

Responsibility for Recommendations

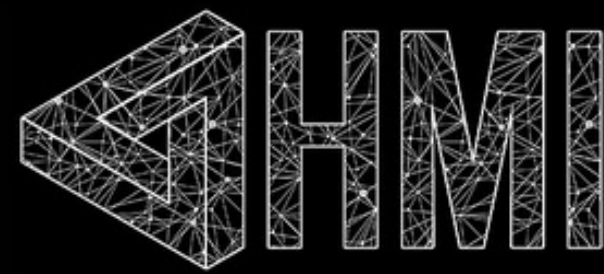
please note: this talk will contain some discussions of ethnic and gender-based violence

Michael Randall Barnes

slides available at michaelrandallbarnes.com/talks



Australian
National
University



mintLab
Machine Intelligence and Normative Theory

The New York Times

A Genocide Incited on Facebook, With Posts From Myanmar's Military

"In 2017, the Rohingya were killed, tortured, raped, and displaced in the thousands as part of the Myanmar security forces' campaign of ethnic cleansing.

In the months and years leading up to the atrocities, Facebook's algorithms were intensifying a storm of hatred against the Rohingya which contributed to real-world violence."

- Agnès Callamard, *Amnesty International's* Secretary General

NEWS

Home | Israel-Gaza war | War in Ukraine | Climate | Video | World | Asia | UK | Business | Tech

Asia | China | India

Facebook admits it was used to 'incite offline violence' in Myanmar



ENGLISH

WHO WE ARE

WHAT WE DO



MYANMAR: FACEBOOK'S SYSTEMS PROMOTED VIOLENCE AGAINST ROHINGYA; META OWES REPARATIONS



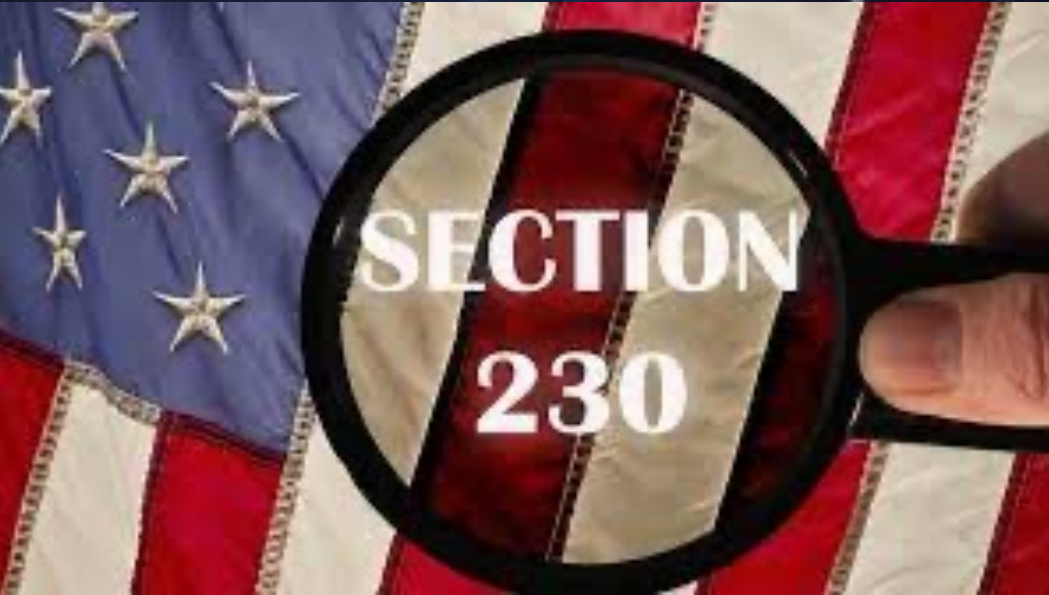
I'm Maung Sawyeddollah.

Hard Questions About (Online Intermediary) Responsibility Are Unsettled

Different Legal Regimes Taking Shape:



- While the European Union's **Digital Services Act** is steadily coming into force, applications of key provisions (e.g., its transparency and audit requirements) are yet to be fully implemented.



- In the US, the Supreme Court is set to hear two cases (***Moody v. NetChoice*** and ***NetChoice v. Paxton***) that could upend **Section 230 of the Communications Decency Act**, which offers broad intermediary liability protections for online service providers.



What's Going On
When We Post On
Social Media
Platforms?

And Who Bears
Responsibility When
Things Go Wrong?



A Naïve View:

Platform Companies (Meta, Google, X, TikTok, etc.) are simply not responsible for the content that appears on their sites/apps.

This is because—on one view—they are mere intermediaries, connecting speakers and audiences.



How To Fix Section 230

Danielle K. Citron

PUBLISHER

Boston University Law Review

DATE

2023-10

A More Nuanced and Targeted Intermediary Liability Framework



New obligation structure for online intermediaries as the DSA comes into force



The Plan

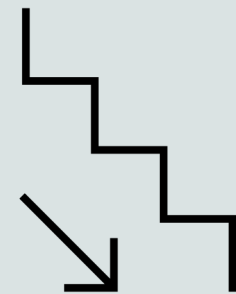
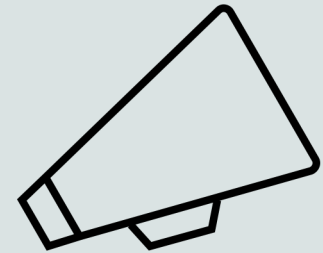
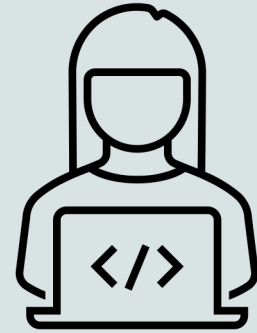
Set up ~~= Why This Matters~~

Part 1: Background

1. Clarify: My Aims
2. Clarify: Why Social Media?
3. Clarify: How Speech Act Theory Can Help Here?

Part 2: Speech Acts on Social Media

1. Algorithms
2. Affordances
3. Amplifications (Recommendations)



Wrap Things Up

Part 1.1

My aims

My claims

This talk argues **social media companies share in the responsibility** for the harms that occur on their platforms by demonstrating how they are **constitutive intermediaries**.

Platforms Bear Responsibility

I use **speech act theory** and the concept of **affordances** to illuminate the contribution that platform companies make to our communicative acts, showing how **platforms shape users' speech**, and also **perform speech acts themselves**.

I focus on **recommendations**, and show how these are **second-personal speech acts** that implicate the recommender in morally relevant ways, by **embedding values**.

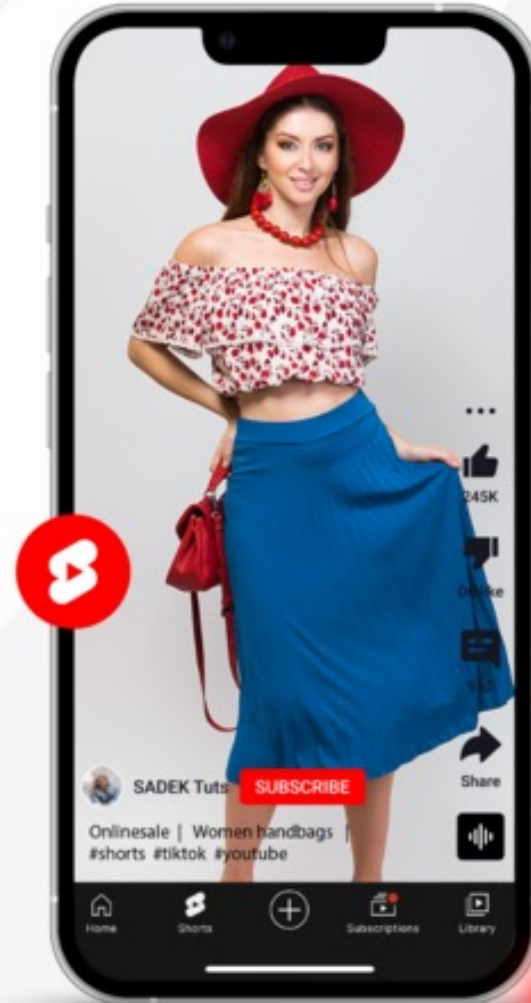
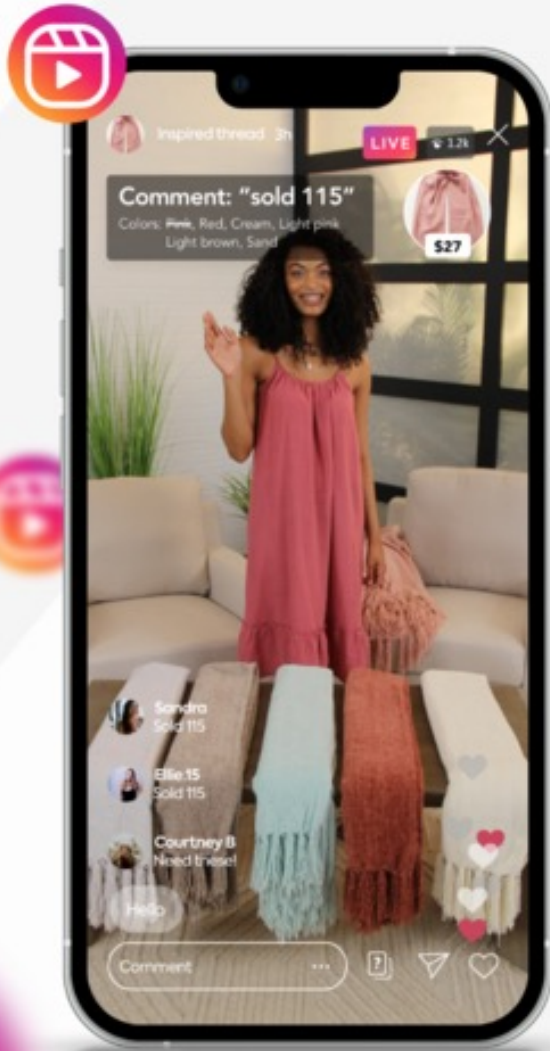


Seth Lazar: platforms exercise 'intermediary power' over their users. This is a type of governing power with significant impacts for users and society, and so requires justification (or elimination).

Jeff Howard: social media platforms have a duty to moderate content based on "natural rescue duties to defend those wronged by such speech [and] duties to avoid complicity with users' wrongful speech."

Part 1.2

What is
Social Media?



Social Media is the Media. And It's Growing



European
Commission

DSA Transparency Database



4 291 342 933

Total number of statements of reasons submitted



16

Number of active platforms



Most Reported Violations

1. Scope of platform service
2. Illegal or harmful speech
3. Unsafe and/or illegal products



73%

of fully automated decisions



Top Restriction Types

1. Disabling access to content
2. Removal of content
3. Other restriction (please specify)

Part 1.3

How Can Speech Act Theory Help?

Speech is Action.

Through speech, we do more than transfer information. We *request, command, insult, invite, warn, inquire, propose*, etc. In doing so, we make concrete changes to the social landscape.

Speech performs these actions—these normative social transformations of obligations, permissions, etc.—only when performed in the right context, following the right conventions, and when the speaker has the right entitlements and/or authority.

Speech Act Theory



J.L Austin

The Locutionary Act

- The making of a meaningful utterance.
The act of “saying something”.

The Illocutionary Act

- An act of doing something with words (like promising or betting).
The act performed “in speaking”.

The Perlocutionary Act

- An act brought about by doing something with words (like upsetting your partner).
The act effected “by speaking”.

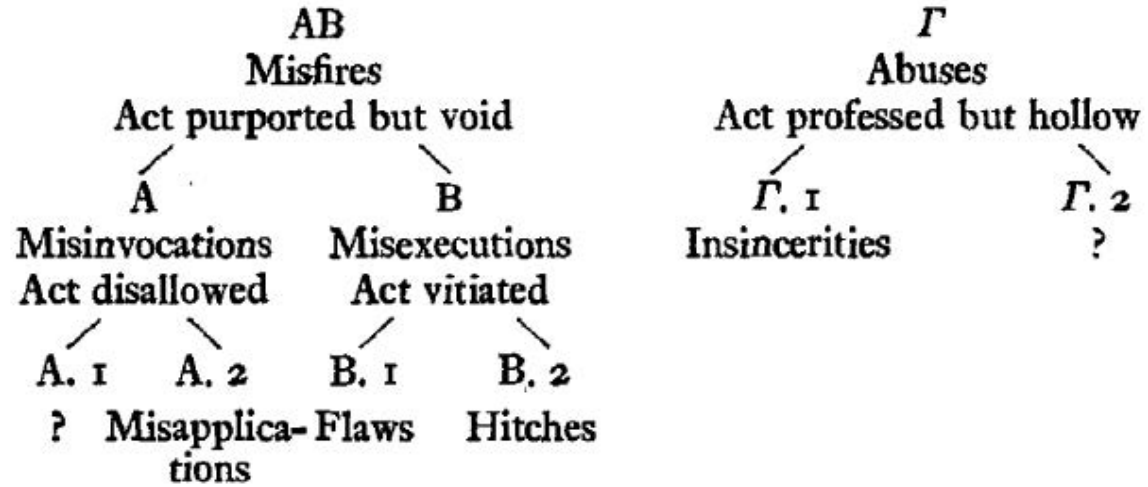
Speech Act Theory



J.L Austin

Kinds of Infelicity

Infelicities



- Austin is a **conventionalist**: what matters for the performance of a speech act is that the speaker acts according to publicly defined conventions.
- Like all conventions, felicity conditions do not have perfectly sharp boundaries of application.

They shade into one another.

(Daniel W. Harris)

Speech Act Theory



J.L Austin

Speech Act Theory

Hornsby and Langton: “some kinds of speech can set the conditions for other kinds of speech: they make some speech acts possible for some, and impossible for others”



Three Observations From Speech Act Theory

1. Illocutionary acts are acts that occur in particular **contexts**:


Illocutionary acts are **uses**; in different contexts the same locution may be put to different uses.

2. The **audience** of a speech act (as part of its context) partly determines illocutionary force:

An utterance among equals can be different from the same utterance occurring in a hierarchy.

3. The **uptake** an audience gives can play an important (and sometimes determinative) role in constituting a speech act:

Ex: refusals, jokes, bets, questions, etc.



If features like *context*, *audience*,
uptake matter for our offline speech
acts, then how do these matter
online?

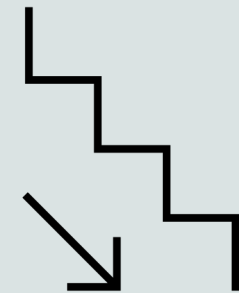
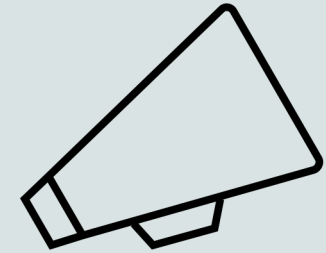
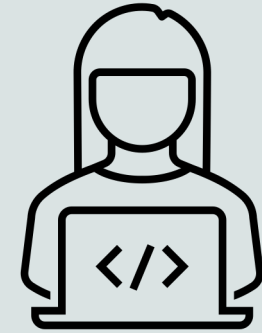
Speech Act Theory

As **algorithmic intermediaries** Social Media Platforms **control the context** of our discursive interactions and are therefore **constitutive intermediaries** of our online speech.

Through various **features** and their **affordances**, platforms **shape user actions**, including **uptake** that further shapes our speech acts.

With **amplification algorithms** – or rather, **recommender systems** – platforms **control the audience** of an utterance, inserting it into contexts of their choosing, making them **co-producers** (of a sort) of whatever speech acts that utterance constitutes.

On The Internet





The Plan

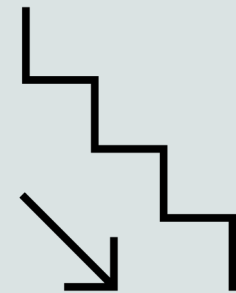
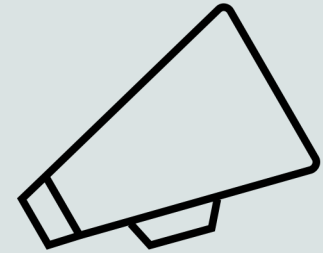
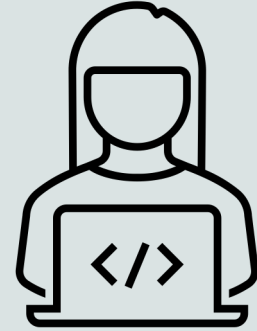
Set up ~~= Why This Matters~~

Part 1: Background

- ~~1. Clarify: My Aims~~
- ~~2. Clarify: Why Social Media?~~
- ~~3. Clarify: How Speech Act Theory Can Help Here?~~

Part 2: Speech Acts on Social Media

1. Algorithms
2. Affordances
3. Amplifications (Recommendations)



Wrap Things Up

Part 2.1

Algorithms

Algorithms

- Unlike offline, in real life (IRL) speech, Online Speech is only made possible because of the existence of **(algorithmic) intermediaries**.
- These intermediaries have an impact on how we act.



Algorithms

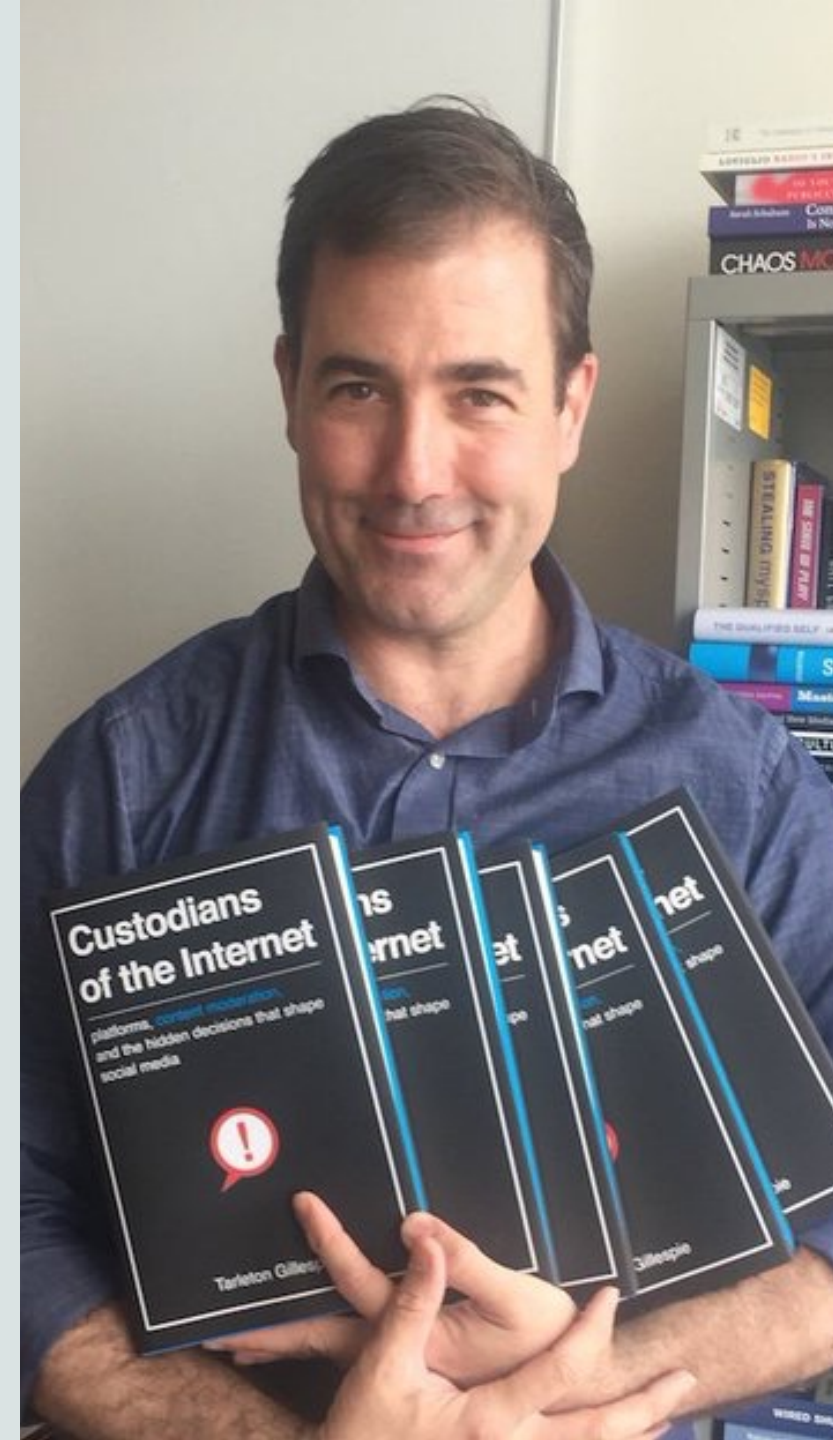
- Unlike offline, in real life (IRL) speech, Online Speech is only made possible because of the existence of (algorithmic) intermediaries.
- These intermediaries have an impact on how we act.



Algorithms impact how we act

- 1. Identity Management.**
- 2. Content Creation and Sharing.**
- 3. Interaction and Feedback.**

With these features **platforms provide the context in which our acts make sense.**



With these features **platforms** provide the context in which our acts make sense.



“context is not just the backdrop against which a person speaks; rather, the **context and the subject mutually determine each other ongoingly.**”

- Mary Louise Pratt, 1986,
“Ideology and Speech-Act Theory”

With these features **platforms** provide the context in which our acts make sense.



“context is not just the backdrop against which a person speaks; rather, the **context and the subject mutually determine each other ongoingly.**”

- Mary Louise Pratt, 1986,
“Ideology and Speech-Act Theory”

Online platforms are **Constitutive Intermediaries**

With these features **platforms provide the context in which our acts make sense.**



ChatGPT

A better term for "constitutive intermediaries" could be:

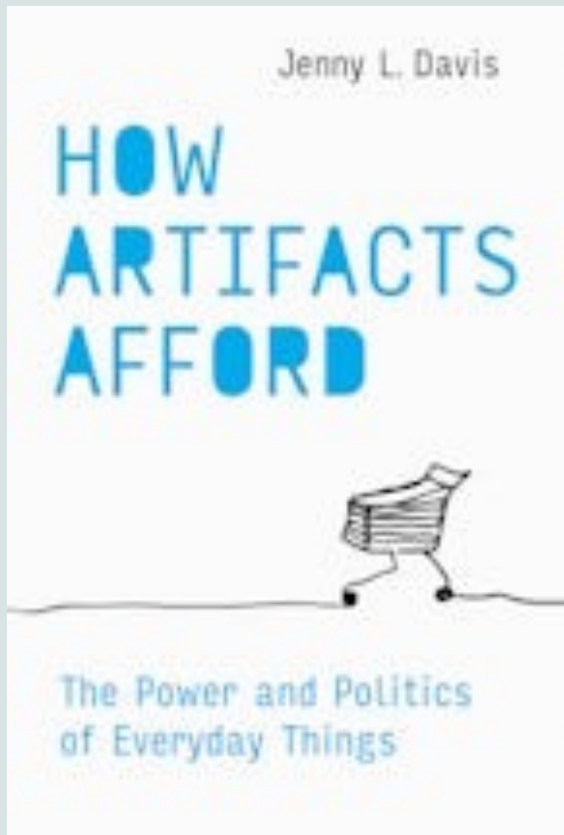
1. Foundational connectors
2. Elemental liaisons
3. Core intermediaries
4. Key linkages
5. Essential bridges
6. Primary agents
7. Fundamental mediators
8. Baseline facilitators
9. Integral intermediaries
10. Central conduits



Online platforms are **Constitutive Intermediaries**

Part 2.2

Affordances

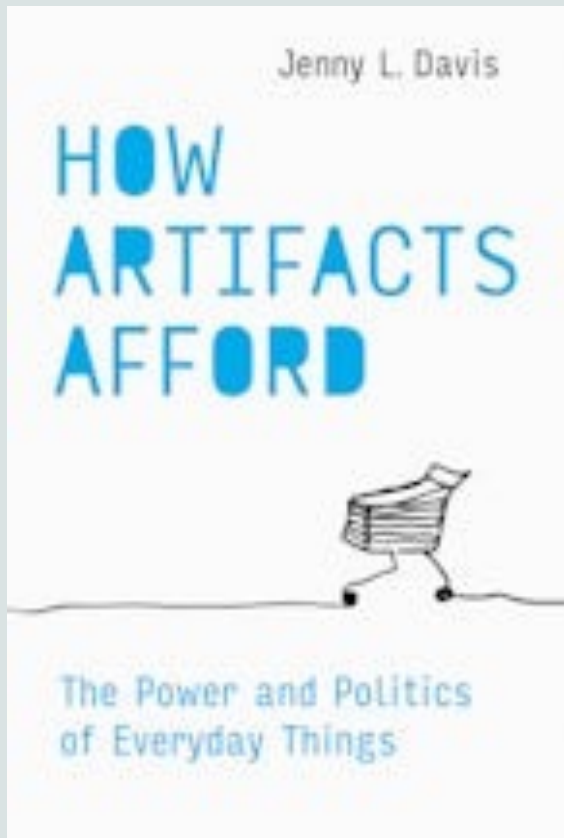


Affordances

Defined as: *the multifaceted relational structure between an object/technology and the user that enables or constrains potential behavioural outcomes in a particular context.* (Evans et al, 2017)

“Affordances mediate between a technology’s features and its outcomes. Technologies don’t make people do things but instead, push, pull, enable, and constrain. **Affordances are how objects shape action for socially situated subjects.**”

- Jenny Davis



Affordances

Definition: *the relationship between an object's properties and the capabilities of the actor that allows the actor to perform an action.* (Evans et al, 2017)

... a technology's affordances don't just exist, they are enacted. Instead, push, pull, tap, etc. Affordances are how technologies are socially situated

Three uses I want to get out of this concept:

affordances of social media shape speech acts ...

01

by encouraging
and enabling some
acts over others.

02

by encouraging
and enabling types
of uptake that
impacts the force.

03

by themselves being
2nd Personal speech
acts.

1. Affordances Shape Speech Acts



1. Affordances Shape Speech Acts

TWITTER - X / TECH / MOBILE

X stops showing headlines because Elon Musk thinks it will make posts look better



/ On iOS, when you try to post a link, you'll only see an image with the included attachment.

By [Jay Peters](#), a news editor who writes about technology, video games, and virtual worlds. He's submitted several accepted emoji proposals to the Unicode Consortium.

Oct 5, 2023, 7:42 AM GMT+10 | [119 Comments](#) / [119 New](#)



1. Affordances Shape Speech Acts



pudding person @JUNLPER · 50m



NEW SCOOP: evidence showing elon musk is a pedophile mounting quickly



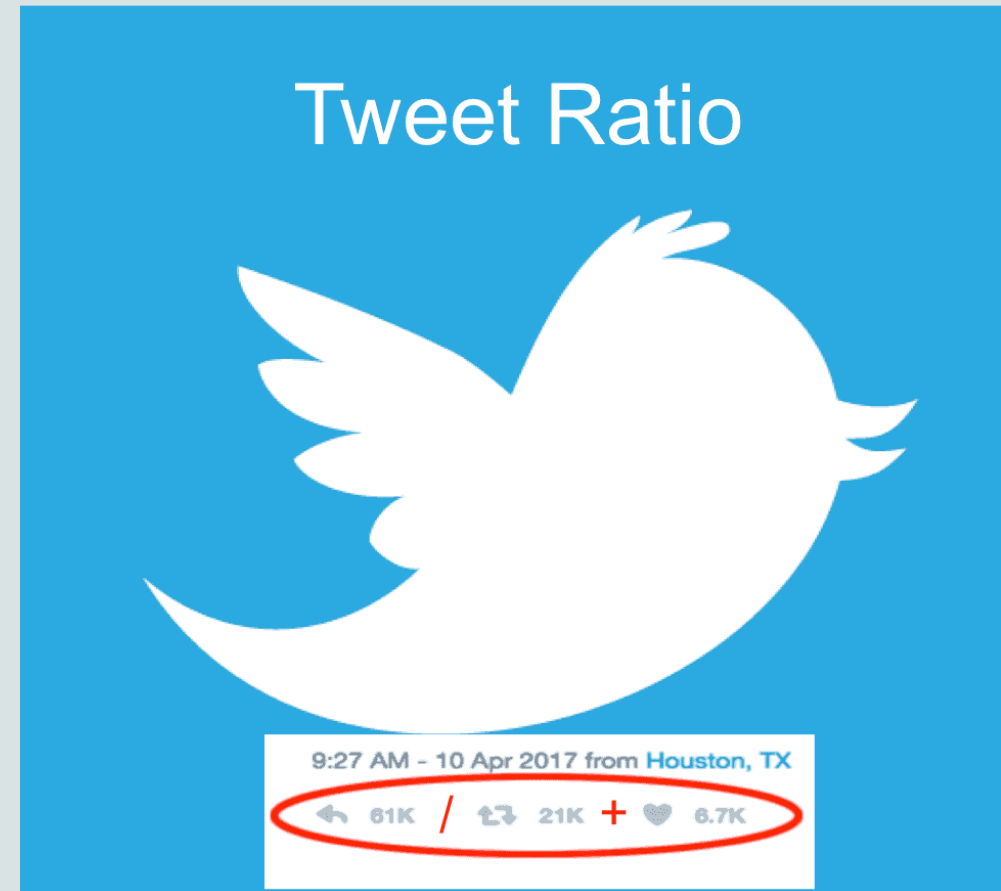


1. Affordances Shape Speech Acts (in a non-neutral way)

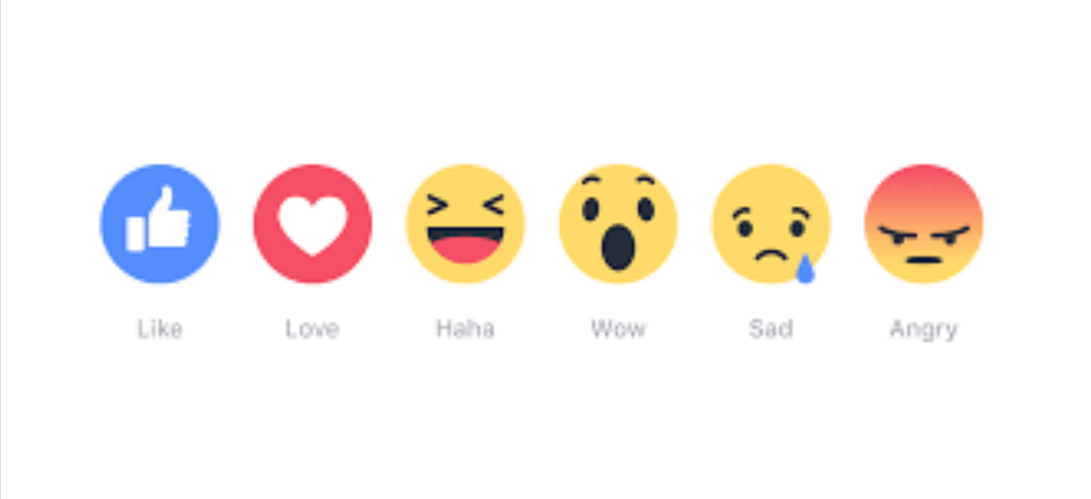
In Algorithmic Environments,
"there is **no 'natural' baseline** to
fall back on," and "**no neutral
middle ground** between prohibiting
an option and endorsing [it]."

- Seth Lazar, forthcoming

2. Affordances shape speech acts by **enabling uptake** that impacts the force of an utterance



2. Affordances shape speech acts by **enabling uptake** that impacts the force of an utterance



3. Platform Affordances are Second-Person calls, Asking for Interaction (and data)

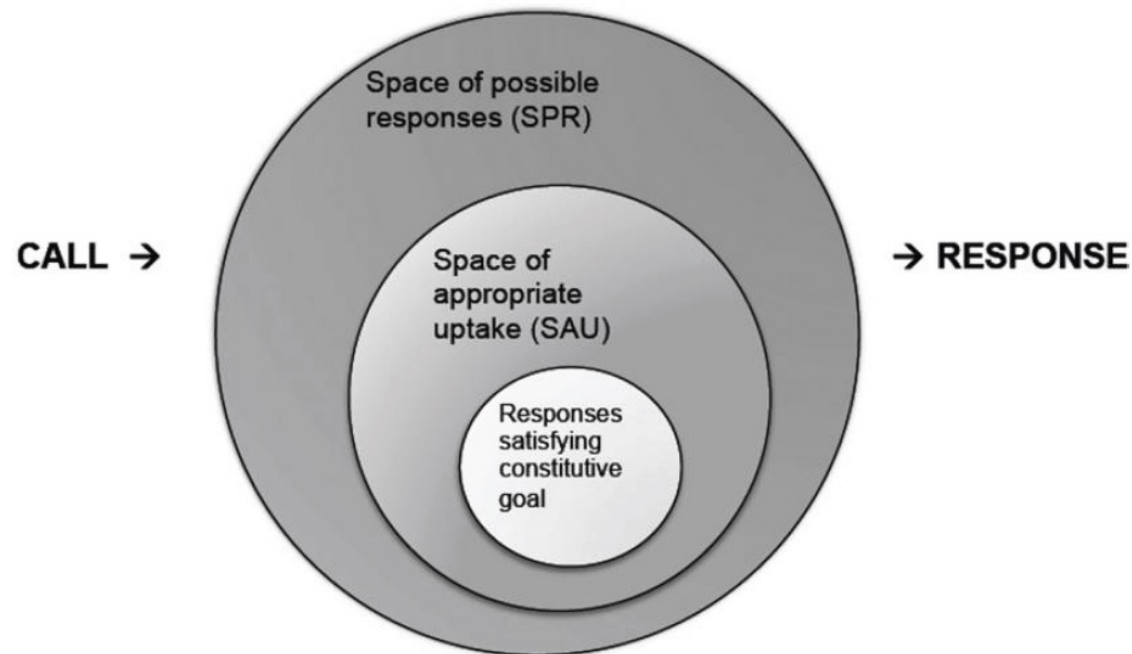
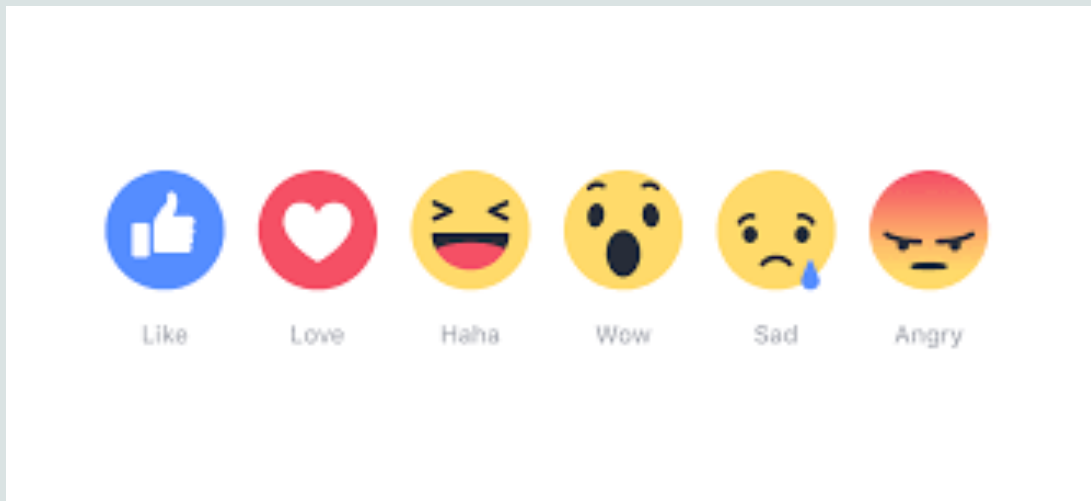


FIG. 1

3. Platform Affordances are Second-Person calls, Asking for Interaction (and data)

Table 3: Interaction-type weights for the MSI formula in 2020.

Interaction type	weight
Like	1
Reaction	1.5
Reshare	1.5
Comment	15-20

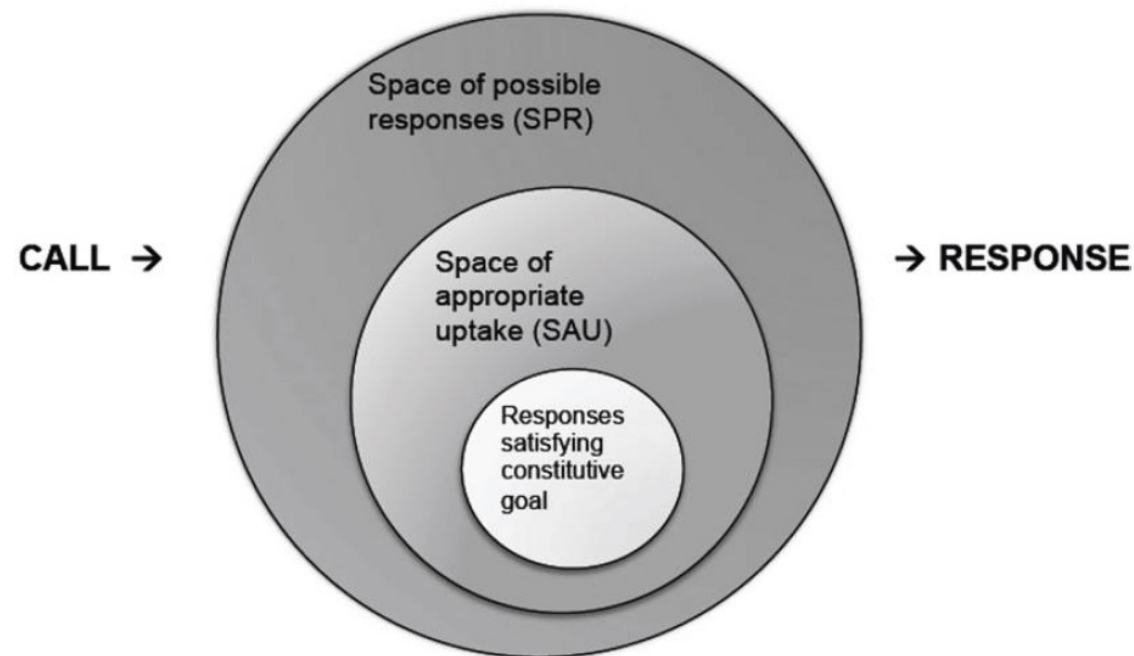
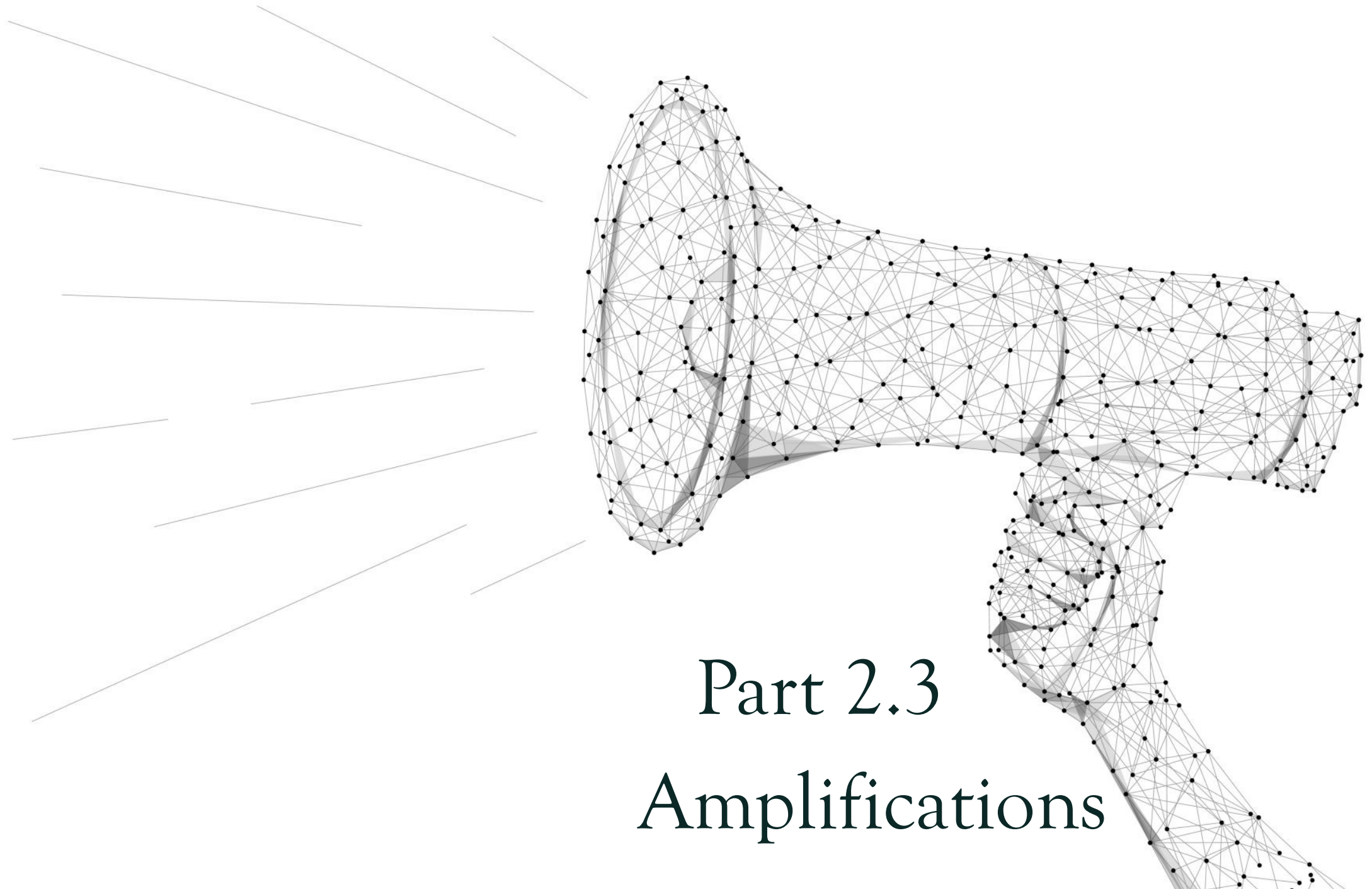


FIG. 1

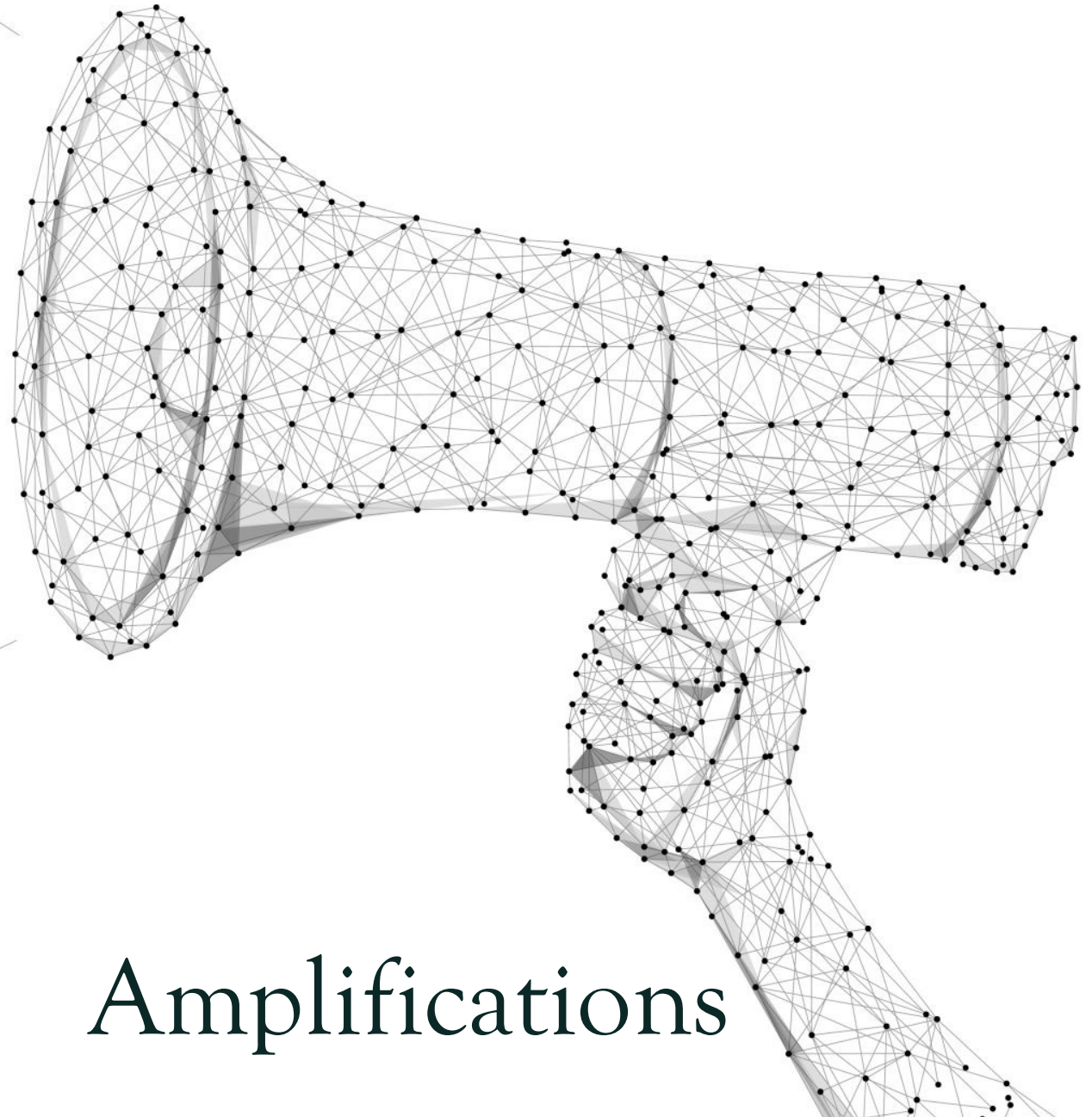


Part 2.3
Amplifications

The Plan

Part 2: Speech Acts on Social Media

1. Algorithms (Platforms)
2. Affordances
3. Amplifications (Recommendations)
 - A. Some useful distinctions;
 - B. How Recommendation Systems Work;
 - C. How this design impacts Speech Acts on Social Media.
 - D. What kind of Speech Act recommendations themselves are.



Amplifications

Wrap Things Up

Retweets (shares)

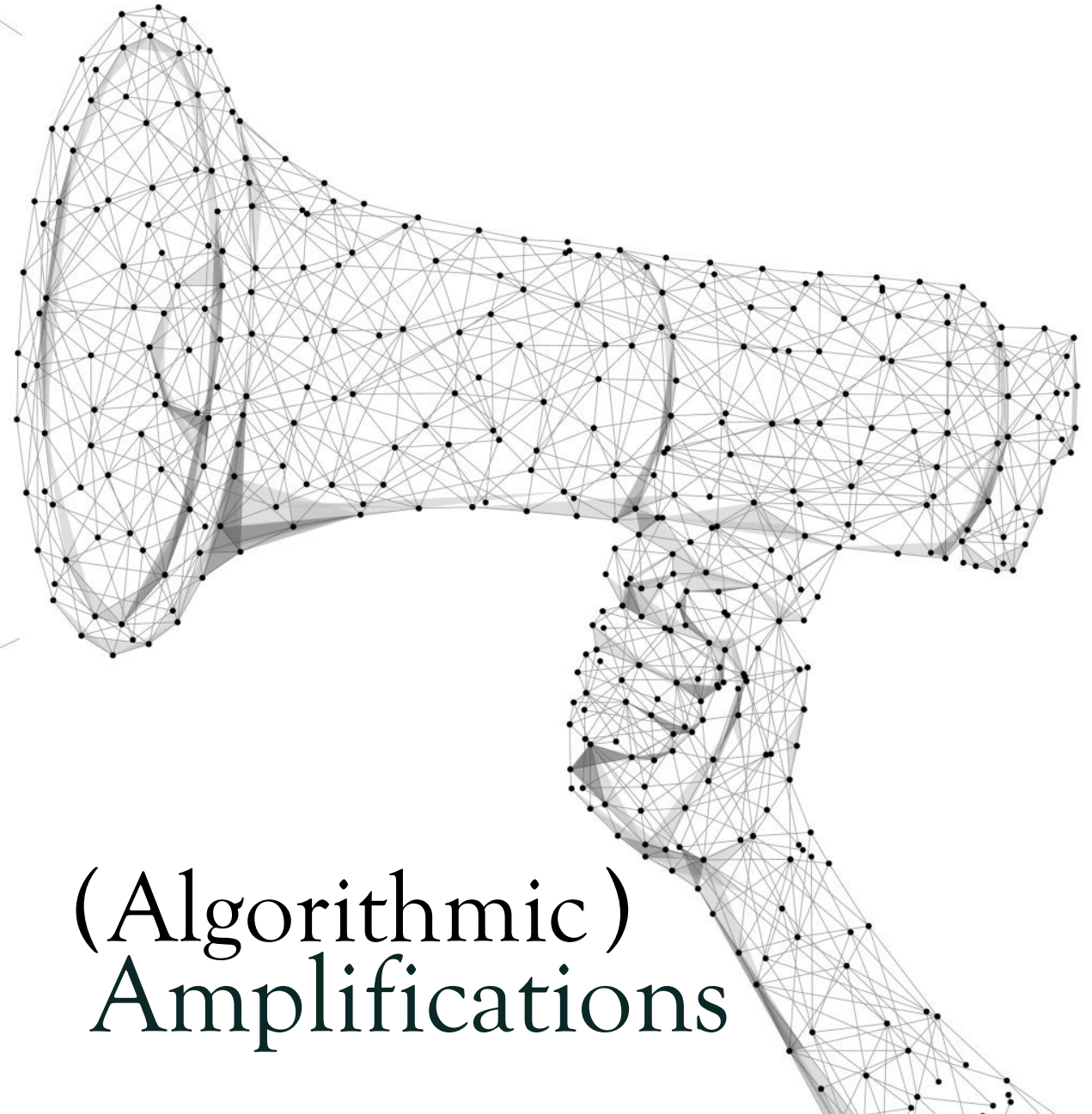
vs.

QuoteTweets (shares with commentary)

vs.

Recommendation Algorithm(s)

All enable different sorts of
'amplification'

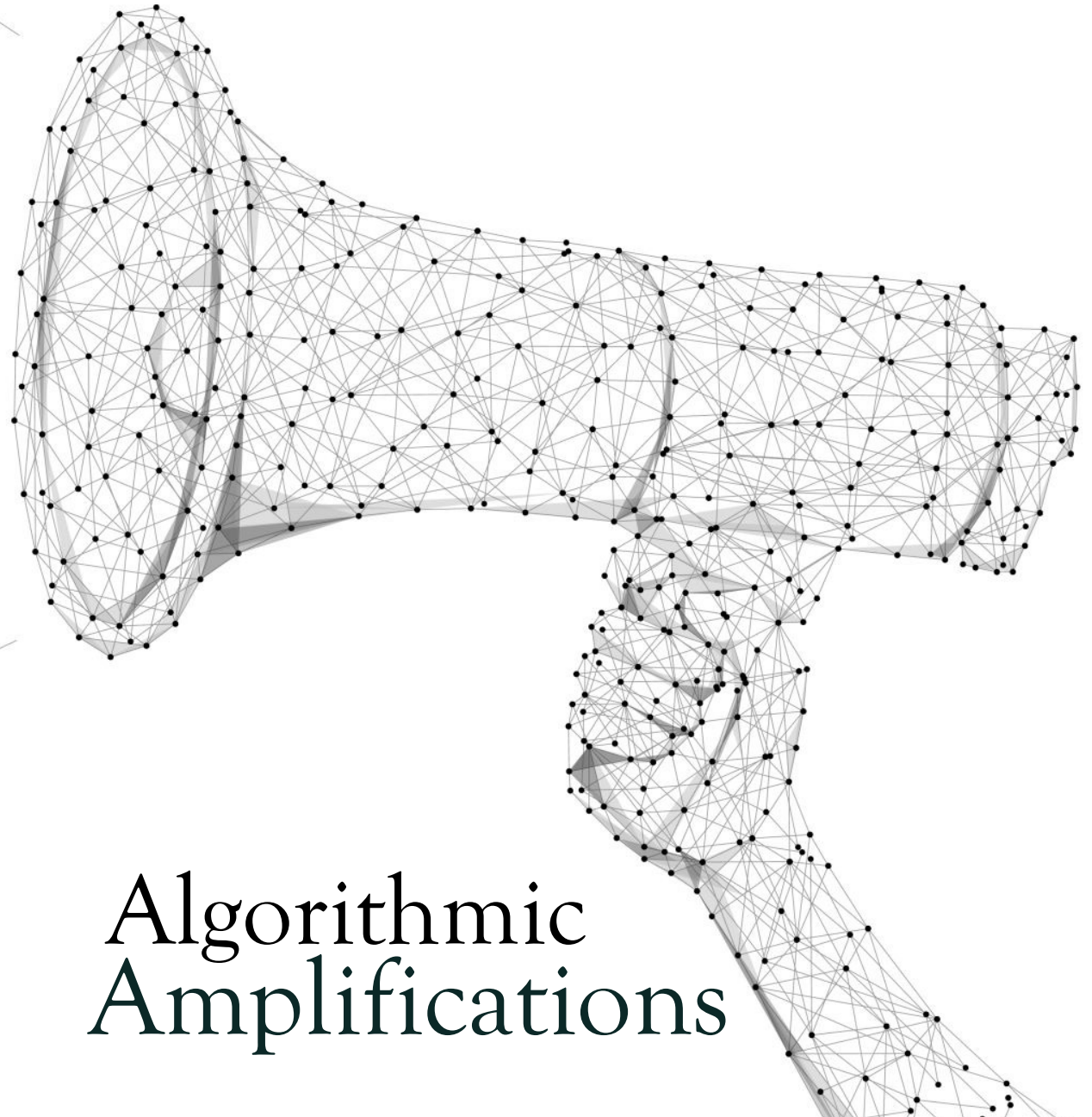


(Algorithmic)
Amplifications

Algorithmic exposure: “the ability of algorithmic systems to expose users to overtly harmful content, such as extremist or radicalizing content and misinformation”

Algorithmic inequality: “the concern that social media platforms unfairly allocate more influence to some types of people than others”

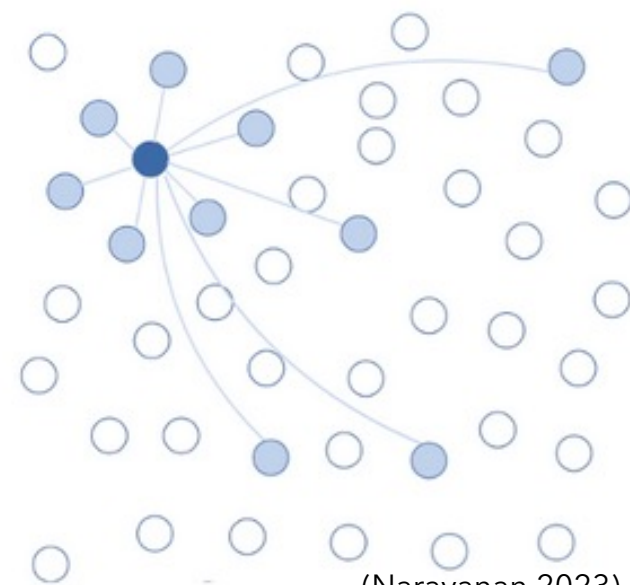
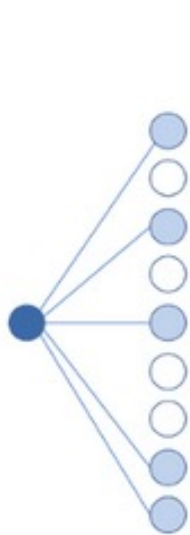
(Lum and Lazovich, 2023)



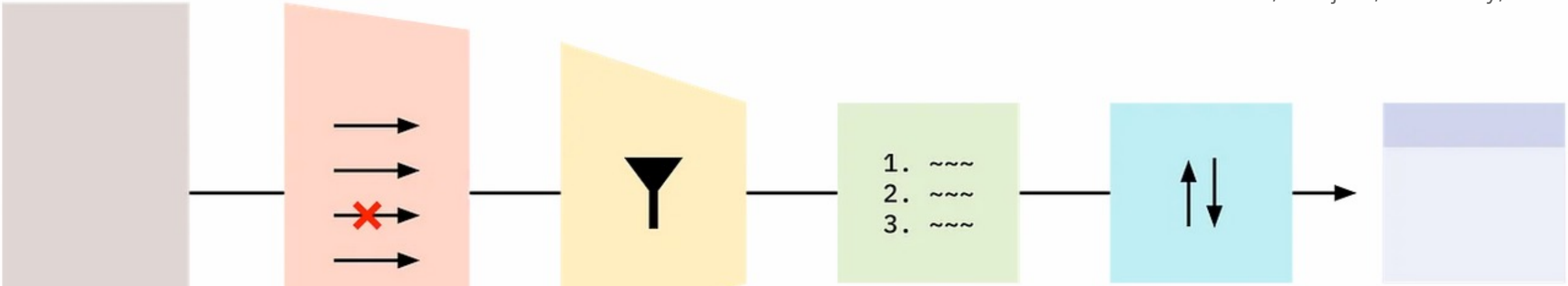
Algorithmic Amplifications

Table 2: Three stylized models of information propagation.

	Subscription	Network	Algorithm
What a user sees	Posts by those they've subscribed to	Posts by (or shared by) those they've subscribed to	Posts the algorithm predicts the user will like best
Examples	Newspapers, Substack, FB pre-2009, IG pre-2022	Word of mouth, the web, Twitter pre-2016, Mastodon	TikTok, Google Discover, YouTube
What impacts a post's reach	Poster's subscriber count	Both subscriber count and content	The content of the post



(Narayanan 2023)



Moderation

**Candidate
Generation**

Ranking

Re-ranking

All > 10^8 items

> 10^8 items

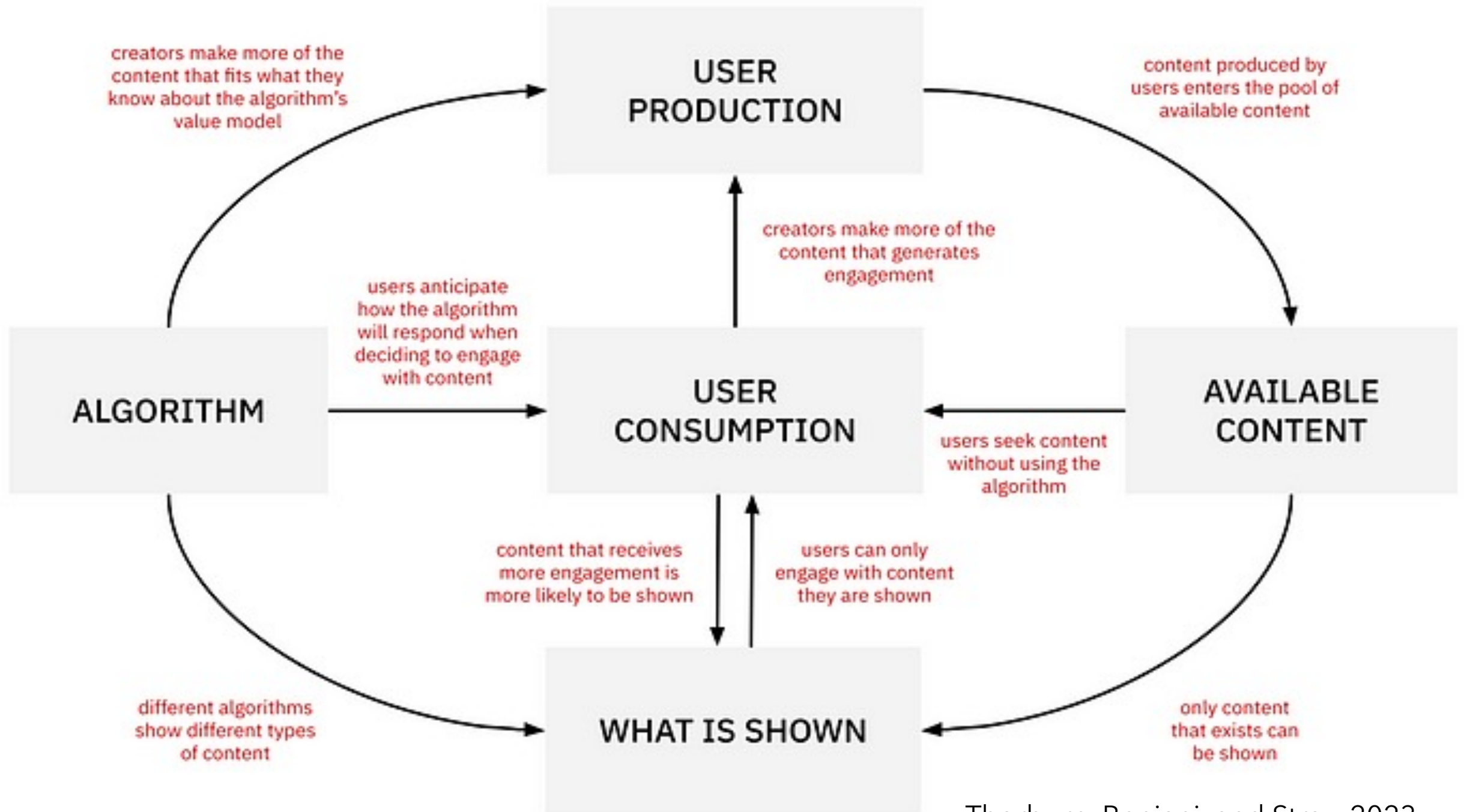
$10^8 \rightarrow 10^2$ items

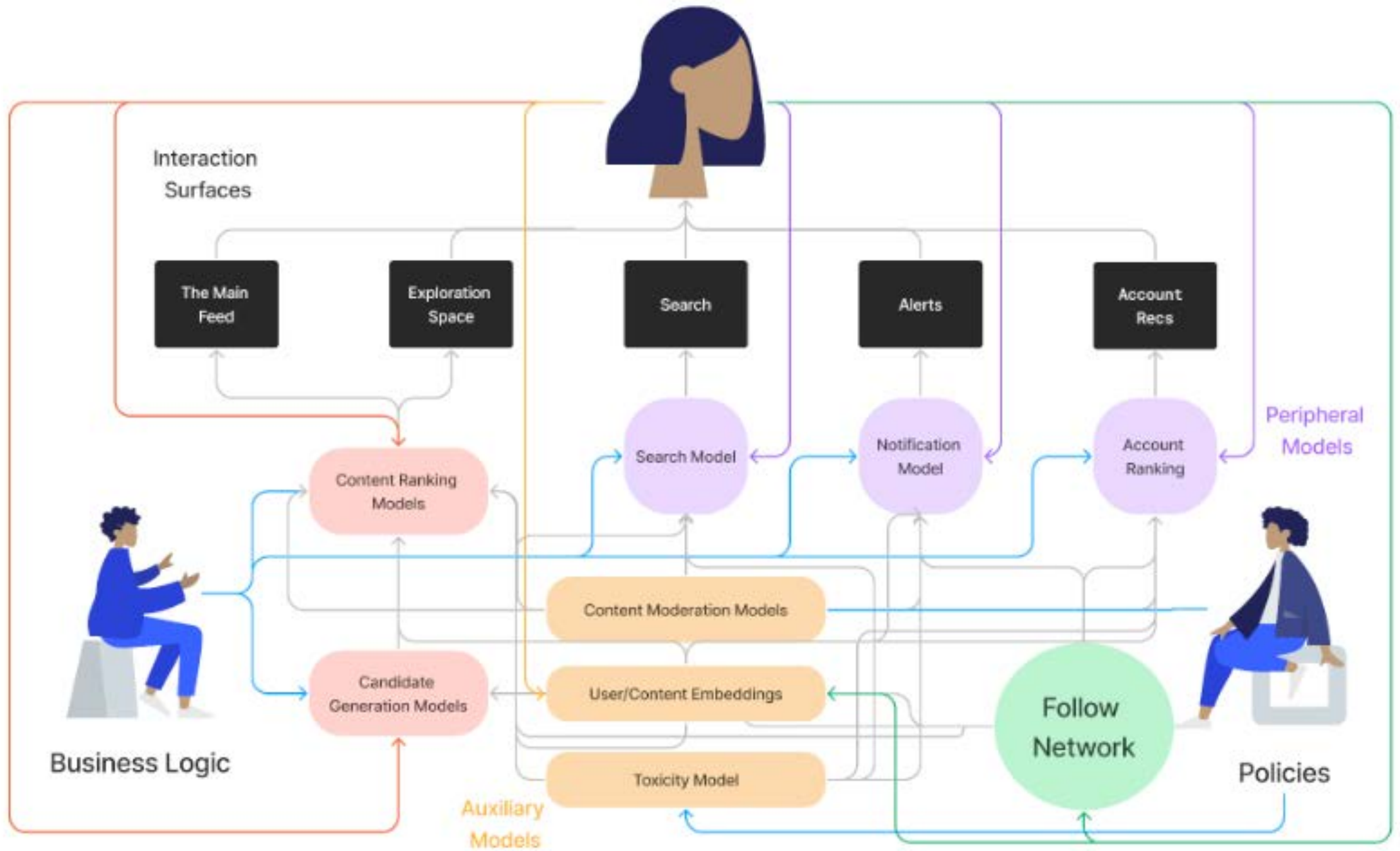
Same 10^2 items

Same 10^2 items

Top 10^1 items
shown to user

A typical recommender pipeline, along with the approximate number of items retained at each stage for a large platform.





(Lum and Lazovich, 2023)

The New York Times

A Genocide Incited on Facebook, With Posts From Myanmar's Military

“Meta’s algorithms proactively amplified and promoted content which incited violence, hatred, and discrimination against the Rohingya - pouring fuel on the fire of long-standing discrimination and substantially increasing the risk of an outbreak of mass violence.”

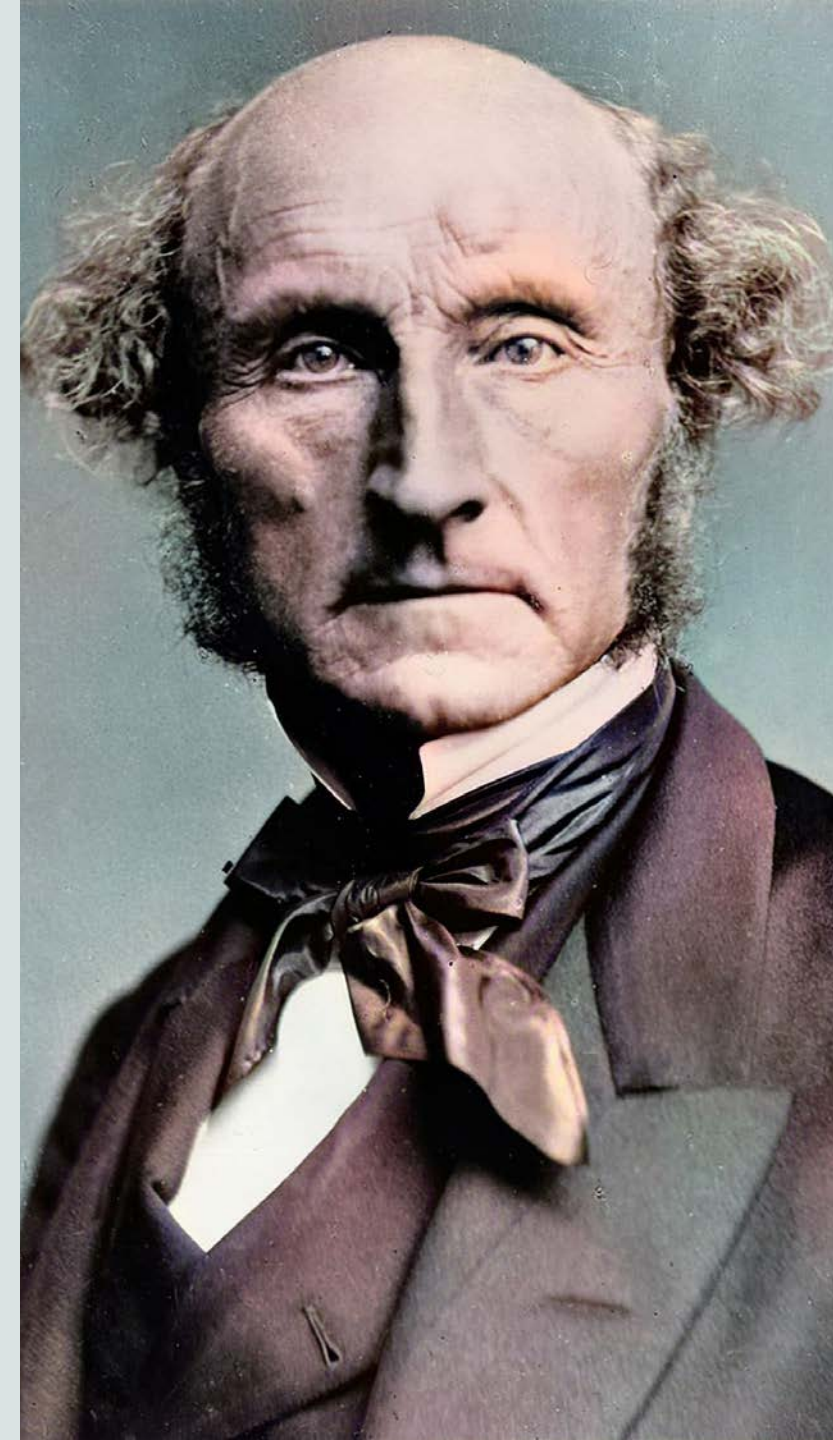
- Amnesty International

“Meta’s algorithms proactively amplified and promoted content which incited violence, hatred, and discrimination against the Rohingya – pouring fuel on the fire of long-standing discrimination and substantially increasing the risk of an outbreak of mass violence.”

- Amnesty International

Consider the ‘corn-dealers’ from John Stuart Mill’s *On Liberty*

“even opinions lose their immunity, when the **circumstances** in which they are expressed are such as to **constitute** their expression a positive instigation to some mischievous **act**. An opinion that corn-dealers are starvers of the poor, or that private property is robbery, ought to be unmolested when simply circulated through the press but may justly incur punishment when delivered orally to an excited mob assembled before the house of a corn-dealer, or when handed about among the same mob in the form of a placard.



Incitement (Online)

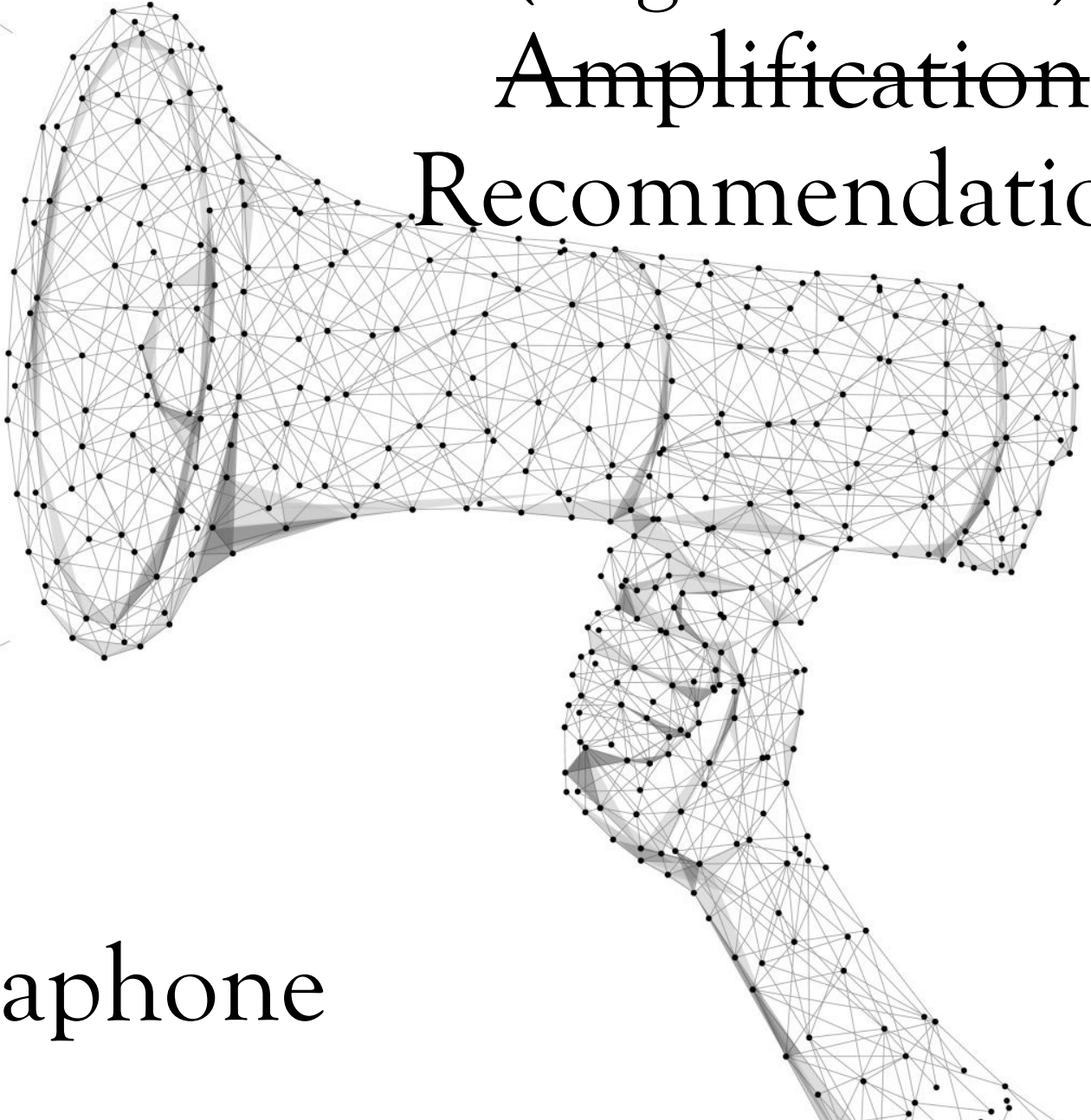
“when handed about among the same mob in the form of a placard.”

Note: Time of production differs from time of use.

In cases of ***delayed communication***, the illocutionary force of written text or a recorded utterance isn't determined by the context of production, but something else.

- *Intended Context vs. Decoded Context*

(Algorithmic)
~~Amplification~~
Recommendation



Not a Megaphone

Jeff Howard



- “My claim is that when a platform amplifies wrongful speech, increasing its visibility, it thereby **makes a greater causal contribution to the speaker’s wrongdoing**—making his principal wrongdoing worse than it would otherwise be.
- Like the gun vendor who sells the terrorist a larger weapon, enabling him to kill more people, **platform amplification enables wrongful speakers to commit a greater wrong.**”

Recommender Systems Transform Speech, and Make a Moral Difference

Recommendation does not simply increase the reach of stable speech acts.

So, they do more than “enables wrongful speakers to commit a greater wrong.”

By choosing the **audience**, and placing speech acts in new **contexts**, and by being speech acts themselves, they enable **something different**.

Recommender Systems Transform Speech, and Make a Moral Difference

Recommendation does not simply increase the reach of stable speech acts.

So, they do more than “enables wrongful speakers to commit a greater wrong.”

By choosing the **audience**, and placing speech acts in new **contexts**, and by being speech acts themselves, they enable **something different**.

Joint Speech Acts?

Co-Speakers?

Collective Speech?

The Speech Act(s) of Recommendations

The Speech Act(s) of Recommendations

“The *content of recommendations is not truth-evaluable*: when the speaker offers a recommendation, she is *not describing* things as being a certain way, but *rather inviting the audience to do something*.”

Javier González de Prado Salas & Ivan Milić (2018).

Recommendations have verdictive and exercitive force

The Speech Act(s) of Recommendations

Recommendations have verdictive and exercitive force

- Recommendations function as **directives**.
 - They differ in **strength** and **structure** for nearby speech acts (**invitations**, **requests**, etc.)
- They are **second-personal speech acts**.
 - They are issued by **one agent** to **another**: the **recommender** the **recommendee**.
- They presuppose the recommendee has reason to **trust** the **recommender's** judgment.
 - That is, the recommender is **entitled** to issue their recommendation.

The Speech Act(s) of Recommendations

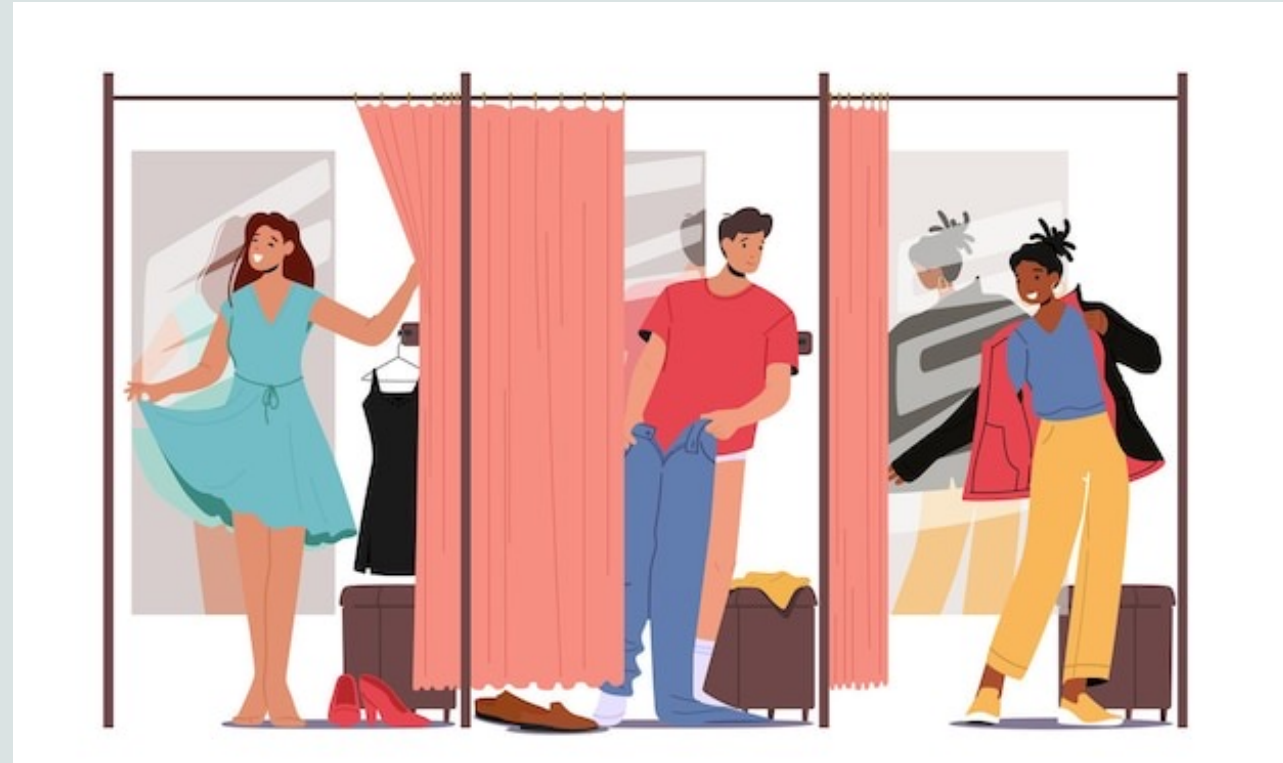
- Recommendations get the speaker on the hook.



The Speech Act(s) of Recommendations

- Recommendations get the speaker on the hook.

But? What if they are quite *thin*?

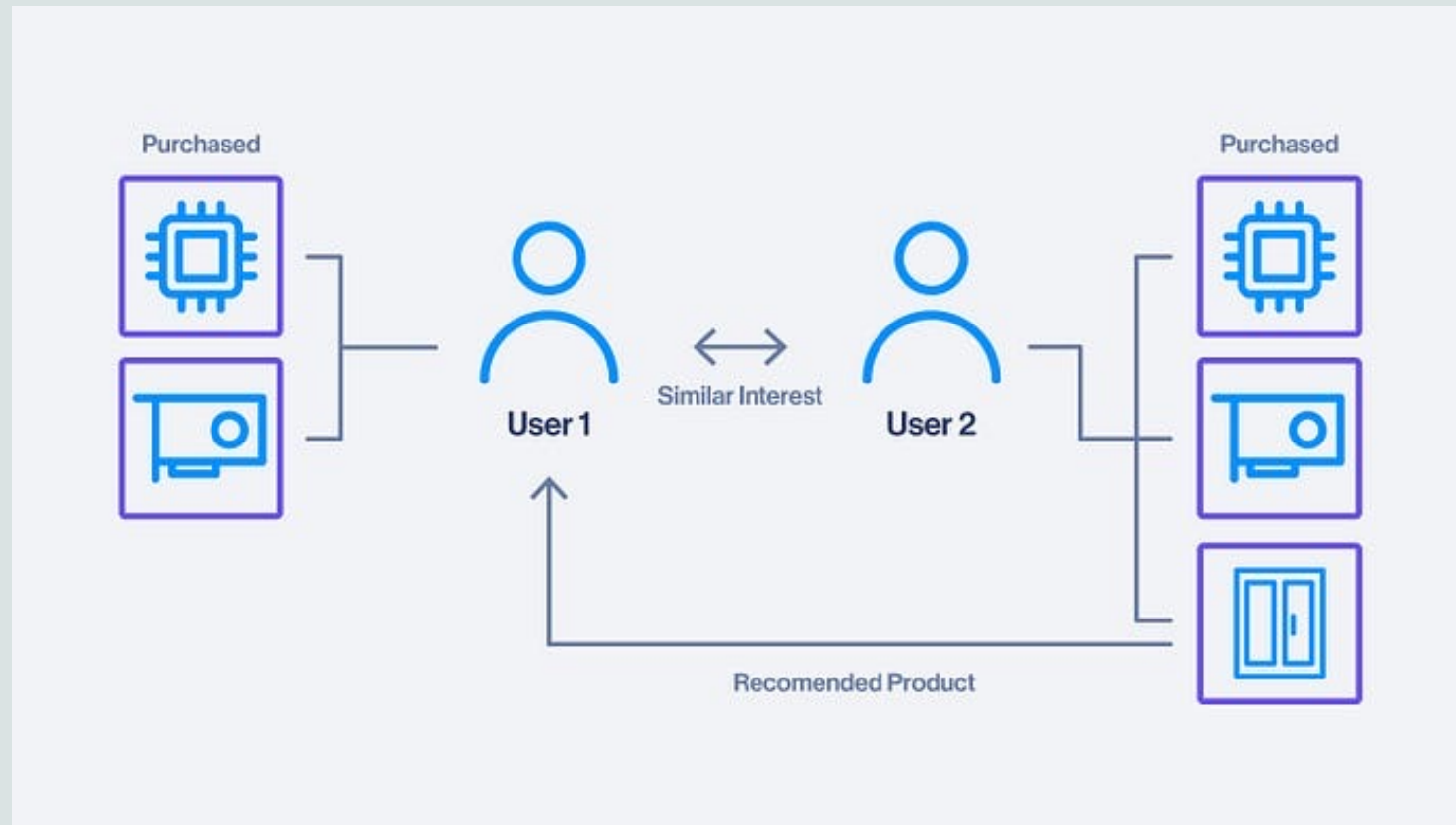


The Speech Act(s) of Recommendations

- Recommendations get the speaker on the hook.

But? What if they are quite *thin*?

- My Claim: Even when based simply on behavioural data – and so not intended as an evaluation of the content – recommendations can still be taken as value-laden.



The Speech Act(s) of Recommendations

Implicit vs. Explicit Feedback Mechanisms



Behavioral data—data on past engagement—is the critical raw material for recommendation engines. The more data, the better the model is able to drive future engagement. So platforms emphasize feedback types that are more frequent. An example of this viewpoint from YouTube researchers in 2016: “Although explicit feedback mechanisms exist on YouTube (thumbs up/down, in-product surveys, etc.) we use the implicit feedback of watches to train the model, where a user completing a video is a positive example. This choice is based on the orders of magnitude more implicit user history available.

The Speech Act(s) of Recommendations

Recommendations **embed values**. But the **opacity** about **whose values** drive recommender systems, limits user agency and therefore responsibility.

So long as **platforms** make the decisions about which **values** their recommenders serve, they are on the **moral hook** for what those recommendations do.

The Speech Act(s) of Recommendations

Leveraging Large Language Models for Recommendation and Explanation

Itallo Silva¹, Alan Said^{2,*}, Leandro Balby Marinho¹ and Martijn Willemsen³

¹Federal University of Campina Grande, Brazil

²University of Gothenburg, Sweden

³TU Eindhoven & JADS, The Netherlands



Large Language Models for Recommendation: Progresses and Future Directions

Authors:  [Keqin Bao](#),  [Jizhi Zhang](#),  [Yang Zhang](#),

 [Wang Wenjie](#),  [Fuli Feng](#),  [Xiangnan He](#) [Authors](#)

[Info & Claims](#)



The Plan

Set up ~~= Why This Matters~~

~~Part 1: Background~~

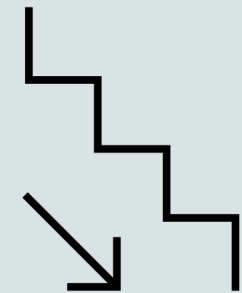
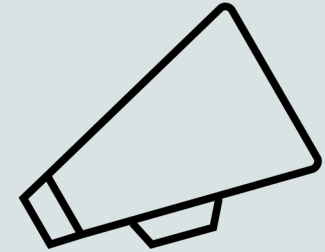
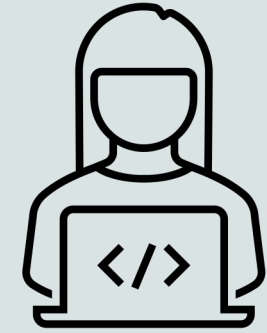
- ~~1. Clarify My Aims~~
- ~~2. Why Social Media?~~
- ~~3. How Speech Act Theory Can Help Here?~~

~~Part 2: Speech Acts on Social Media~~

- ~~1. Algorithms (Platforms)~~
- ~~2. Affordances~~
- ~~3. Amplifications (Recommendations)~~



Wrap Things Up





Wrapping Up

Objection: Isn't this far too over-inclusive of what counts as a speech act?

The Aims and Purposes of
Speech Act Theory:

Political not Metaphysical



Wrapping Up

The Upshot:

Social Media Platforms are *not* mere conduits; they are **constitutive intermediaries** of our online speech acts. They **shape our speech acts** through their choices regarding **context, audience, and uptake**.

They speak themselves through their **recommendations** and various **affordances**, both of which express their **values** about how users should use the platform.

They share **responsibility** for what occurs on their platforms, not because they are (merely) complicit, but because they are **co-producers** of the acts that occur there.

They represent **something new**. We must collectively come to a decision about what that thing is, and how we want it to act.

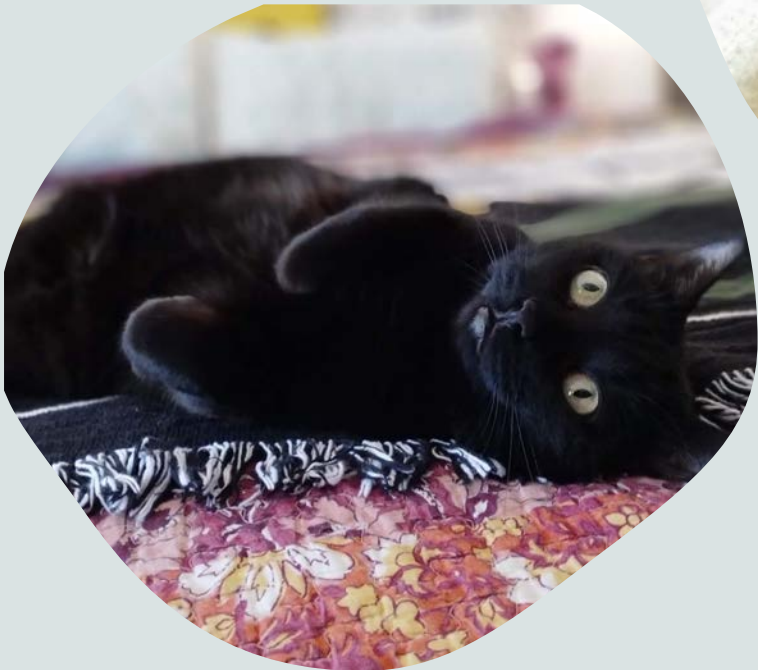
Thank You



soda



maki



michaelrandallbarnes.com

michael.barnes@anu.edu.au

REFERENCES:

J.L. Austin, 1972. *How to Do Things with Words*.

C. Bianchi, 2014, "How to do things with (recorded) words"

J. Davis. 2020. *How Artifacts Afford*.

D. Fogal, D. Harris, and M. Moss, 2018, "Speech Acts: The Contemporary Theoretical Landscape."

J. Gibson, 1979, *The Ecological Approach to Visual Perception*.

T. Gillespie, 2018, *Custodians of the Internet*.

D. Harris and R. McKinney, 2021, "Speech Act Theory: Social and Political Applications."

J. Hornsby and R. Langton, 1998, "Free Speech and Illocution."

D. Keller, 2021, "Amplification and Its Discontents."

Q. Kukla, fc, "'Don't @ Me'".

S. Lazar, fc, *Connected by Code*.

K. Lum and T Lazovich, 2023, "The Myth of 'The Algorithm': A system-level view of algorithmic amplification."

C. MacKinnon, 1993, *Only Words*.

E. Michaelson, J. Pepp, and R. Sterken, 2021, "Online Communication."

S. Milano, M Taddeo, and L. Floridi. 2020. "Recommender Systems and their Ethical Challenges."

J.S. Mill, 1859, *On Liberty*.

A. Narayanan. 2023. "Understanding Social Media Recommendation Algorithms: A discussion guide."

D. Norman, 1998, *The Design Of Everyday Things*.

J. Saul, 2006, "Pornography, Speech Acts and Context."

L. Thorborn, P. Bengani, J. Stray, 2022, "How Platform Recommenders Work."

L. Thorborn, P. Bengani, J. Stray, 2023, "Making Amplification Measurable."



Wrapping Up

Some Questions:

1. *Algorithmic Intermediaries*

Given that **users** can **push back** against the top-down governance of algorithmic intermediaries, how should we understand the **power dynamics** at play?

2. *Affordances*

Who is the '**speaker**' of the various features that have affordances of interest? And how do we go about analyzing their **force**, given their ambiguity?

3. *Amplifications Recommender Systems*

If the **decoded context** of recommended utterances fixes illocutionary force, how do we narrow that down given the sometimes-literal **millions of options**?