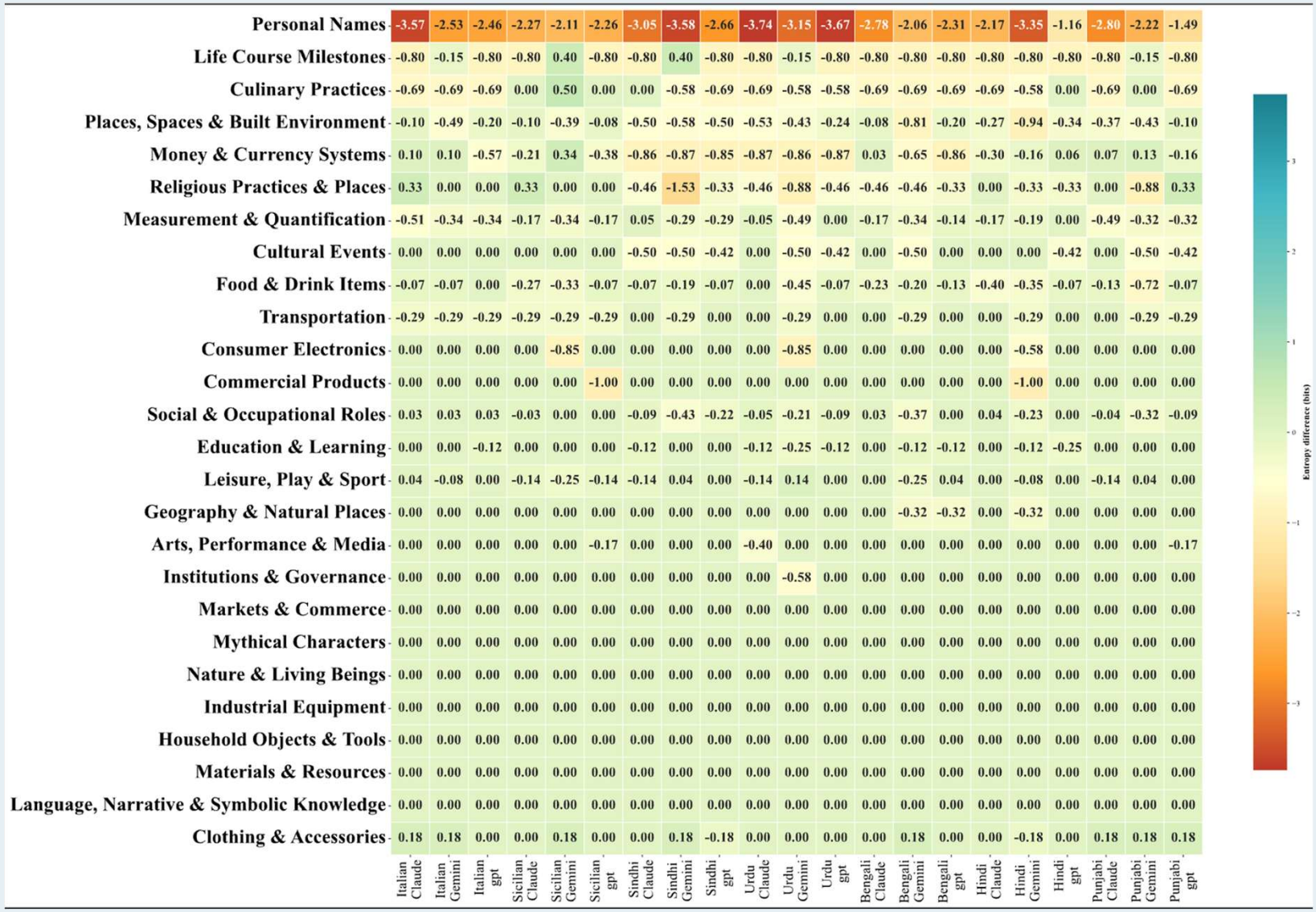
Surface plausibility makes culturally adapted problems look correct. That's exactly what makes deeper failures easy to overlook.

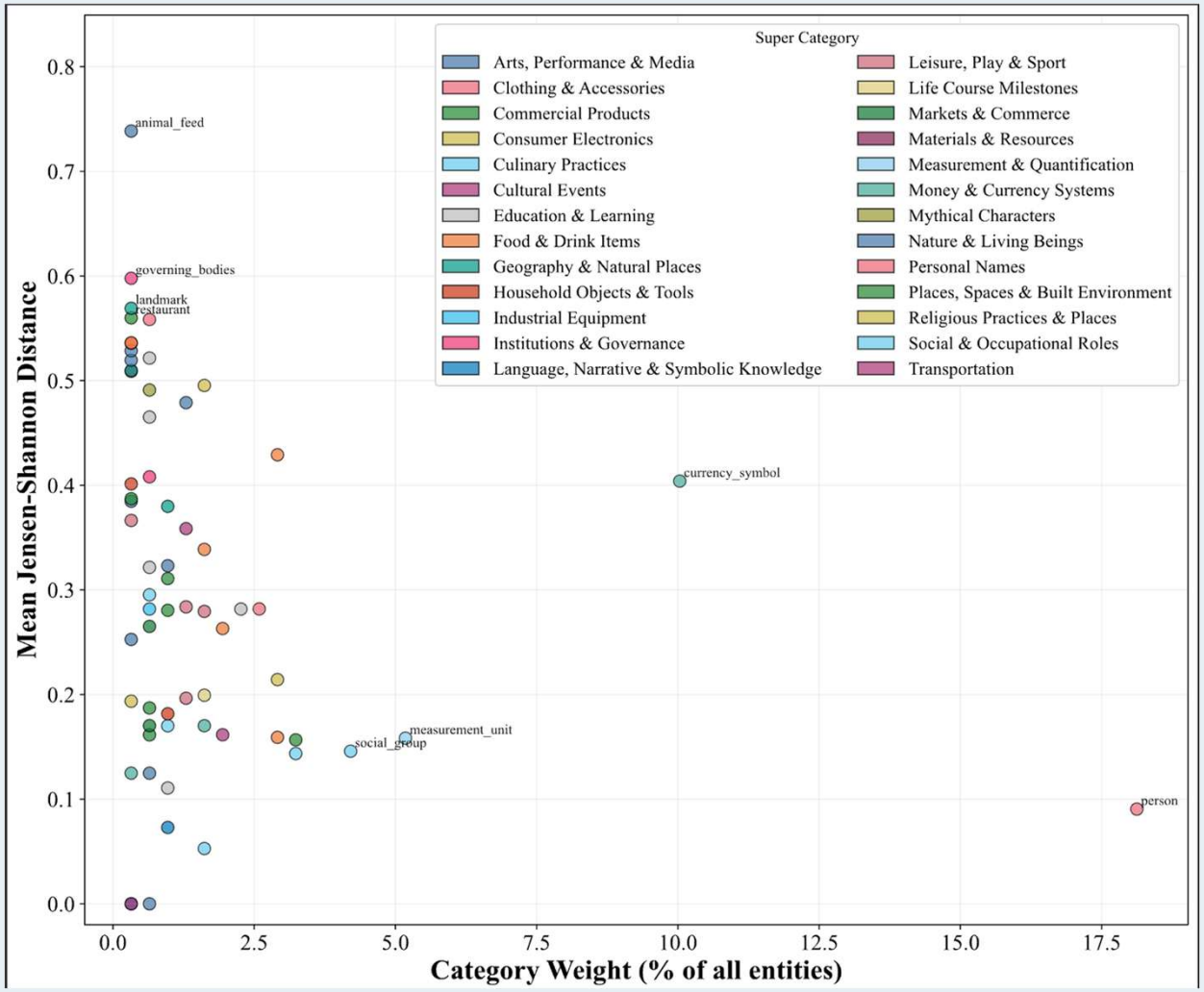
QUESTIONS, REMARKS, RECOMMENDATIONS?

RQ1: MODELS DISAGREE ON CULTURAL OUTPUTS






Personal Names -3.57 -2.53 -2.46 -2.27 -2.11 -2.26 -3.05 -3.58 -2.66 -3.74 -3.15 -3.67 -2.78 -2.06 -2.31 -2.17 -3.35 -1.16 -2.80 -2.22 -1.49
Life Course Milestones -0.80 -0.15 -0.80 -0.80 0.40 -0.80 -0.80 0.40 -0.80 -0.80 -0.15 -0.80 -0.80 -0.80 -0.80 -0.80 -0.80 -0.80 -0.80 -0.15 -0.80
Culinary Practices -0.69 -0.69 -0.69 0.00 0.50 0.00 0.00 -0.58 -0.69 -0.69 -0.58 -0.58 -0.69 -0.69 -0.69 -0.69 -0.58 0.00 -0.69 0.00 -0.69
Places, Spaces & Built Environment -0.10 -0.49 -0.20 -0.10 -0.39 -0.08 -0.50 -0.58 -0.50 -0.53 -0.43 -0.24 -0.08 -0.81 -0.20 -0.27 -0.94 -0.34 -0.37 -0.43 -0.10
Money & Currency Systems 0.10 0.10 -0.57 -0.21 0.34 -0.38 -0.86 -0.87 -0.85 -0.87 -0.86 -0.87 0.03 -0.65 -0.86 -0.30 -0.16 0.06 0.07 0.13 -0.16
Religious Practices & Places 0.33 0.00 0.00 0.33 0.00 0.00 -0.46 -1.53 -0.33 -0.46 -0.88 -0.46 -0.46 -0.46 -0.33 0.00 -0.33 -0.33 0.00 -0.88 0.33
Measurement & Quantification -0.51 -0.34 -0.34 -0.17 -0.34 -0.17 0.05 -0.29 -0.29 -0.05 -0.49 0.00 -0.17 -0.34 -0.14 -0.17 -0.19 0.00 -0.49 -0.32 -0.32
Cultural Events 0.00 0.00 0.00 0.00 0.00 0.00 -0.50 -0.50 -0.42 0.00 -0.50 -0.42 0.00 -0.50 0.00 0.00 0.00 -0.42 0.00 -0.50 -0.42
Food & Drink Items -0.07 -0.07 0.00 -0.27 -0.33 -0.07 -0.07 -0.19 -0.07 0.00 -0.45 -0.07 -0.23 -0.20 -0.13 -0.40 -0.35 -0.07 -0.13 -0.72 -0.07
Transportation -0.29 -0.29 -0.29 -0.29 -0.29 -0.29 0.00 -0.29 0.00 0.00 -0.29 0.00 0.00 -0.29 0.00 0.00 -0.29 0.00 0.00 -0.29 -0.29
Consumer Electronics 0.00 0.00 0.00 0.00 -0.85 0.00 0.00 0.00 0.00 0.00 -0.85 0.00 0.00 0.00 0.00 0.00 -0.58 0.00 0.00 0.00 0.00
Commercial Products 0.00 0.00 0.00 0.00 0.00 -1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 -1.00 0.00 0.00 0.00 0.00
Social & Occupational Roles 0.03 0.03 0.03 -0.03 0.00 0.00 -0.09 -0.43 -0.22 -0.05 -0.21 -0.09 0.03 -0.37 0.00 0.04 -0.23 0.00 -0.04 -0.32 -0.09
Education & Learning 0.00 0.00 -0.12 0.00 0.00 0.00 -0.12 0.00 0.00 -0.12 -0.25 -0.12 0.00 -0.12 -0.12 0.00 -0.12 -0.25 0.00 0.00 0.00
Leisure, Play & Sport 0.04 -0.08 0.00 -0.14 -0.25 -0.14 -0.14 0.04 0.00 -0.14 0.14 0.00 0.00 -0.25 0.04 0.00 -0.08 0.00 -0.14 0.04 0.00
Geography & Natural Places 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 -0.32 -0.32 0.00 -0.32 0.00 0.00 0.00 0.00
Arts, Performance & Media 0.00 0.00 0.00 0.00 0.00 -0.17 0.00 0.00 0.00 -0.40 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 -0.17
Institutions & Governance 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 -0.58 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Markets & Commerce 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Mythical Characters 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Nature & Living Beings 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Industrial Equipment 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Household Objects & Tools 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Materials & Resources 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Language, Narrative & Symbolic Knowledge 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Clothing & Accessories 0.18 0.18 0.00 0.00 0.18 0.00 0.00 0.18 -0.18 0.00 0.00 0.00 0.00 0.18 0.00 0.00 -0.18 0.00 0.18 0.18 0.18
Italian Claude
Italian Gemini
Italian gpt
Sicilian Claude
Sicilian Gemini
Sicilian gpt
Sindhi Claude
Sindhi Gemini
Sindhi gpt
Urdu Claude
Urdu Gemini
Urdu gpt
Bengali Claude
Bengali Gemini
Bengali gpt
Hindi Claude
Hindi Gemini
Hindi gpt
Punjabi Claude
Punjabi Gemini
Punjabi gpt
Entropy difference (bits)
3
2
1
0
−1
−2
−3


Super Category
Arts, Performance & Media
Clothing & Accessories
Commercial Products
Consumer Electronics
Culinary Practices
Cultural Events
Education & Learning
Food & Drink Items
Geography & Natural Places
Household Objects & Tools
Industrial Equipment
Institutions & Governance
Language, Narrative & Symbolic Knowledge
Leisure, Play & Sport
Life Course Milestones
Markets & Commerce
Materials & Resources
Measurement & Quantification
Money & Currency Systems
Mythical Characters
Nature & Living Beings
Personal Names
Places, Spaces & Built Environment
Religious Practices & Places
Social & Occupational Roles
Transportation
animal_feed
governing_bodies
landmark
restaurant
currency_symbol
measurement_unit
social_group
person
Mean Jensen-Shannon Distance
Category Weight (% of all entities)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5