

One Style Does Not Regulate All: Moderation Practices in Public and Private WhatsApp Groups



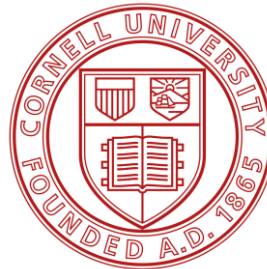
Farhana Shahid



Dhruv Agarwal



Aditya Vashistha



Motivation



WhatsApp has more
than 3 billion users

Platform with **second largest** userbase

End-to-end
encryption

Messages and calls stay between you.
No one else can read or listen to them,
not even WhatsApp.

How WhatsApp Destroyed A Village

In July, residents of a rural Indian town saw rumors of child kidnappers on WhatsApp. Then they beat five strangers to death.

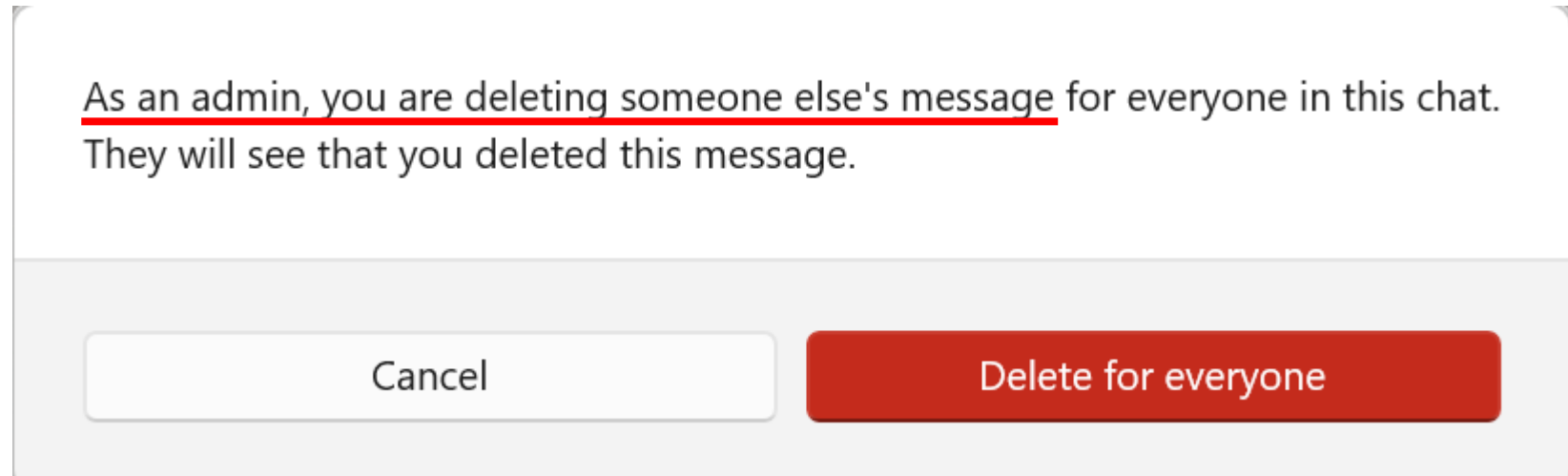
Mob Violence Against Palestinians in Israel Is Fueled by Groups on WhatsApp

Come to the ‘war cry party’: How social media helped drive mayhem in Brazil

Researchers detected a surge in aggressive rhetoric from election denialists in far-right channels online ahead of Sunday’s rioting

Research Question

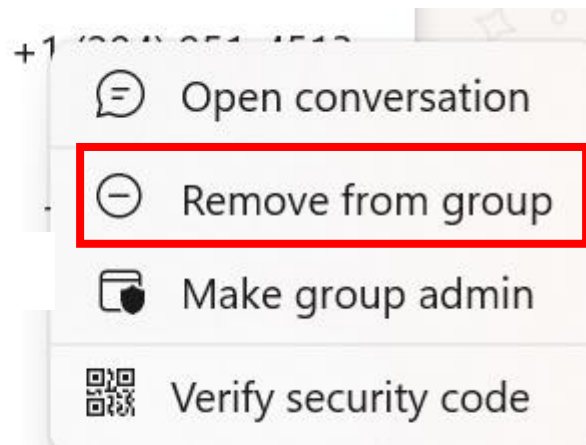
How do **admins** address harmful content when managing WhatsApp groups?



Prakash



Gautam



Methods



Semi-structured interviews with 32 **public and private** group admins in India and Bangladesh



Observation of user activities in 30 **public** WhatsApp groups



Private groups: default setting, only admin can manually add group members



Public groups: group join links are posted online for anyone to join

Theoretical Framework



Diana Baumrind

“Parenting” as a metaphor for moderation
(Gillespie et al. 2018, Register et al. 2023)

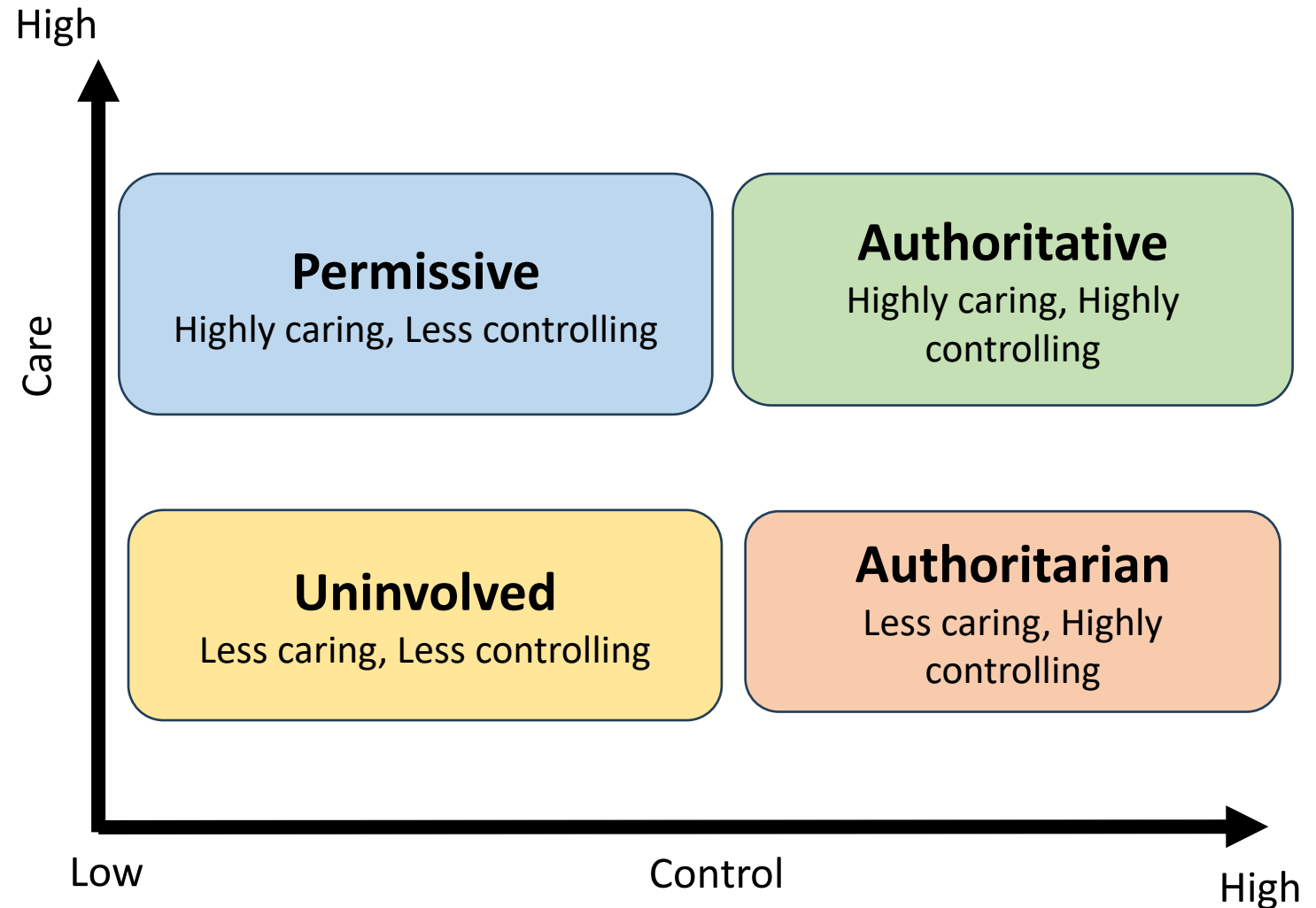
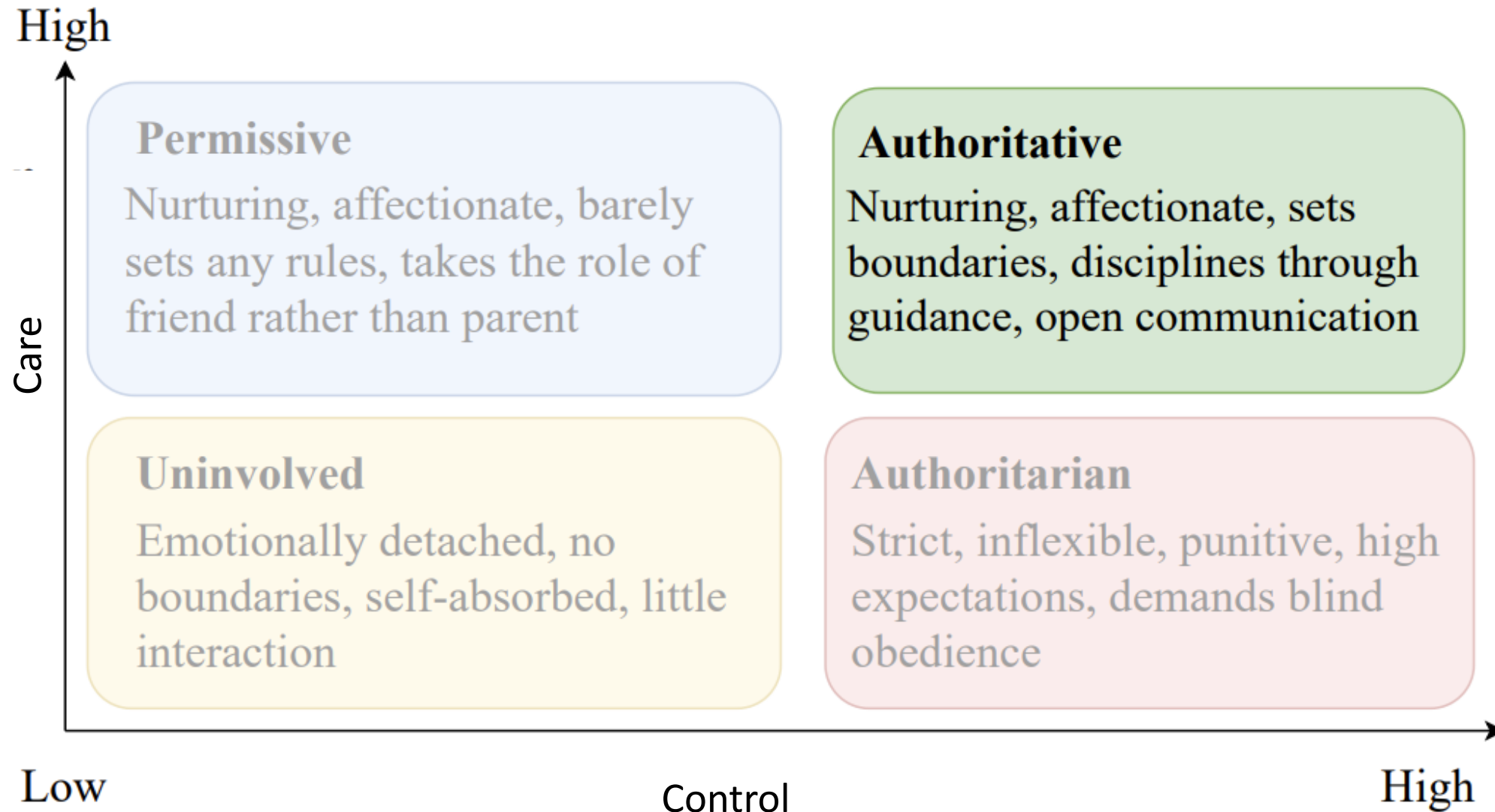


Figure: Baumrind's typology of **parenting** styles

Moderation Styles in **Private** WhatsApp Groups



Authoritative Moderation Style



Private groups with **weaker** social ties: colleagues, classmates, neighbors etc.



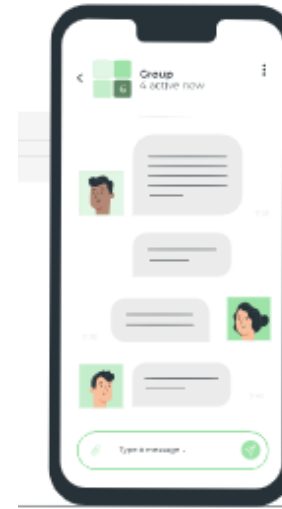
Fact-checks misinformation



Indirect, polite correction **privately** when the offender outranks the admin in social status



Sets group **rules**



Direct correction of offenders who are **peer** or **subordinate**

Authoritative Moderation Style



Deletes messages they perceive to be problematic



Removes repeat offenders from the group

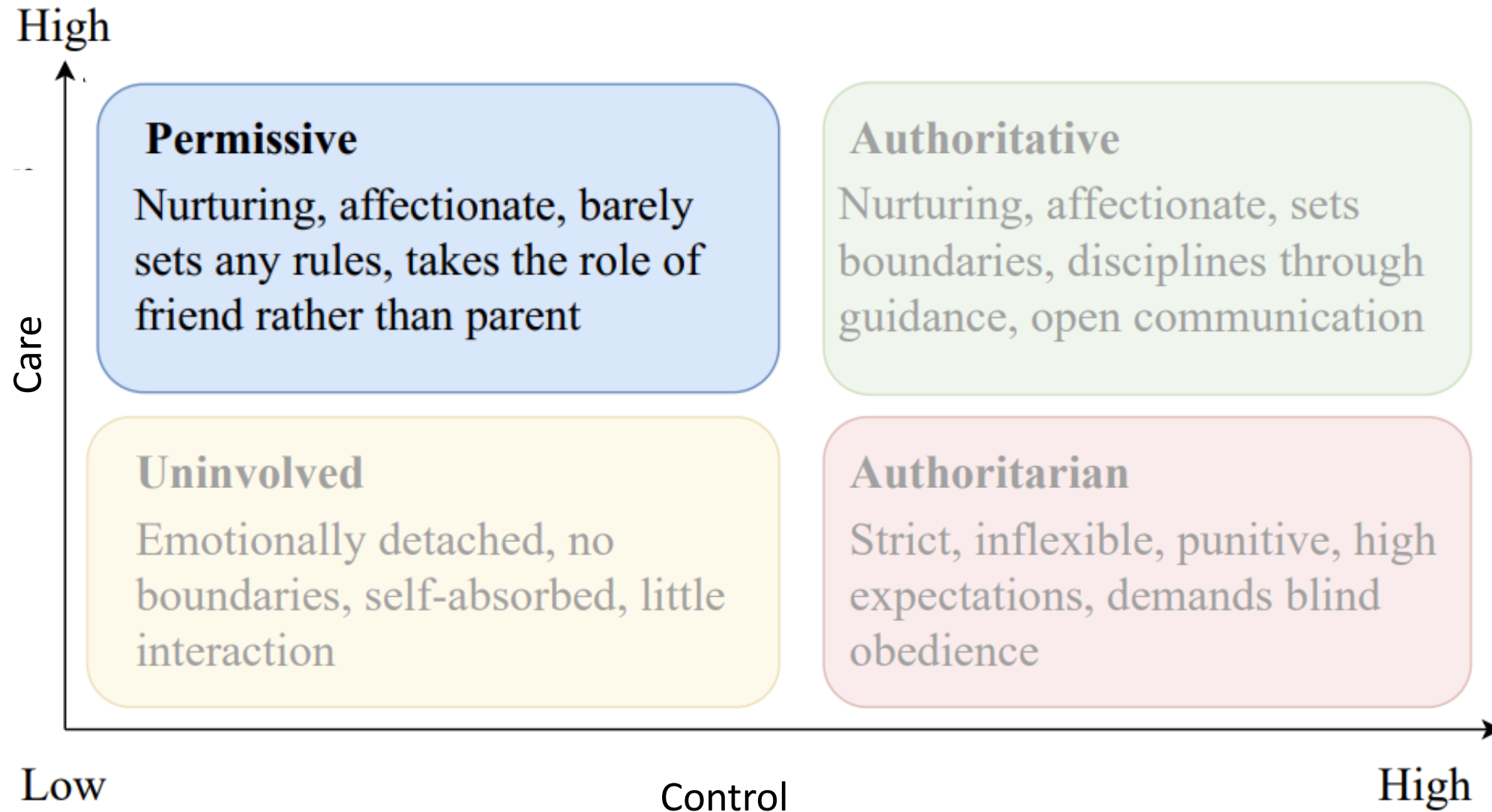


Offline mediation with repeat offenders



Critiqued by offender and group members

Moderation Styles in **Private** WhatsApp Groups



Permissive Moderation Style



Private groups with **strong** social ties: family and friends' groups

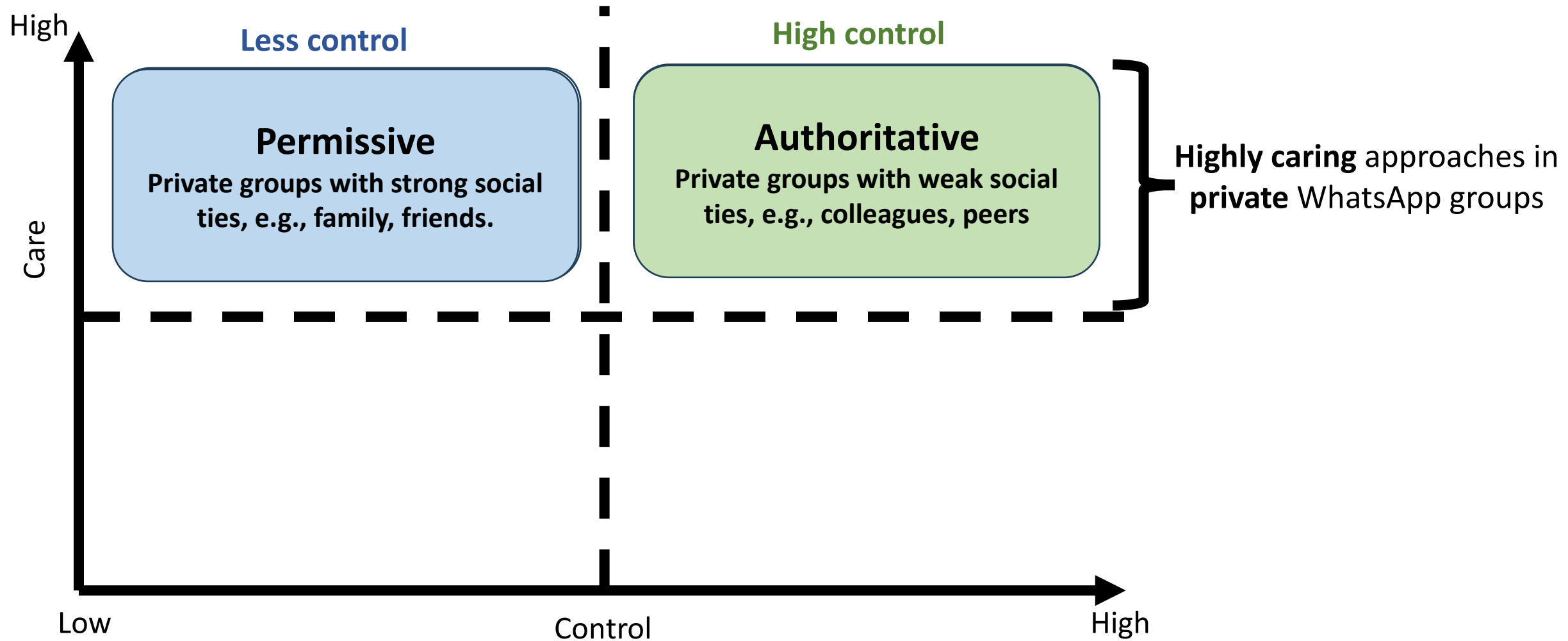


Reluctance to moderate problematic content shared by **elderly relatives**



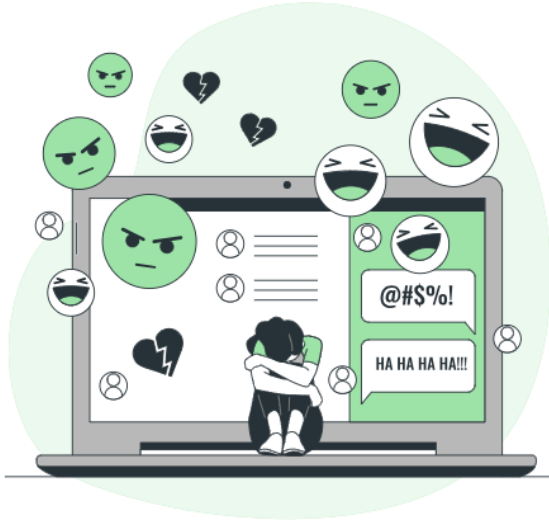
Eco-chambers of problematic content

Moderation Styles in **Private** WhatsApp Groups



Implication for **Private** WhatsApp Groups

Member and content removal do not account for the relationship dynamics in private WhatsApp groups

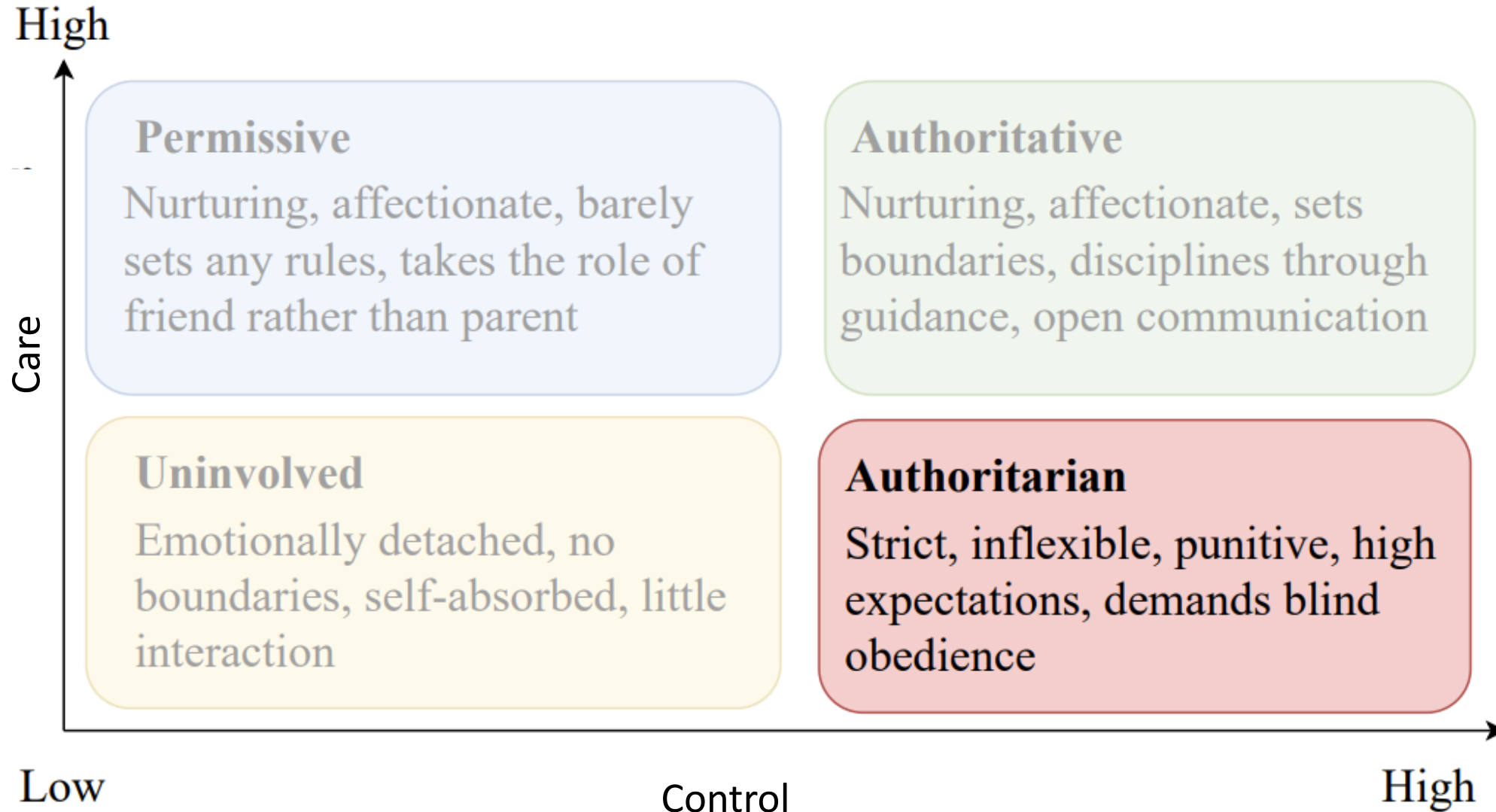


Criticism and harassment of admins who remove problematic messages and users



In close-knit groups, removing problematic messages and users is too sensitive

Moderation Styles in **Public** WhatsApp Groups



Authoritarian Moderation Style

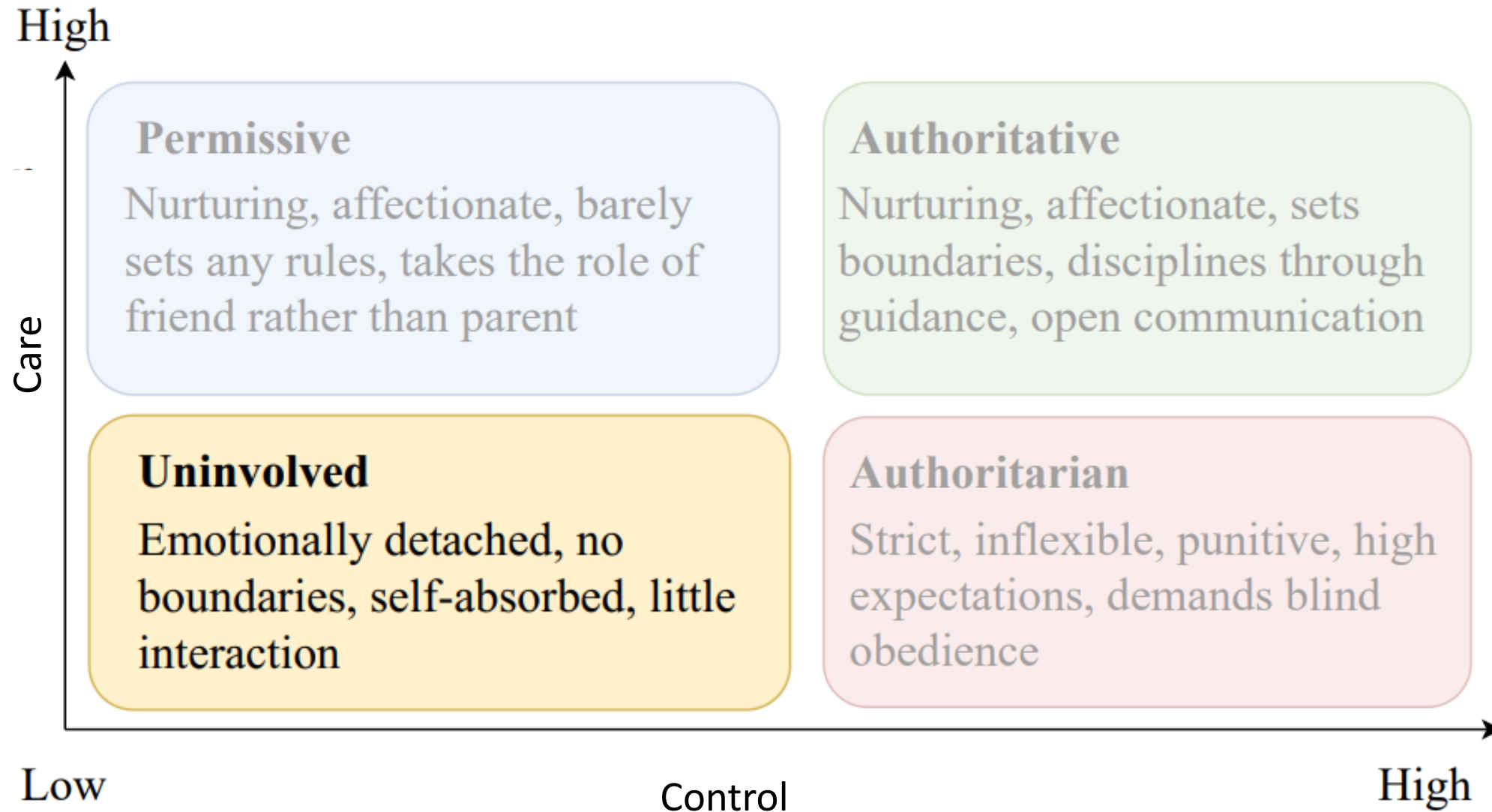


Public groups e.g., business, job opportunity groups



Collapsed interaction and communication within the group

Moderation Styles in **Public** WhatsApp Groups



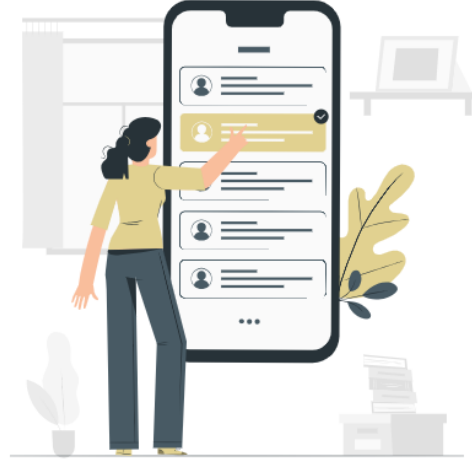
Uninvolved Moderation Style



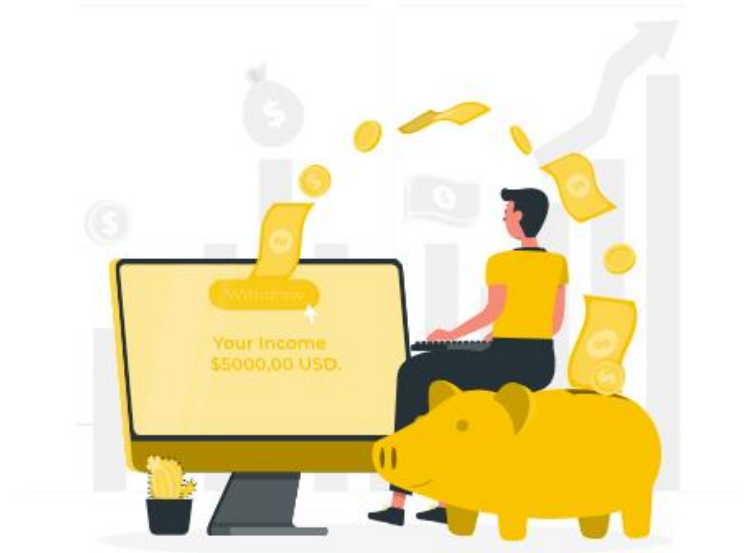
Public groups



Abandoned the group to
avoid problematic content

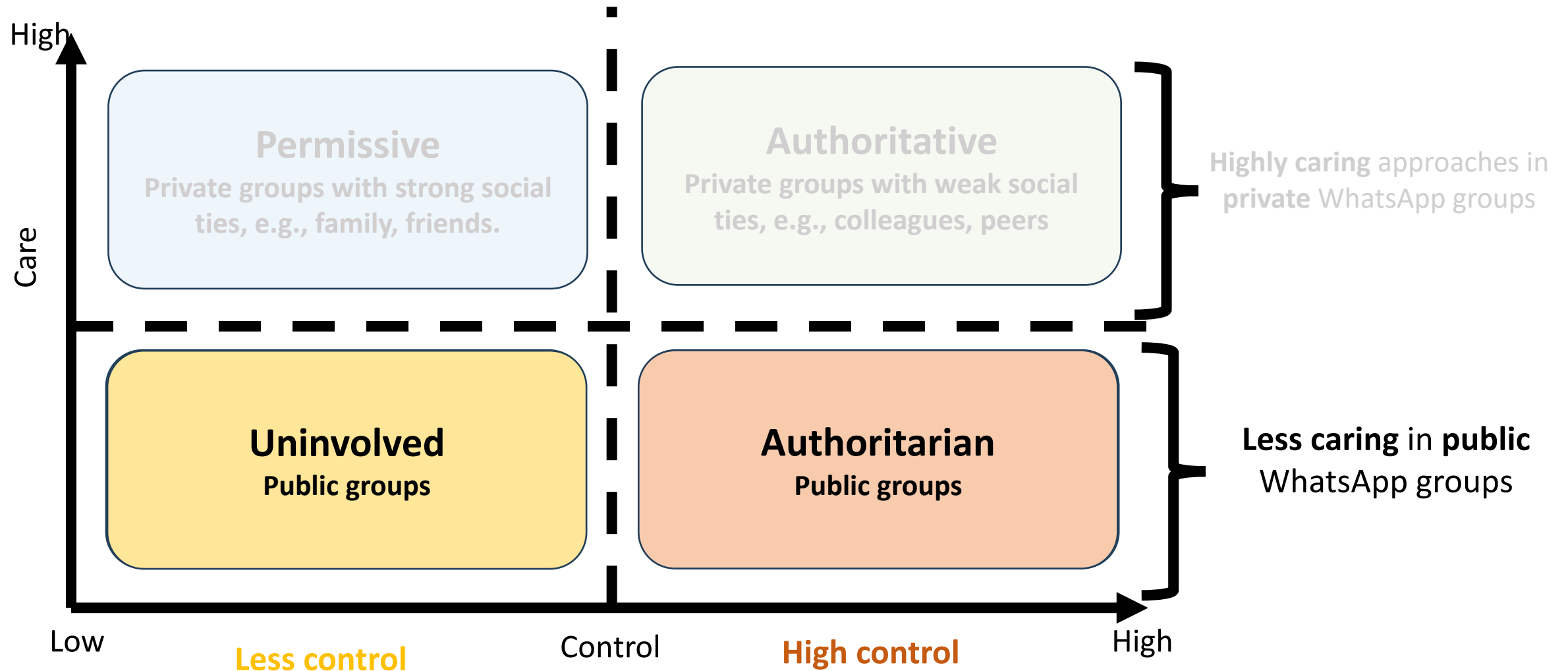


Sell the large public group
to make profit



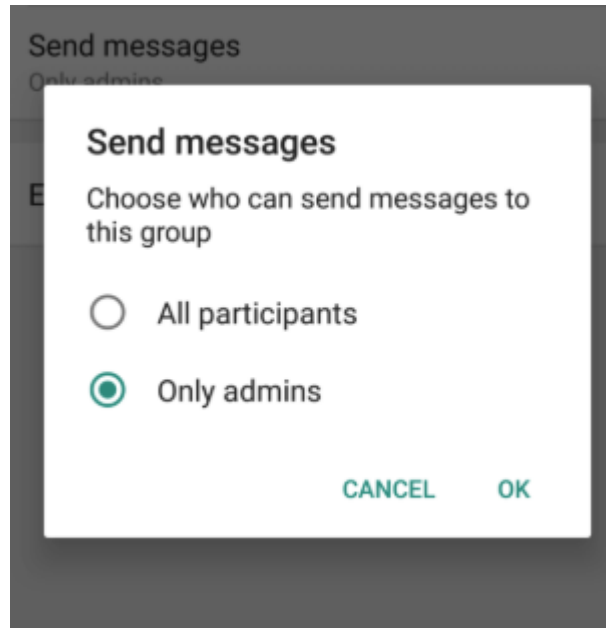
Make money by posting
scam messages

Moderation Styles in WhatsApp Groups



Implication for **Public** WhatsApp Groups

No moderation tools to help admins deal with the **overflow of problematic content** in **public** groups



Authoritarian measures



Absolutely ignoring group

Design Implication

Punitive moderation tools are inadequate

- 3Billion WhatsApp users without any effective intervention
- 853.8 Million WhatsApp users in India
- Designing culturally sensitive tools for collectivist societies
- Diverse needs across different types of groups



Full Paper

Thank You!
Farhana Shahid (fs468@cornell.edu)