



DEALING WITH HATE SPEECH & MODERATION

RECAP AND REFLECTION





REFLECTION

What do you do once you have encountered a hateful message?

HOW TO DEAL WITH HATE & DISINFO ONLINE?

1. Flag/Report
2. Delete/Hide
3. Respond/Counter
4. Ignore



AUDIENCES

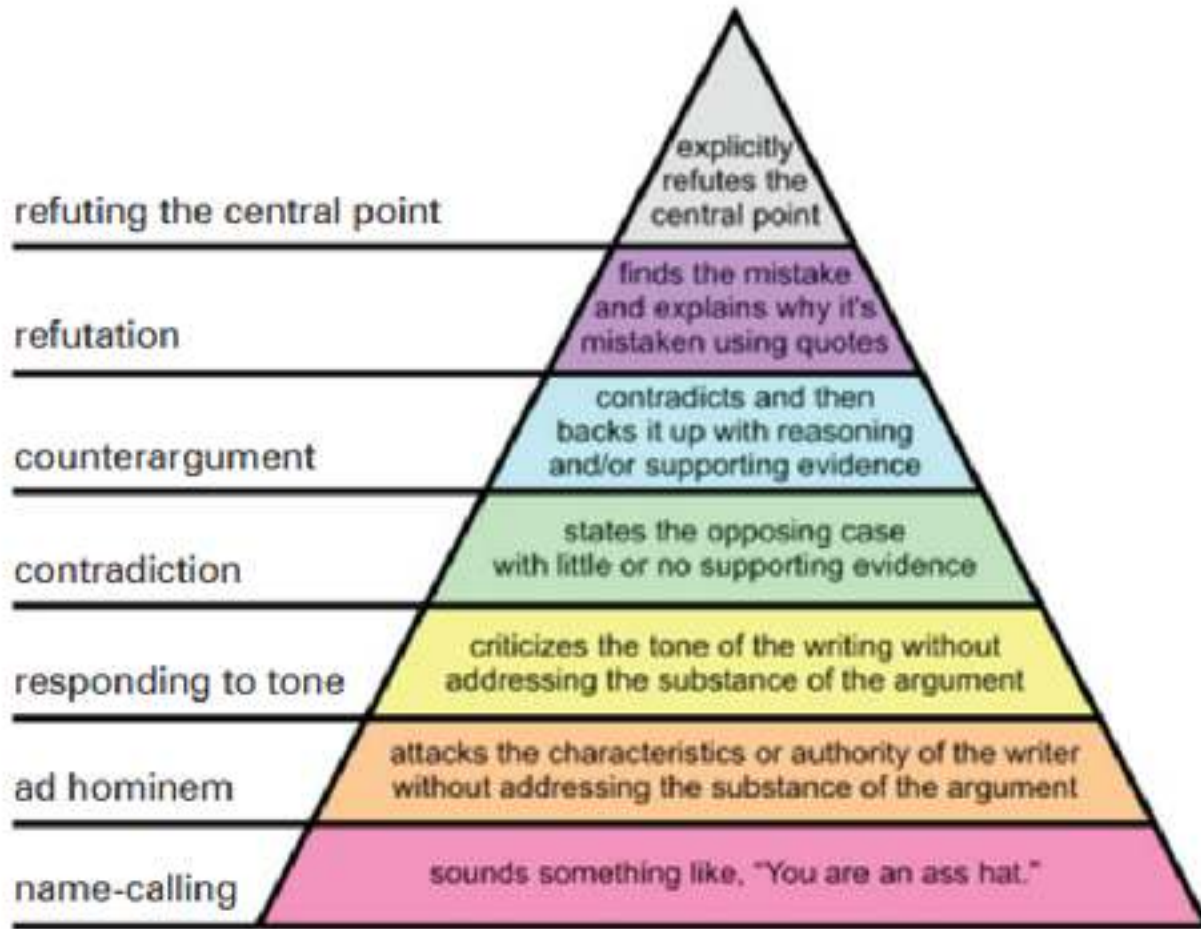


- How does the presence and type of audience dictate our responses?
- How do we recognise who we are able to influence?

MESSAGES

Who is speaking, how are they speaking and what are they saying is one of the key aspects of persuasion.

- Are you critically trying to encourage people to think about a source of claims?
- Are you looking to challenge prejudicial messages about other races or religions?
- Or are you attempting to highlight the hypocrisy of an extremist group or ideology?



User Comments

| Supportive / Constructive | Inquisitive | Negative / Unconstructive | Antagonistic | Abusive / Offensive | MODERATOR RESPONSE |
|---------------------------|-------------|---------------------------|--------------|---------------------|---|
| X | | | | | Like / React |
| X | | | | | Positive Affirmation |
| | X | | | | Answer Question |
| | X | X | | | Inform / Clarify |
| X | | X | X | | Ignore |
| | | X | X | | Challenge |
| | | X | X | | Enforce Community Guidelines (Publicly) |
| | | | X | X | Hide Comment |
| | | | X | X | Hide (with DM to user) |
| | | | | X | Hide & Block |

Moderation Guidelines

- Always answer user questions.
- Contribute in respectful, non-judgemental way – but don't be afraid to challenge users.
- Ignoring comments, i.e. not replying, is a form of moderation. Especially on social media, not replying to a comment automatically diminishes the potential reach of that comment.
- We never delete comments – and we only hide them if they are abusive, violent, or offensive.
- We focus our moderation on norm/core topics – although dealing with comments of an abusive, violent, or offensive nature are to be dealt with regardless.
- Violent, abusive, or offensive comments have no place on Citizens' Voice communities – they should be immediately hidden, and private messages sent to offenders.
- Spam should always be hidden, and repeat offenders blocked.
- There should be short and clear behaviour guidelines on the site/page, that you can direct users to.



How to deal with polarisation

1. **Change the target audience.** This could take the form of ignoring extreme, polarised positions and looking and highlighting the opinions of the middle.
2. **Change the topic.** Move away from the identity construct chosen by the pushers and start a conversation on the common concerns and interest of those in the middle.
3. **Change position:** Stop trying to build bridges (positioning above the poles) but rather to a position in the middle (connected and mediating).
4. **Change the tone.** This is not about right or wrong or facts. Use mediating speech and try to engage and connect with the diverse middle ground.

QUESTIONS





REFLECTION

**What is your key
takeaway from this
session?**



THANK YOU!



