

The Disparate Effects of Recommending to Strategic Users



Andreas Haupt (haupt@mit.edu), Chara Podimata (podimata@mit.edu), Dylan Hadfield-Menell (dhm@csail.mit.edu)

Q1: Are users aware of feedback loop? Do they act in response?

large-scale survey

(MTurk)

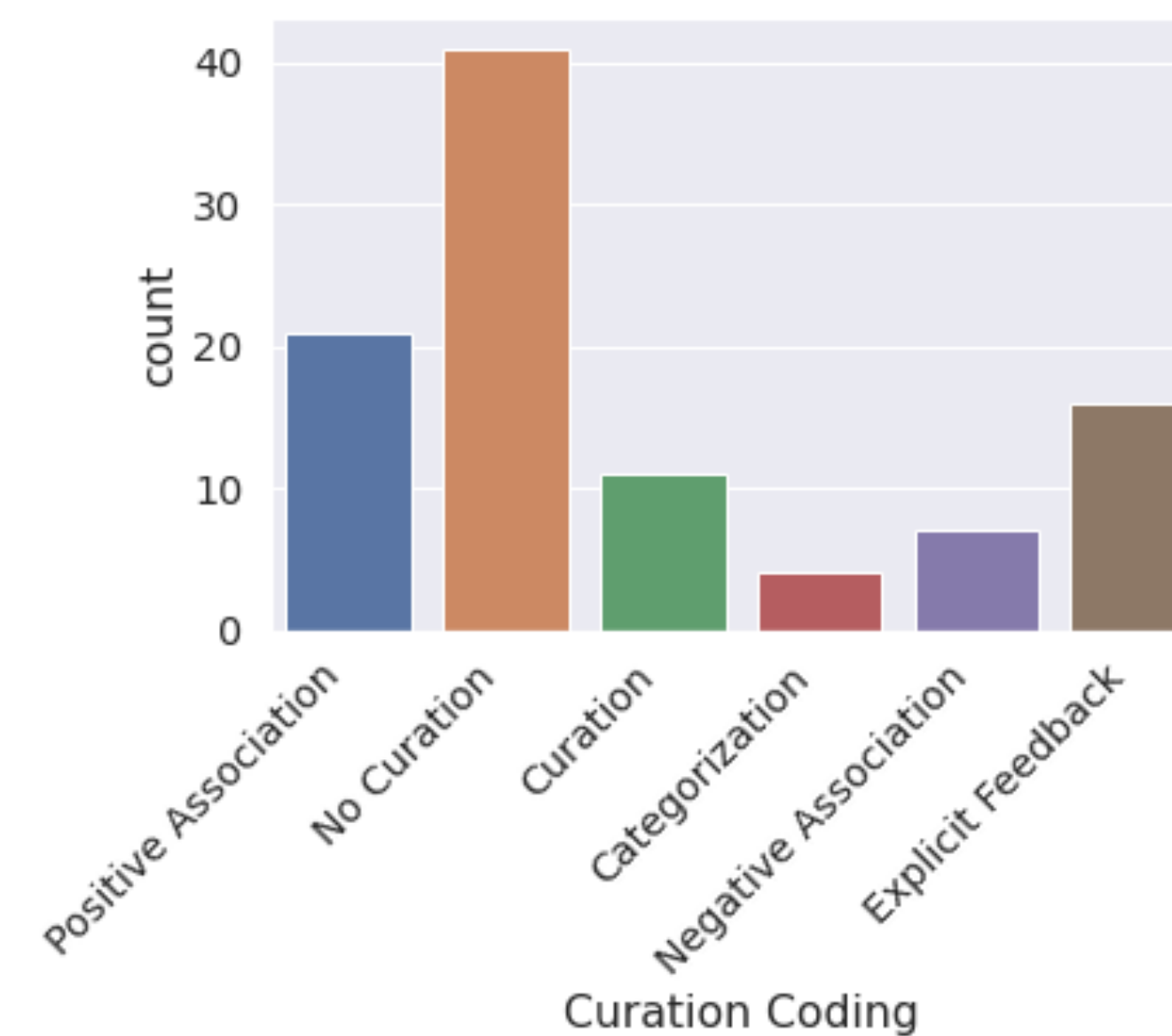
users are **non-myopic**

Why do you think TikTok recommends these categories?

Actions you take to curate your feed?

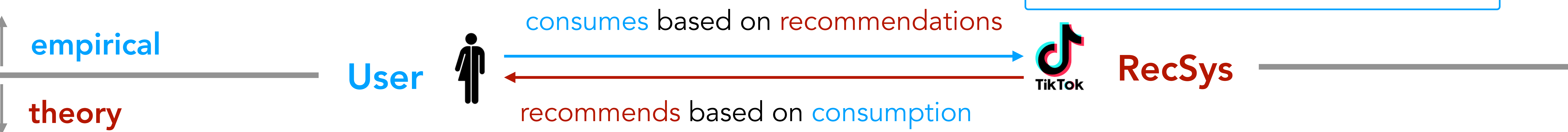
"I feel that TikTok continues to put these in my feed because I almost always get sucked into watching them. **That tells the algorithm that I like them**, even though I am mostly just using them for background noise and have seen most of them before"

"I think because I liked a video once of this type of content. I believe by me liking the video the algorithm thought I would like to see more videos like that one."



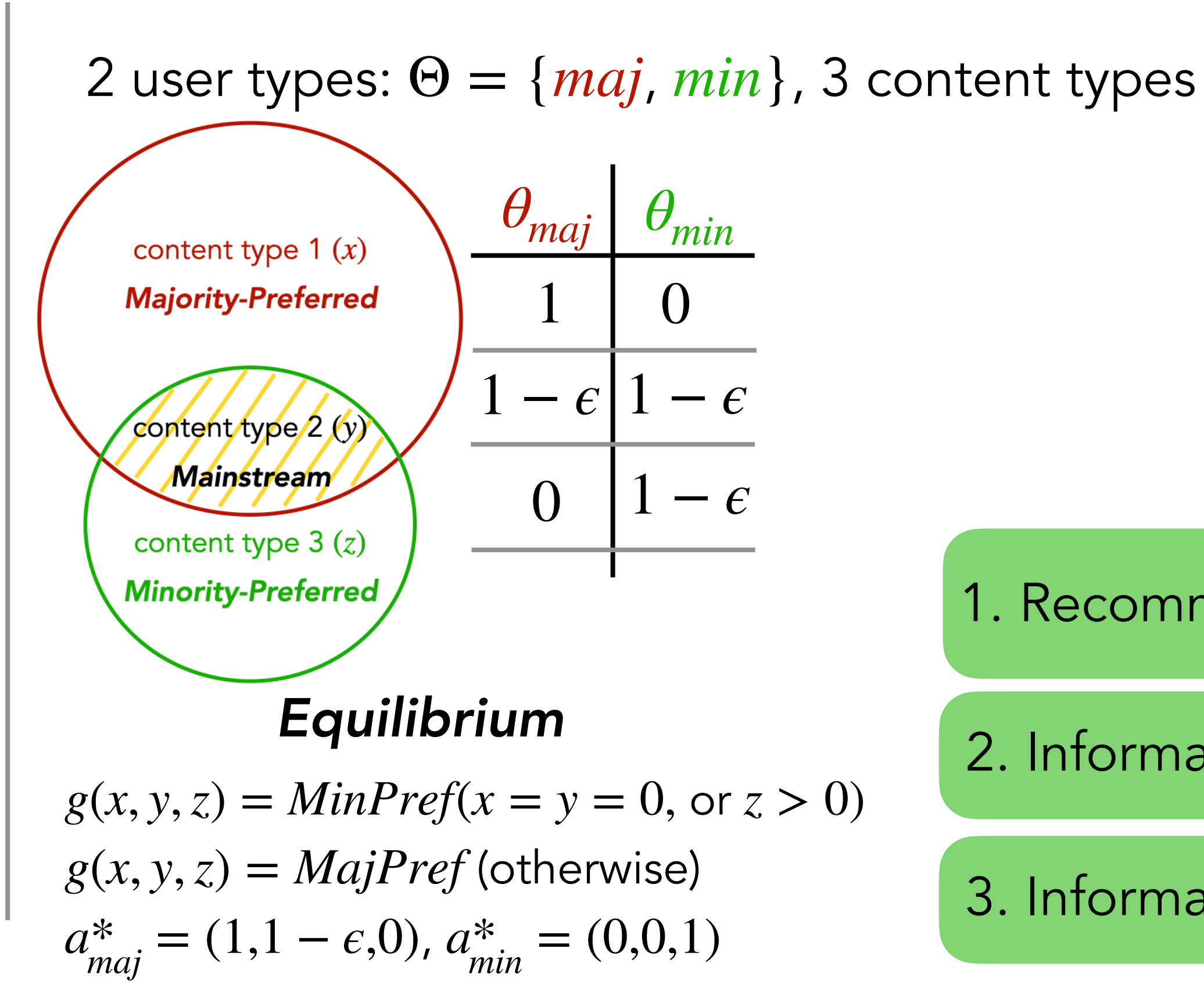
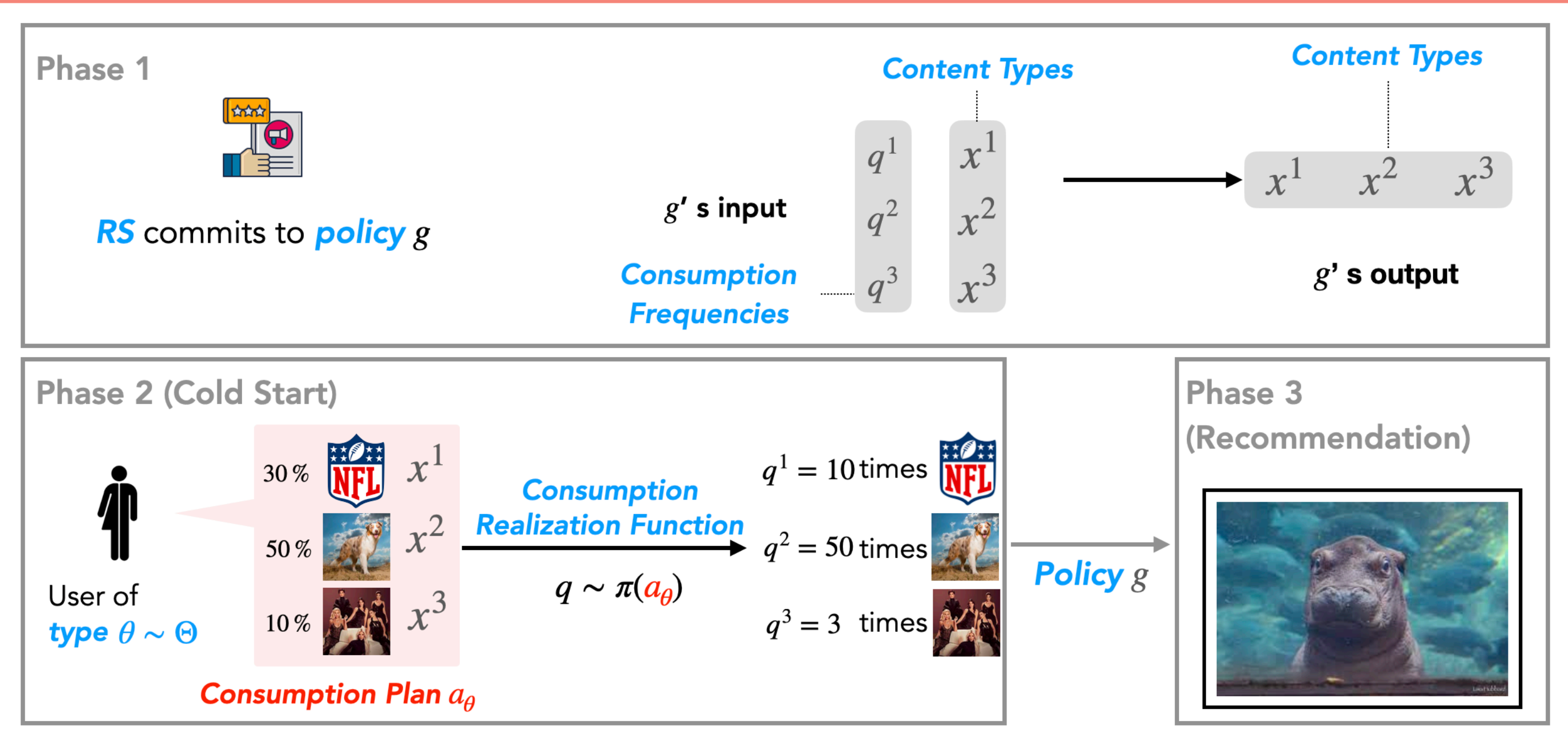
"I also like stuff just to see more of that type stuff evn though I don't like it. Llke soemtimes if my content gets to dark I try to like animal videos and **comedy more to get off the darker content for a bit.**" [sic]

"I am **cognizant of what category** of video I think material falls under. I am careful to watch completely videos that fall under the correct category (**even if I am not interested in that particular video**). I am careful to **skip** over videos from "wrong" categories."



Q2: Harms to users if RecSys does not adapt? Interventions?

Disparate Impact



- Implications for Minority**
- 1) *maj* users: **no need to strategize!**
 - 2) *min* users: act more **stereotypically**
 - 3) *min* users: **mainstream abstention**

Interventions

1. Recommendation intervention: **over-representing minorities.**
2. Information design intervention: **automatic incognito mode.**
3. Information gathering intervention: **Cold Start improvement.**

Users and RecSys have **aligned, separable utilities.**

$$u_{RS}(q, x, a; \theta) = u_U(q, x, a; \theta) := u^{CS}(q, a; \theta) + u^{Rec}(x; \theta)$$