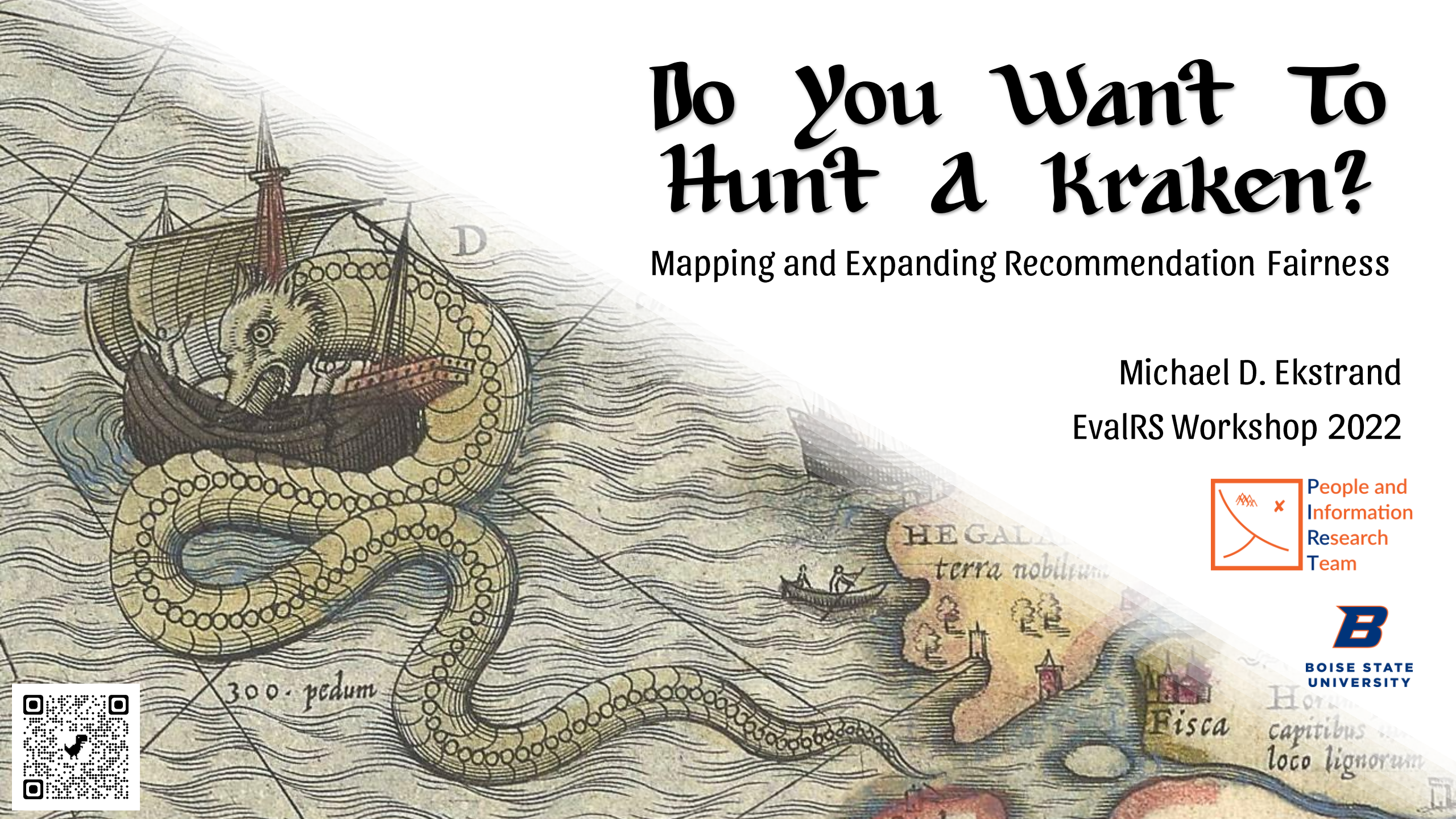


Do You Want To Hunt A Kraken?

Mapping and Expanding Recommendation Fairness

Michael D. Ekstrand

EvalRS Workshop 2022

People and Information Research Team

BOISE STATE UNIVERSITY

concerning monsters

Medieval maps put sea monsters in unexplored space

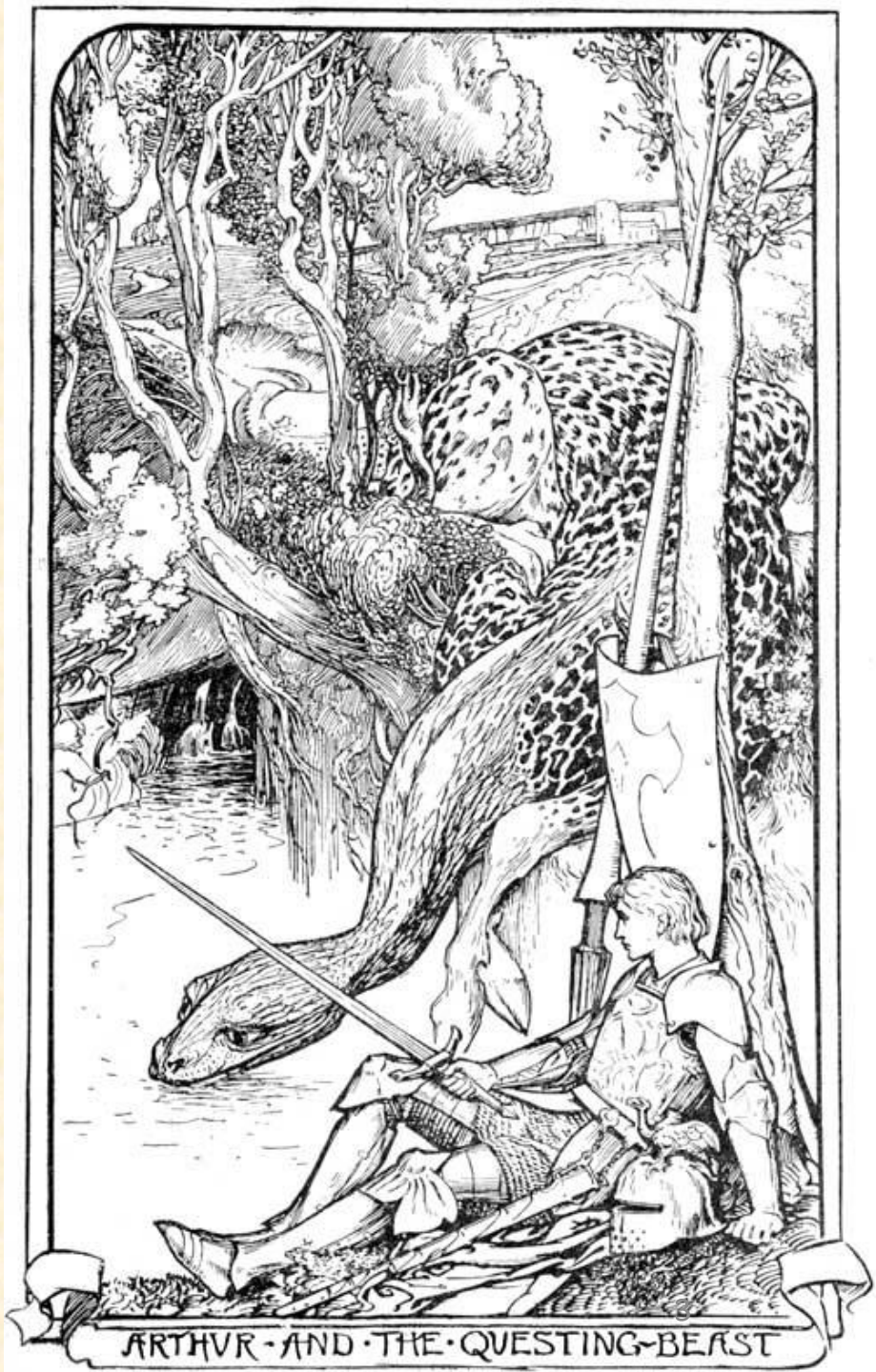
The map of **fair recommendation**:

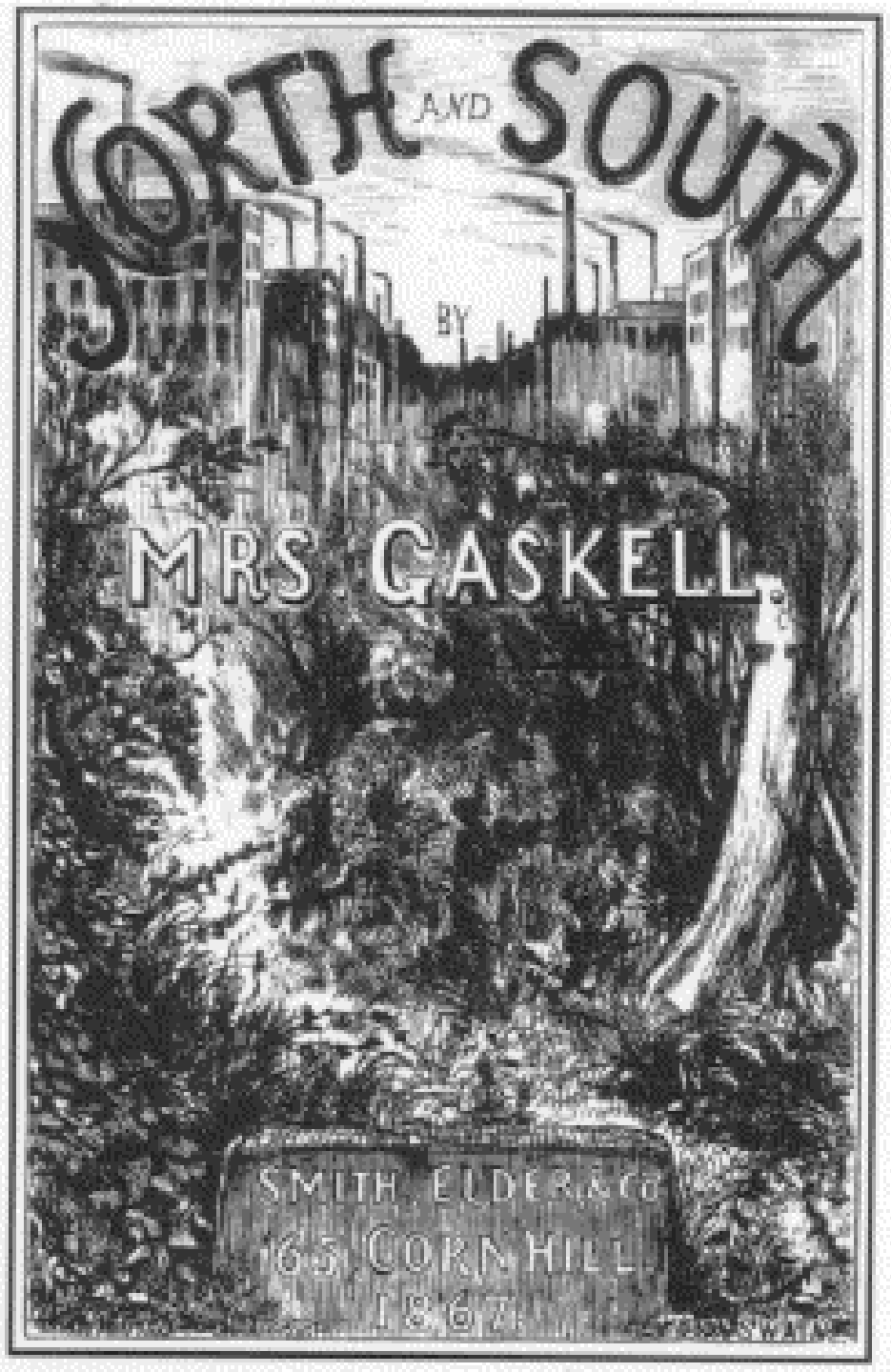
- Broad space of problems
- A few points studied
- Much is unknown!

our quest

Point you to the monsters!

- What does the broad map look like?
- Where do EvalRS objectives fit?
- Where do we need to go exploring?

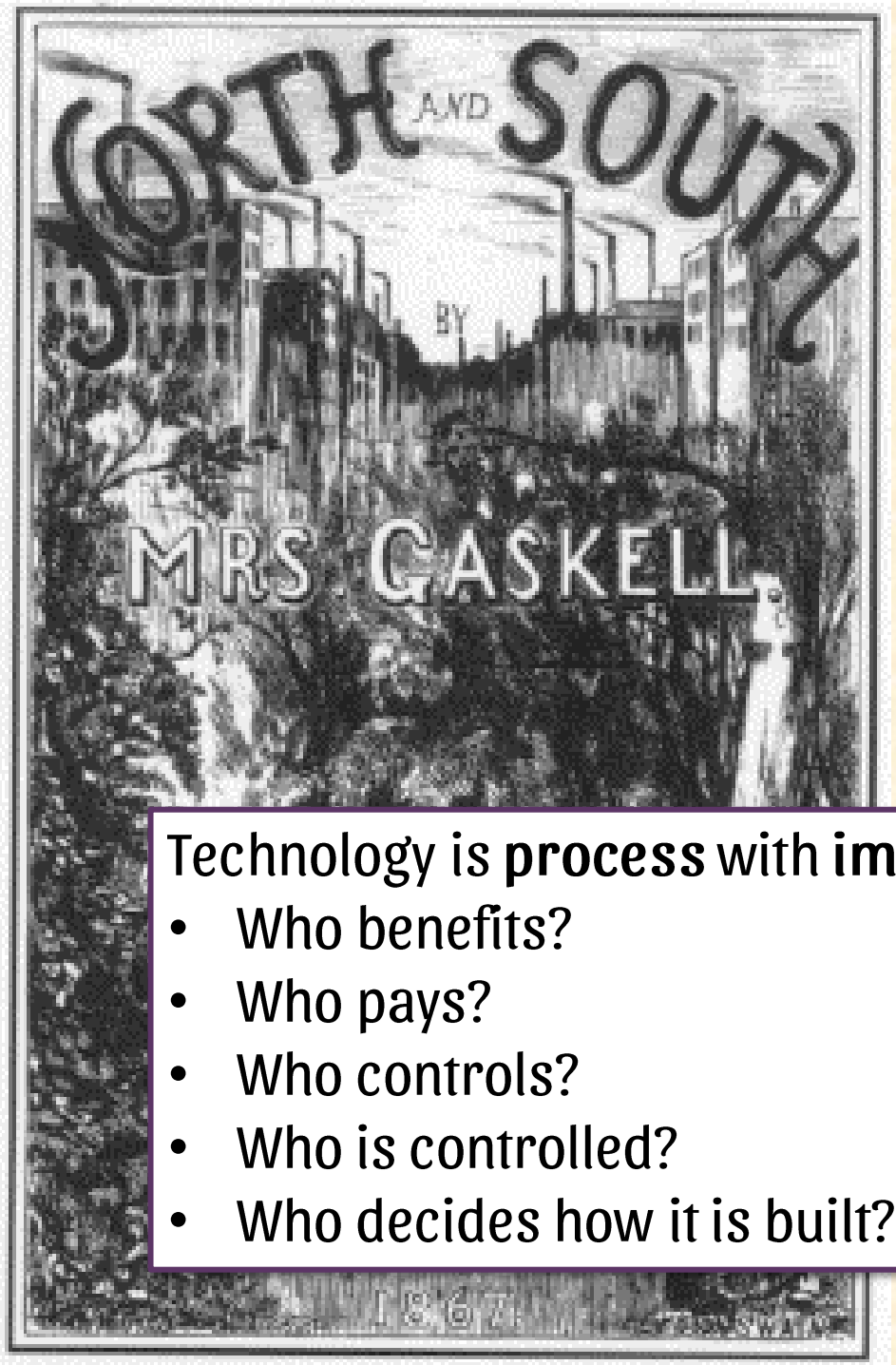




North and South

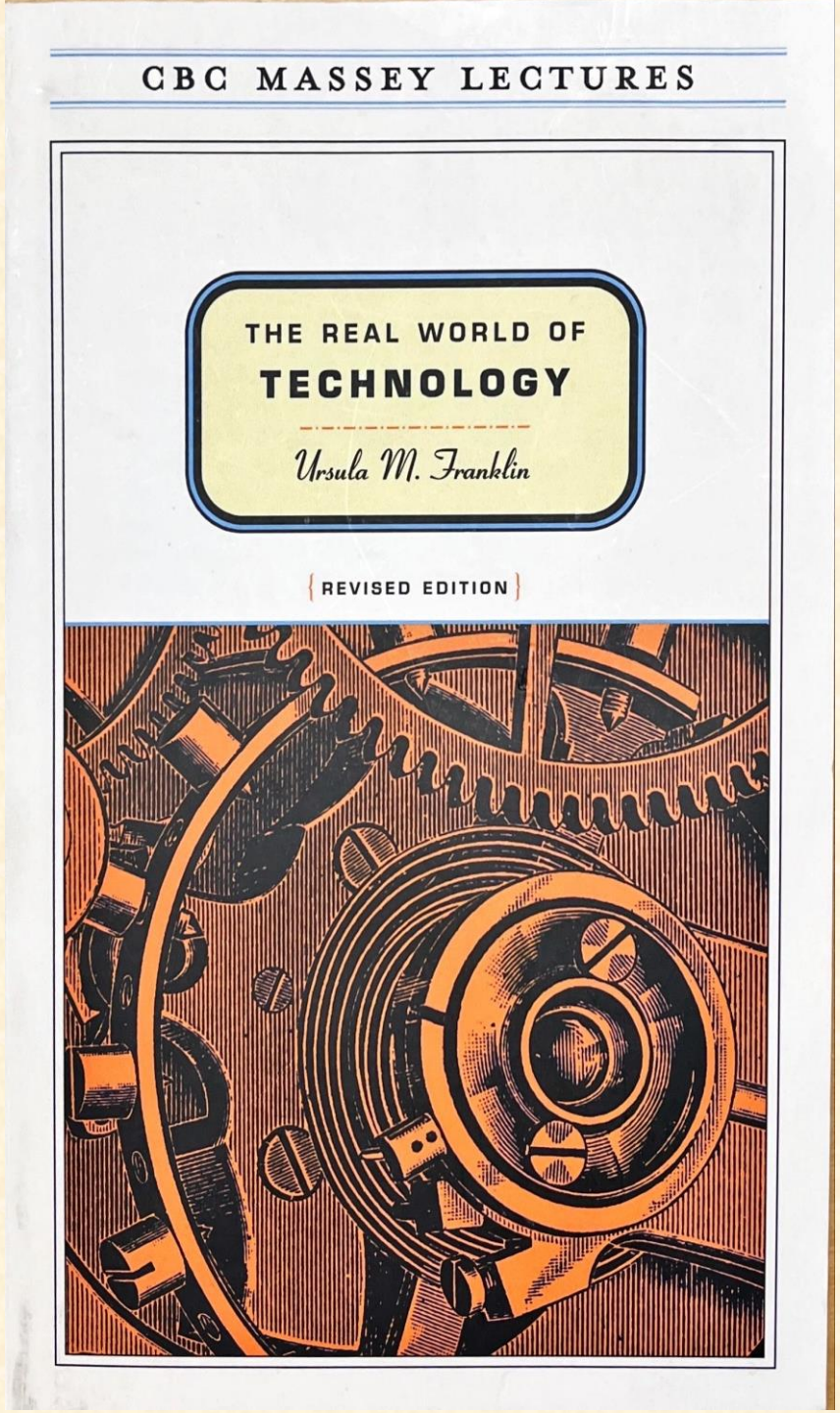
Elizabeth Gaskell, 1854

(cover art from 1867 edition, drawn by George du Maurier)

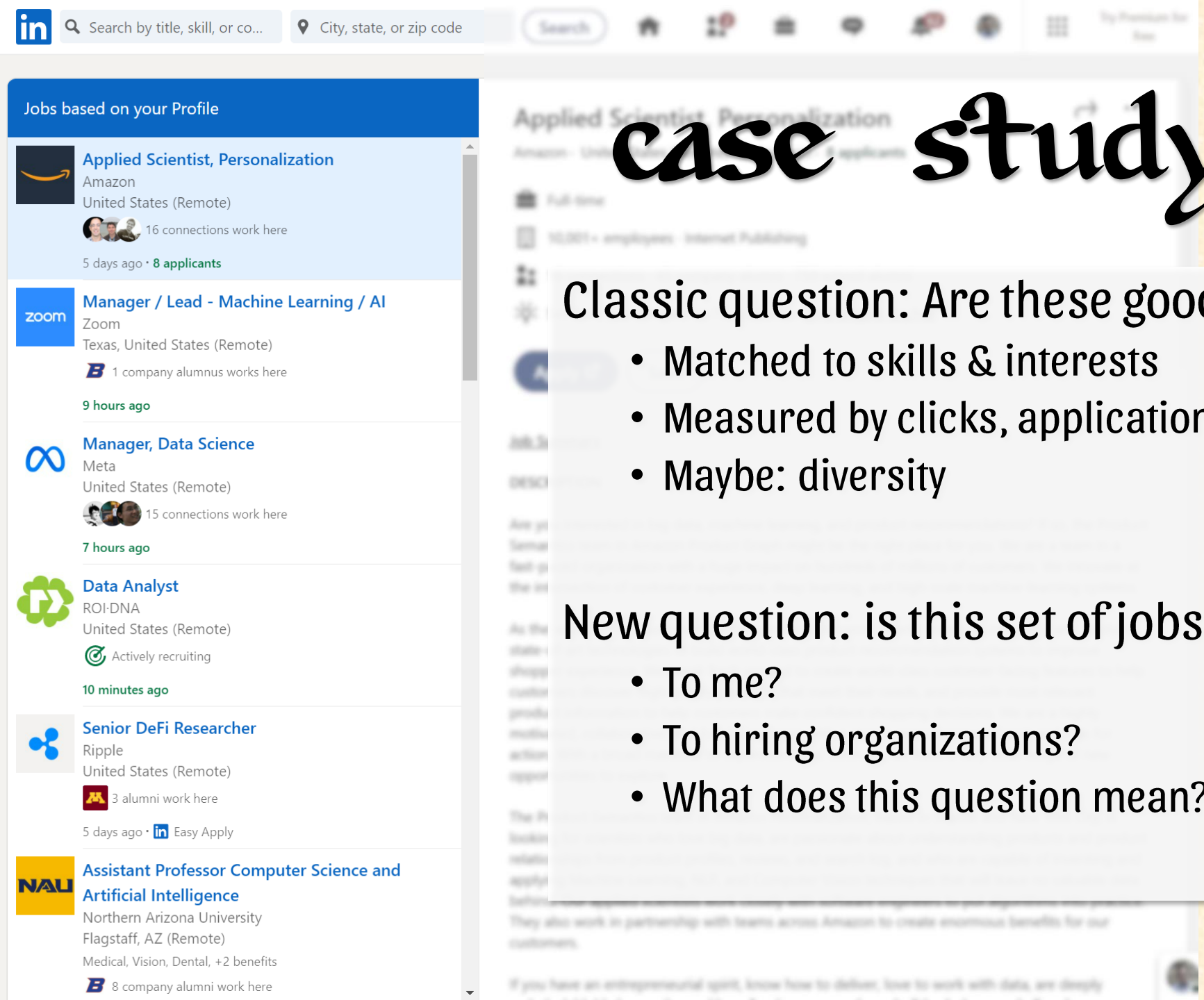


Technology is **process** with **impacts**

- Who benefits?
- Who pays?
- Who controls?
- Who is controlled?
- Who decides how it is built?



case study - jobs



Classic question: Are these good recommendations?

- Matched to skills & interests
- Measured by clicks, applications
- Maybe: diversity

New question: is this set of jobs *fair*?

- To me?
- To hiring organizations?
- What does this question mean?

no final answer

There isn't "an answer" to bias, fairness, or ethics.

Andrew Selbst et al. FAccT 2019, § "The Formalism Trap":

> Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms

Further research:

- Fairness depends heavily on assumptions and goals
- Not all reasonable goals mutually compatible (but situation complex!)
  - See: Friedler et al. 2021, Mitchell et al. 2020, Chouldechova 2017, Binns 2020

specific harms

We **can't**:

define fairness as a clear, universal objective

We **can**:

Identify, measure, and mitigate specific harms
Study how harms relate to each other
Look for general strategies and principles

Target specific kinds of **unfairness**

mapping harms

- Identify dimensions for locating fairness-related information access harms
  - Not orthogonal or hierarchical
- Provide guidance for positioning these harms
- Survey a lot of research

Ekstrand et al., FnT IR 2022

fair for who?

Consumers
(users)

Providers
(artists, authors, etc.)

Vendors

Subjects

Society

Shareholders

Publishers

individuals and groups

Individual fairness:

- Give each individual what they "deserve"
- Treat similar individuals similarly

Group fairness:

- Avoid systematic disparities between groups
- Can target explicit treatment, outcomes, or errors
- Common groups: gender, race, religion, age
- Also relevant: popularity, genre, rookie / established

EvalRS: fair utility

Utility metric: miss rate

Computed by consumer, producer, or item groups

Computed as: $|\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} \left| \bar{M}_g - \bar{M} \right|$ (mean of abs. diff. group vs. all)

group equity of utility

Goal: each consumer group should have comparable utility

Ekstrand and Pera [FAccTRec 2022]: *equity of utility*

Groups:

- User gender
- User country
- User activity level

consumer fairness

Scoping [Beattie et al. 2022]

What is your **goal**?

What is your **unit of analysis**? (individual, group; what groups?)

What is your **metric**?

What is your **aggregation**?

consumer fairness goals

Many goals possible [Ekstrand and Pera, FAccTRec 2022]:

- Equity of utility
- Equity of usability
- Fair representation
- Stereotype avoidance
- … many more possible!

Many strategies can pursue these goals

Ponder: what if goal is for *ecosystem* to be fair, not just each system?

aggregates and equity

Difference (for multiple groups):

$$|\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} \left| \bar{M}_g - \bar{M} \right|$$

How to minimize? All groups same metric

zero-sum, but is consumer utility rivalrous?

Alternative: **positive-sum** [Wang and Joachims, ICTIR 2021]: $\sum_{g \in \mathcal{G}} \log \Upsilon_g$

Improve by adding utility to **least-utility group** — highest payoff in final metric

provider equity

Goal: be equally accurate at retrieving items from each provider group

Item equity works the same way—song popularity

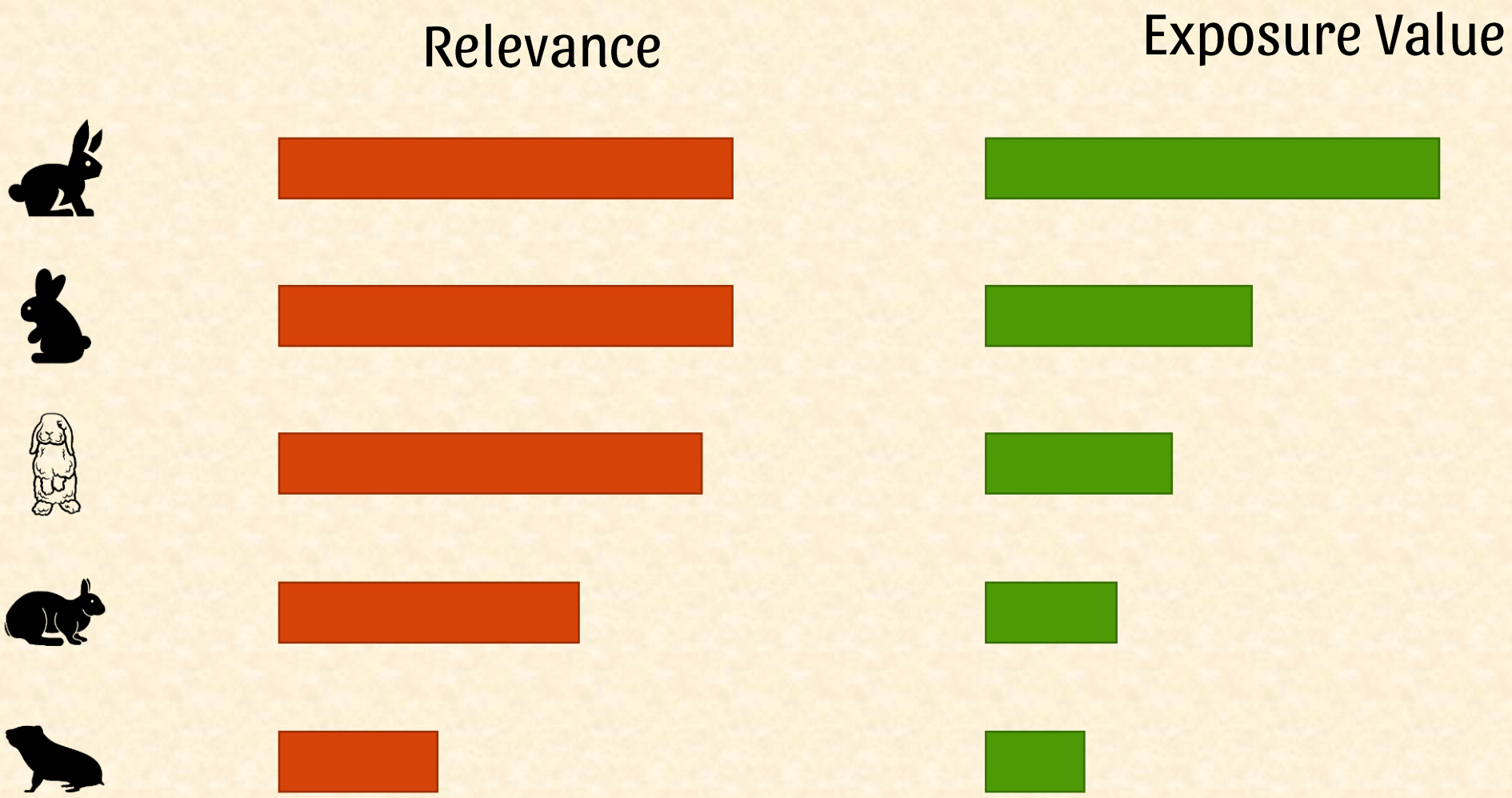
Groups: artist popularity (can apply to any other groups!)

Principle:

If the user wants a less-popular artist
Then the system should provide it just as well

Consumer-focused view of provider equity (cross-group harm)

exposure and fairness



problems

- Comparably-relevant items get different exposure
  - Violates individual fairness
  - If systemic by group: violates group fairness
- Only one item gets first position in any one ranking
  - So we can't fix the problem!
  - … at least at one time

expected exposure

Diaz et al., CIKM 2020

Problem: items get exposure based on rank position, is it fair?

- measure their **exposure**
- look for **equity**

Principle of Equal Expected Exposure:

> Given a fixed information need, no item should receive more or less expected exposure than any other item of the same relevance grade.

expected exposure

Diaz et al., CIKM 2020

Problem: items get exposure based on rank position, is it fair?

- measure their **exposure**
- look for **equity**

Principle of Equal Expected Exposure:

> Given a fixed information need, no item should receive more or less **expected exposure** than any other item of the same relevance grade.

One ranking can't be fair – take expectation over **stochastic rankings**

eee and goals

Operationalizes **equality of opportunity**:

- Equally-relevant documents should have equal opportunity for good ranking

Extends to **group fairness**

- Does a group of providers get more or less exposure than "deserved"?
- Is exposure conditionally independent of group given relevance?

Open questions:

- Relevance (and group) data is often missing
- Relevance data is often likely biased
- Generalizing to multiple information needs & other fairness targets

comparing goals

Provider-disaggregated accuracy metrics:

- Start with consumer-focused utility (are recs good?)
- Ensure recs are good no matter which provider (group) is needed

Exposure/attention metrics:

- Measure provider-side utility (e.g. exposure)
- Ensure providers get comparable utility

See Raj and Ekstrand [SIGIR 2022] for more metrics & comparisons

measuring providers

- Recommendation accuracy (EvalRS)
- Allocation of exposure (Diaz et al. CIKM 2020, Biega et al. SIGIR 2018)
- Equalized (pairwise) odds (Beutel et al. KDD 2019)
- Composition of lists (Ekstrand and Kluver, UMUAI 2021)
- … more

measuring items/subjects

Mathematically, works like provider fairness – look at results

Needs different data

- Item – often easier than provider
  - Item metadata
  - Latent feature spaces (like EvalRS)
- Subject – often harder
  - Wikipedia provides lots of info
  - How do you detect subjects of news article? Research paper?

relating to other concerns

- Utility: may sometimes trade off; not necessarily
- Diversity: very related
  - Subject and provider fairness can be enhanced by diversity
  - Concerns stem from different **normative goals**
- Novelty: can sometimes be a fairness problem!
  - Is our system unfair to new content or new providers?
- Competing fairness concerns

Fairness is **one part** of a **multifaceted** analysis of the system

Sometimes "fairness" may be socially bad!

classes of harms

Distributive/Allocative Harms

Am I harmed by how results or outputs are distributed?

- Disproportionally not recommended
- Lower quality of service

Crawford [NeurIPS 2017]

Representative Harms

Am I represented fairly or accurately in the system?

- Misgendering
- Negatively stereotyped

monster clues

Lots of work on distribution of utility / value

- Provider exposure, consumer utility
- Mostly 1 group (at a time) [exception: TREC Fair Ranking 2021-2022]

Less work on:

- harms of representation!
- multiple groups
- nuances of multiple users / requests for exposure
- statistical analysis of fairness results
- appropriate collection & sharing of data

gender stereotypes

Recent work by Amifa Raj and myself

"Fire Dragon and Unicorn Princess", SIGIR eCom 2022

Problem: society has gender stereotypes

When children are exposed, can have substantial effect!

Question: do search engines reproduce these stereotypes?

Esp. with children!

getting data

Lists of toys with gender stereotypes

- Previous research on stereotypes and children
- Advocacy campaigns for stereotype-free childhoods

Use them as queries

Observe system response

- Query suggestions
- Search response

query suggestions

Provide query "<TOY> for"

Count suggestions with gendered terms

e.g. "<TOY> for boys"

Aligned with gender G if #G greater than other genders

Findings:

E-commerce query suggestions have gender alignments for toys

Alignments correlate with previously-documented stereotypes

generic search results

3 queries (for each of toys, books, games, etc.):

- Toys for kids
- Toys for boys
- Toys for girls

Look at words in gender SERP but *not* neutral SERP (qualitative)

Many words aligned with common gender stereotypes surface

specific search results

Give 3 queries:

- <TOY> for kids
- <TOY> for boys
- <TOY> for girls

Compare Jaccard of gendered results with neutral results

Toy gender-aligned if gender neutral is more similar to one gender than another

Finding: gender stereotypes do propagate

"classical" fairness setting

Shira Mitchell et al. 2020:

- Classification: high/low risk of [crime, default, job failure, etc.]
- Decisions and consequences are individual and independent
- One-shot process (no iterative learning)

Also:

- Process independent of "user" / decision-maker

Exceptions exist, of course.

recsys is different

Decisions are **not independent**

- Only one item gets Rank No. 1

Decisions are **repeated**

- Items have multiple chances to be ranked

Multiple stakeholders have fairness concerns

- Not just data subjects – users, providers, etc.

Decisions are **personalized**

- Different users need different results

Outcome (relevance) is personalized and **subjective**

- The same content won't work for everyone

normative clarity

Be clear about the goal:

- Fair treatment?
- Justice?
- Reparations or reversing oppression?
- Certain conceptions of social welfare?
  - See e.g. Fish et al. 2019

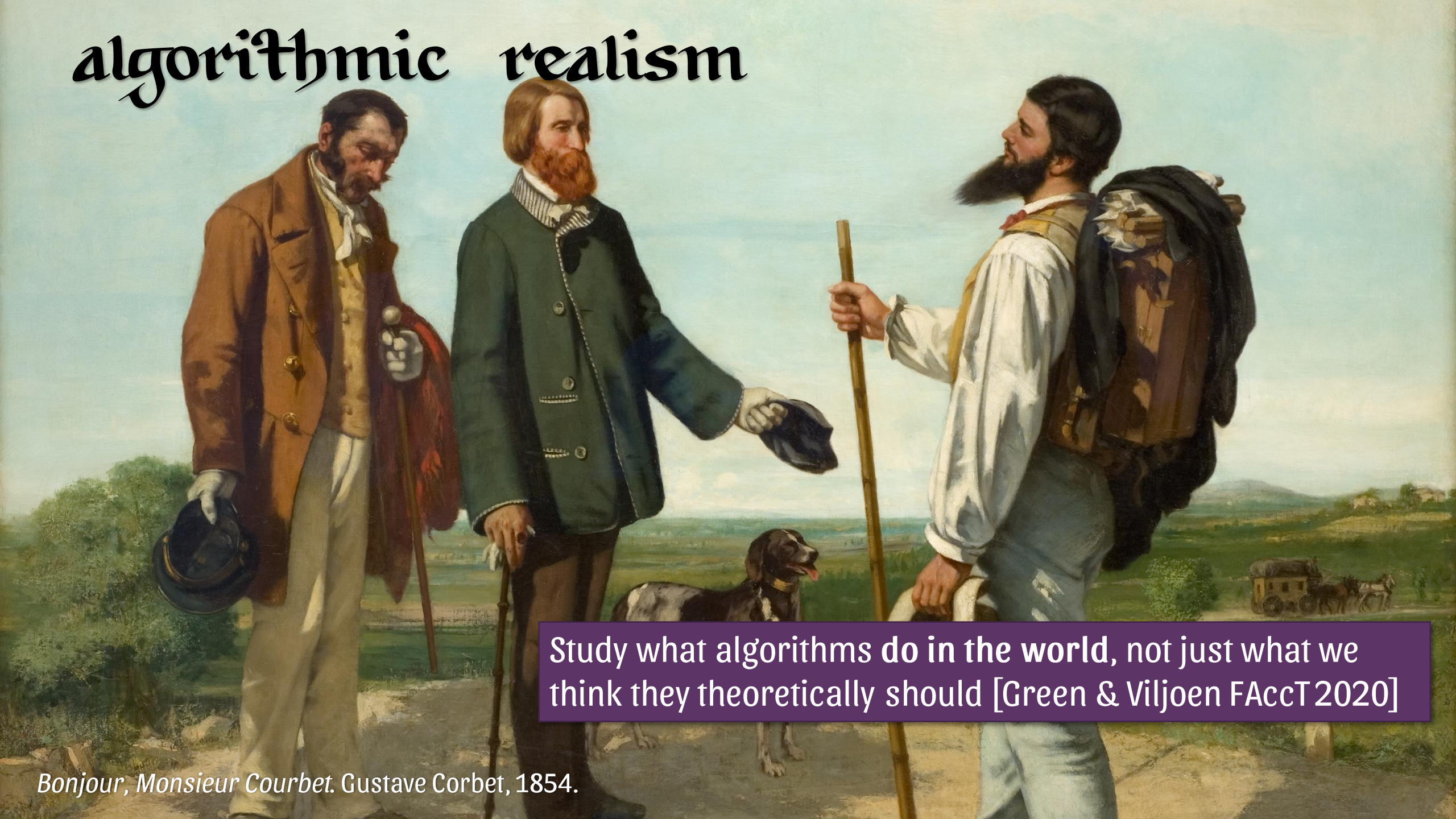
I find it useful to *separate* the **normative** and **descriptive** claims

specifying the problem

Clearly identify:

- Stakeholders
- Type and nature of harm
  - Distributional or representational?
  - Specific question
  - Subtractable or non-subtractable?
- Normative principles
- Limitations of work and methods

algorithmic realism



Study what algorithms **do in the world**, not just what we think they theoretically should [Green & Viljoen FAccT 2020]

Bonjour, Monsieur Courbet. Gustave Corbet, 1854.

hunt for monsters

Many...

... stakeholders: consumers, providers, subjects, etc.

... types of harm: inequity, misrepresentation, on many dimensions

... time scales: snapshot response, repeated over time, etc.

... groups and aggregates / subsets

... open questions on practical details

Good hunting!

thank you

- Workshop organizers for inviting me
- NSF for funding (grant IIS 17-51278)
- Students and colleagues