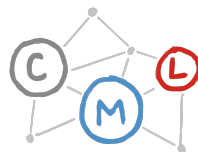


Mapping Topics in 100,000 Real-Life Moral Dilemmas

Josh Nguyen, Georgiana Lyall, Alasdair Tran, Minjeong Shin, Nick Carroll, Colin Klein and Lexing Xie

Computational Media Lab, ANU

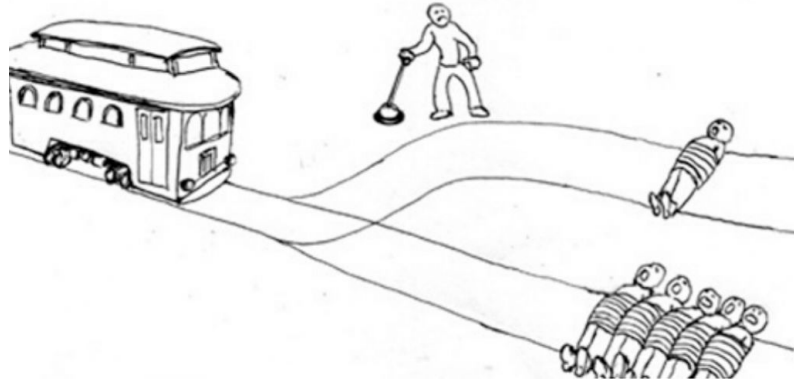
ICWSM 2022, Atlanta, GA, USA



Australian
National
University



Moral decision making: theory and practice



Stark, idealised, dramatic, aimed to test and contrast theories.

Refusing to pay my daughter's tuition.

Unfriending someone who has mental illness.

Quitting my job while we are understaffed.

Hiding my marriage from my family.

Small-scale, low-stakes, full of nuances.

Automating moral decision making

Approach	Work
Linguistic and behavioural features of moral judgement	[Zhou <i>et al.</i> , NAACL'21] [Haworth <i>et al.</i> , ICWSM'21]
Training deep networks to classify moral valence	[Jiang <i>et al.</i> , arXiv'21] [Botzer <i>et al.</i> , IEEE Trans. Comput. Soc. Syst.'22]
Compiling moral rules of thumb	[Forbes <i>et al.</i> , EMNLP'20]

Delphi speculates:
“Ignoring my boss’s phone call during my working hours.”
- ***It's wrong***

v1.0.4

Delphi prototype – Allen AI (Jiang et al. 2021)

- Existing work focuses on automating moral decision making.
- But what are the *common topics/issues* in daily moral dilemmas?

Reddit's r/AmltheAsshole (AITA)

AITA for requesting prayer breaks at work because my religious coworkers do even though I'm not religious?

I'm an atheist who works in an international company in a very busy city.

I have a lot of coworkers who come from a certain religious background that requires them to pray multiple times a day. I would say 90% of my coworkers share this faith and maybe 70-80% of them take the time out to go pray everyday.

1. Author makes a post.

Reddit's r/AmltheAsshole (AITA)

AITA for requesting prayer breaks at work because my religious coworkers do even though I'm not religious?

I'm an atheist who works in an international company in a very busy city.

I have a lot of coworkers who come from a certain religious background that requires them to pray multiple times a day. I would say 90% of my coworkers share this faith and maybe 70-80% of them take the time out to go pray everyday.

1. Author makes a post.

2. Other users comment to judge.



"Not the a-hole"

NTA

I think meditation in this situation is the non-religious equivalent to praying, a moment with yourself to collect your thoughts. It's healthy for a productive work environment and you should be given the same rights as your coworkers.



"You're the a-hole"

YTA.

I'm an atheist as well but to each their own. Who cares if they get an extra break. I view this as the same thing as smokers taking smoke breaks, do you want prayer and smoke breaks? If a nursing mother gets an extra break to pump do you also want a nursing break?

Do your thing let them do theirs. Stop being petty.

Reddit's r/AmltheAsshole (AITA)

AITA for requesting prayer breaks at work because my religious coworkers do even though I'm not religious?

I'm an atheist who works in an international company in a very busy city.

I have a lot of coworkers who come from a certain religious background that requires them to pray multiple times a day. I would say 90% of my coworkers share this faith and maybe 70-80% of them take the time out to go pray everyday.

1. Author makes a post.
2. Other users comment to judge.
3. **Community upvotes/downvotes comments.**



NTA I think meditation in this situation is the non-religious equivalent to praying, a moment with yourself to collect your thoughts. It's healthy for a productive work environment and you should be given the same rights as your coworkers.

↑ 16.3k ↓



YTA. I'm an atheist as well but to each their own. Who cares if they get an extra break. I view this as the same thing as smokers taking smoke breaks, do you want prayer and smoke breaks? If a nursing mother gets an extra break to pump do you also want a nursing break?

Do your thing let them do theirs. Stop being petty.

↑ 2.0k ↓

Reddit's r/AmltheAsshole (AITA)

AITA for requesting prayer breaks at work because my religious coworkers do even though I'm not religious?

Not the A-hole

← **Final verdict = NTA**

I'm an atheist who works in an international company in a very busy city.

I have a lot of coworkers who come from a certain religious background that requires them to pray multiple times a day. I would say 90% of my coworkers share this faith and maybe 70-80% of them take the time out to go pray everyday.

1. Author makes a post.
2. Other users comment to judge.
3. Community upvotes/downvotes comments.
4. **After 18 hours, highest-scored comment becomes the final verdict.**



NTA I think meditation in this situation is the non-religious equivalent to praying, a moment with yourself to collect your thoughts. It's healthy for a productive work environment and you should be given the same rights as your coworkers.



16.3k



YTA. I'm an atheist as well but to each their own. Who cares if they get an extra break. I view this as the same thing as smokers taking smoke breaks, do you want prayer and smoke breaks? If a nursing mother gets an extra break to pump do you also want a nursing break?

Do your thing let them do theirs. Stop being petty.

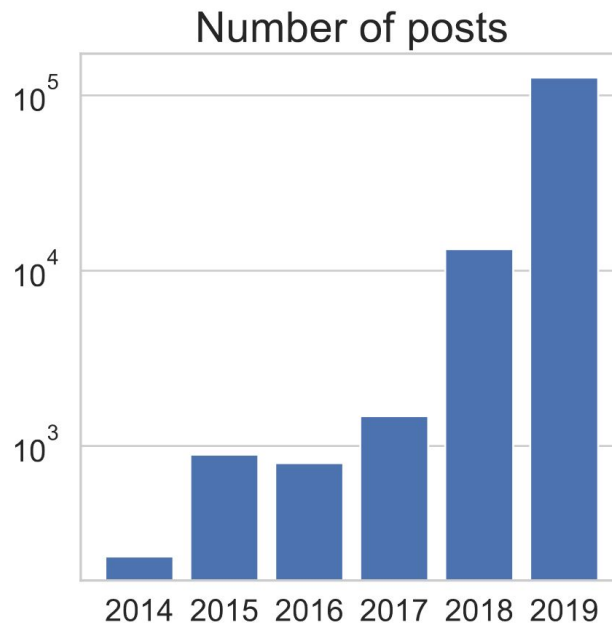


2.0k

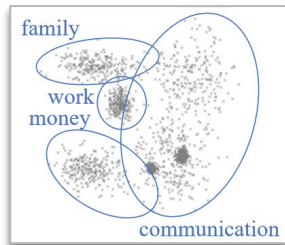


AITA: the dataset

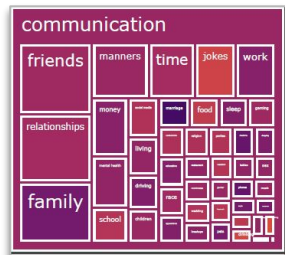
- 102K posts and 8M comments from '14 to '19.
- Subreddit grew fast over the years.
- Most common verdict is “Not The A-hole”.
- “You’re The A-hole” posts attract more comments.



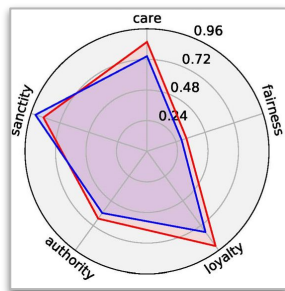
Outline



1. How to discover the topics of discussion on AITA?



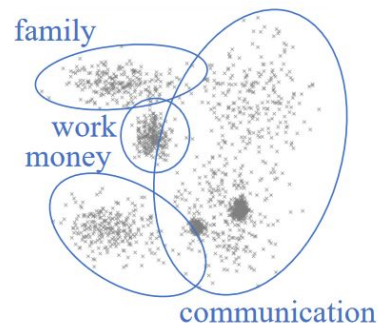
2. How do humans perceive topics on AITA?



3. What characterises the difference among topics?

Discovering ALTA topics using text clustering

- Motivation: characterising the common domains in online moral dilemmas.
- Method: latent Dirichlet allocation and expert validation.
- Result: **102K posts** → **47 topics**/dimensions of moral discussions.



Topic name	Size	Top-10 keywords
appearance	1,664	wear, hair, dress, look, like, shirt, clothe, shoe, makeup, color
-	137	trash, tattoo, review, bug, garbage, bin, product, design, star, throw
babies	893	baby, child, pregnant, kid, month, birth, pregnancy, week, husband, time
breakups	317	ex, break, partner, year, relationship, month, cheat, date, contact, new
celebrations	2,484	birthday, gift, buy, like, year, present, thing, christma, day, card
children	1,257	kid, wife, son, child, old, year, young, boy, parent, nephew
communication	7,855	talk, like, try, thing, time, start, upset, way, apologize, come
-	1,448	text, message, send, texte, day, respond, reply, week, time, talk
-	448	speak, language, english, customer, sorry, coffee, mistake, mobile, write, country
-	52	jane, behavior, session, john, anna, discuss, kate, conversation, issue, therapist

4 topics named *communication*

Validating expert-crafted named topics: a crowd-sourced study

What topic below best describes the theme of the following post? Do not let your ethical judgement of the author affect your choice here.

WIBTA if I sprinkle in just a few lies into my grad school applications?

I want to apply for a PhD program in Aerospace Engineering at the top names. But [I have] only 1 summer of research. [...] I was thinking of adding some more "fake" research experience by going through the aerospace engineering department at my school and listing some of the profs researches and saying I helped them during a certain timeframe. [...] I don't think they will contact every professor, since grad schools get lots of applications. [...]

- ☐ School
- ☐ Mental health
- ☐ Money
- ☐ Education
- ☐ None of the above

Prompt

Randomly chosen post

Top 4 LDA topics, in random order. Can choose more than 1.

Validating expert-crafted named topics: a crowd-sourced study

What topic below best describes the theme of the following post? Do not let your ethical judgement of the author affect your choice here.

WIBTA if I sprinkle in just a few lies into my grad school applications?

I want to apply for a PhD program in Aerospace Engineering at the top names. But [I have] only 1 summer of research. [...] I was thinking of adding some more "fake" research experience by going through the aerospace engineering department at my school and listing some of the profs researches and saying I helped them during a certain timeframe. [...] I don't think they will contact every professor, since grad schools get lots of applications. [...]

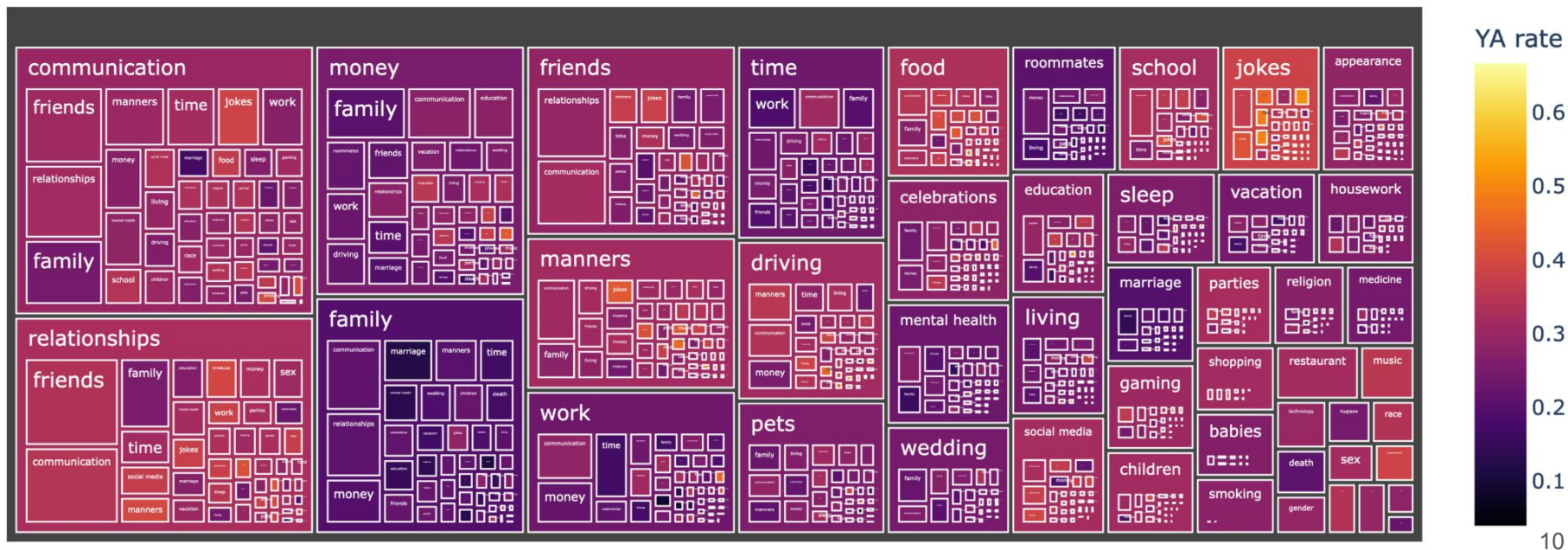
- ☒ **School**
- ☐ Mental health
- ☐ Money
- ☒ **Education**
- ☐ None of the above

- Recruited **285** participants.
- Agreement rate of **83.2%** between humans and topic model.
- On average, **1.70** topics chosen for each post.

This leads **topic pairs** to describe AITA content.

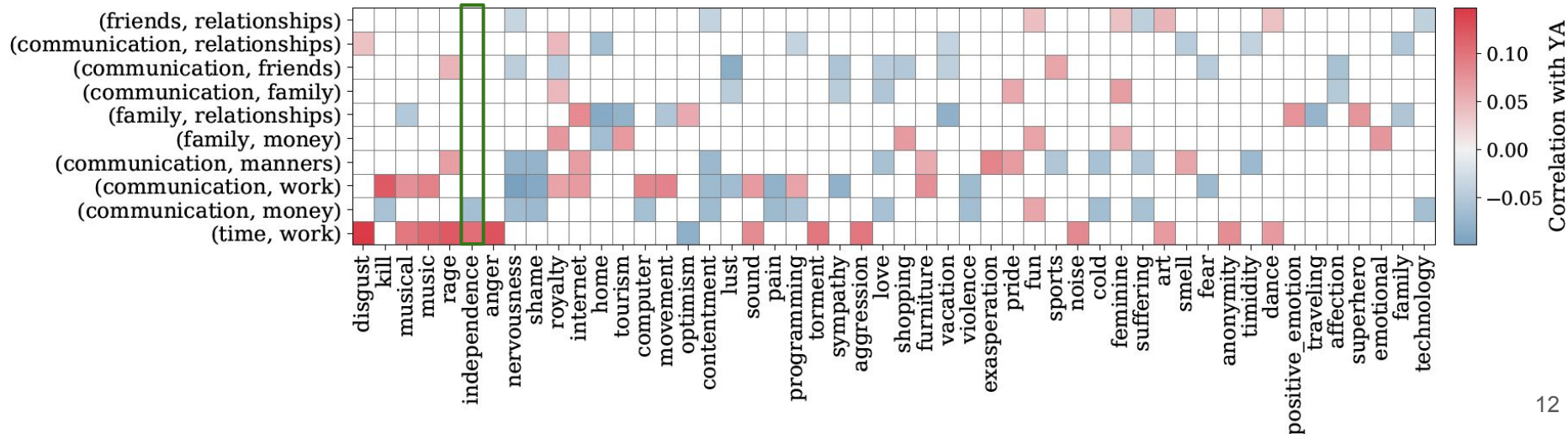
Patterns of judgement in topic (pairs)

- YA = negative judgement against author
- Block size: number of posts
- Block shade: fraction of negatively judged posts



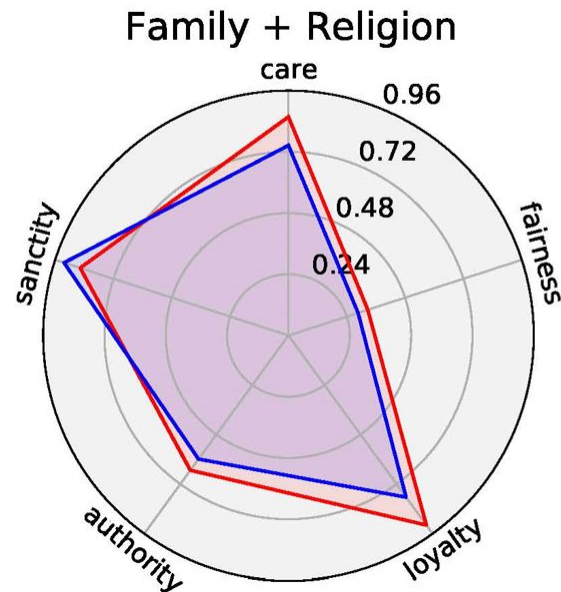
Linguistic patterns in topic (pairs): lexical categories

- **Empath**: 194 linguistic categories, similar to LIWC.
- Correlation between each category and negative judgements.
 - Rows = Empath categories.
 - Columns = 10 most common topic pairs.
- Some same categories correlate with judgement in **opposite ways** among topic pairs.



Linguistic patterns in topic (pairs): moral categories

- Moral Foundations Theory: 5 main “foundations” of morality — *care*, *fairness*, *loyalty*, *authority* and *sanctity*.
- Prevalence of each foundation = proportion of posts/verdicts containing that foundation.



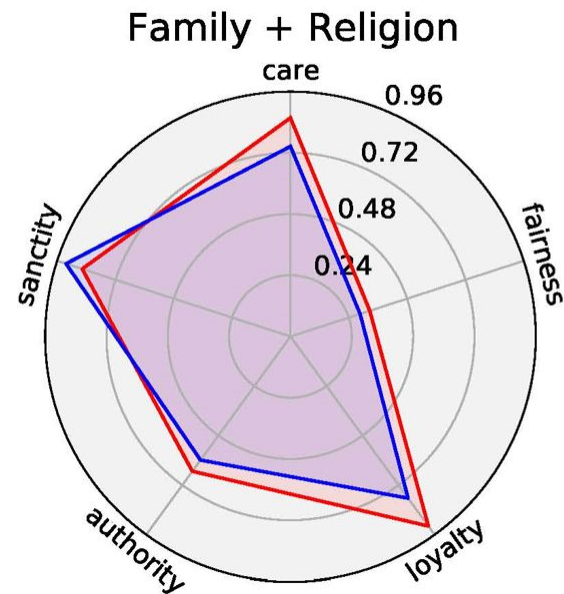
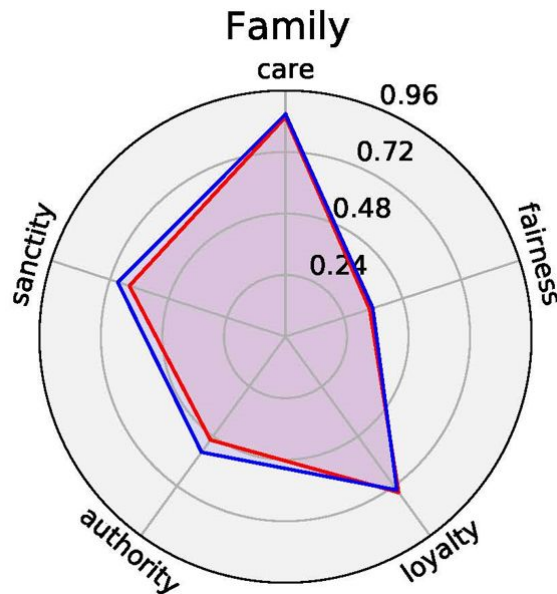
Red = YA-judged posts

Blue = NA-judged posts

Linguistic patterns in topic (pairs): moral categories

Finding 1: Topic pairs modulate the relative strengths of each moral foundation.

E.g., with secondary topic *religion*, more focus on foundations *loyalty* and *authority*.



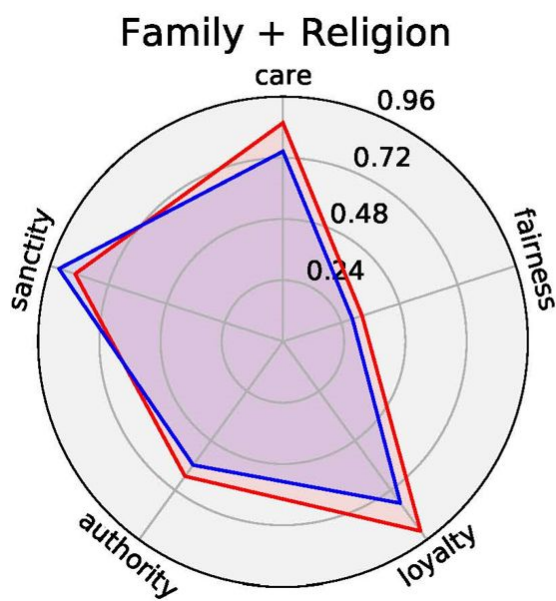
Red = negatively judged

Blue = positively judged

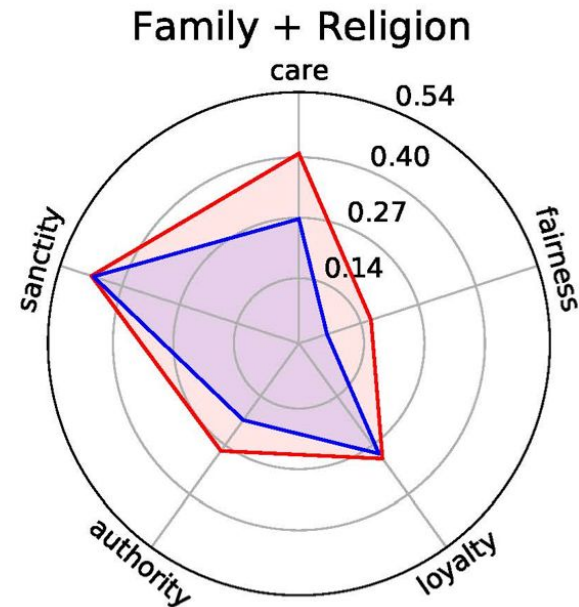
Linguistic patterns in topic (pairs): moral categories

Finding 2: Posts and their verdicts adhere to moral foundations differently.

E.g., verdicts downplay *loyalty* and focus much more on *sanctity*.



Posts

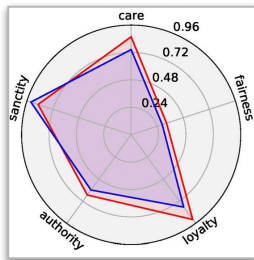
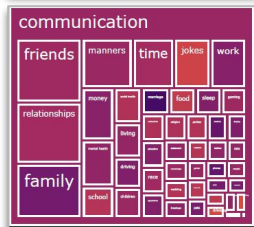
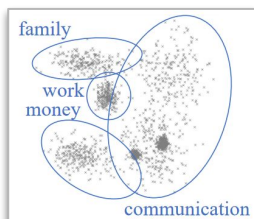


Verdicts

Red = negative

Blue = positive

Summary



1. How to discover the topics of discussion on AITA?

- Topic modeling and expert validation, resulting in 47 named topics
- Lots of topics fall into traditionally non-moral domains

2. How do humans perceive topics on AITA?

- Topic *pairs* describe a post's theme

3. What characterizes the difference among topics?

- The effect of linguistic and moral categories vary among topic pairs

Future work: moral alignment between posts and verdict, framing effects, etc.

Dataset, code and demo can be found at: https://github.com/joshnguyen99/moral_dilemma_topics