



ICQCM

CRITICAL DATA SCIENCE
FOR A DIVERSE WORLD

Defining Critical QCM

Momin M. Malik

Moderated by Ezekiel Dixon-Roman

Thursday, May 27th, 4 - 5:30 PM EDT

William T. Grant AQC SCHOLARS Virtual Seminar Series



Since who we are matters...

Who I am

Defining quantitative methods

Defining "critical"

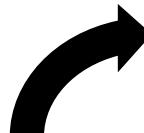
Defining QCM

Problems for QCM

Open questions

About me:

- Second-generation Pakistani-American
 - Physician father, artist/homemaker mother
 - NB spouse Maya: social work academic, Black American descendant of persons enslaved in the US
- UG:  DEPARTMENT OF THE HISTORY OF SCIENCE HARVARD UNIVERSITY
- RA:  Berkman
The Berkman Center for Internet & Society at Harvard University
- MSc:  UNIVERSITY OF OXFORD
- PhD: Carnegie Mellon University
School of Computer Science
- Post-doc:  BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY





My grounding: science studies

Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

- “The declared aim of modern science is to establish a strictly detached, objective knowledge. Any falling short of this ideal is accepted only as a temporary imperfection, which we must aim at eliminating. **But suppose that tacit thought forms an indispensable part of all knowledge, then the ideal of eliminating all personal elements of knowledge would, in effect, aim at the destruction of all knowledge.** The ideal of exact science would turn out to be fundamentally misleading and possibly a source of devastating fallacies.” (Polanyi 1966)
- See: Sismondo 2010, “Questioning functionalism in the sociology of science”

Training: Stats, ML, general modeling

Who I am

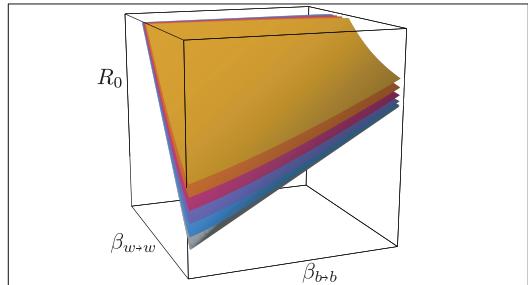
Defining quantitative methods

Defining "critical"

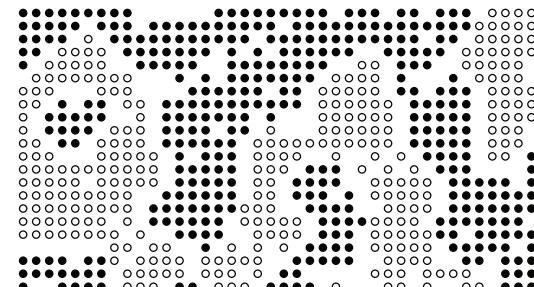
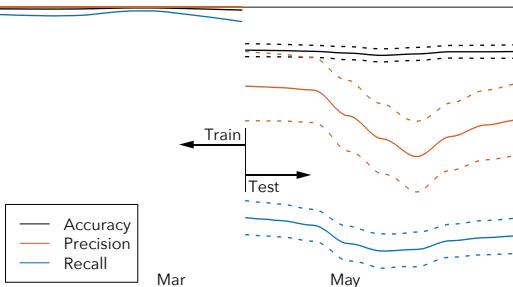
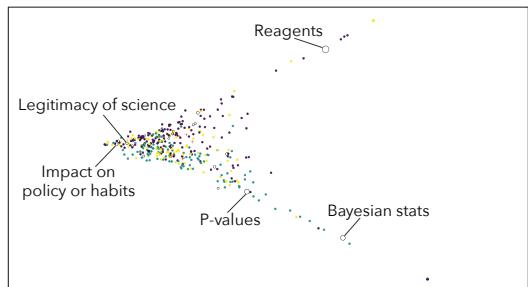
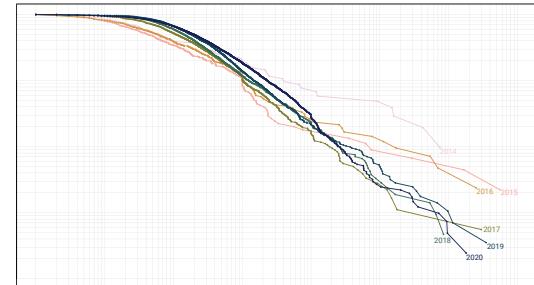
Defining QCM

Problems for QCM

Open questions



$$\begin{aligned}
 \text{Err}(\hat{\mu}) &= \frac{1}{n} \mathbb{E}_f \|Y^* - \hat{Y}\|_2^2 \\
 &= \frac{1}{n} \left[\mathbb{E}_f \|Y^*\|_2^2 + \mathbb{E}_f \|\hat{Y}\|_2^2 - 2 \mathbb{E}_f (Y^{*T} \hat{Y}) \right] \\
 &= \frac{1}{n} \left[\mathbb{E}_f \|Y^*\|_2^2 + \mathbb{E}_f \|\hat{Y}\|_2^2 - 2 \text{tr} \mathbb{E}_f (Y^* \hat{Y}^T) \right] \\
 &\quad + \frac{1}{n} \left[\mu^T \mu + \mathbb{E}_f (\hat{Y})^T \mathbb{E}_f (\hat{Y}) + 2 \text{tr} \mu \mathbb{E}_f (\hat{Y})^T \right] \\
 &\quad + \frac{1}{n} \left[-\mu^T \mu - \mathbb{E}_f (\hat{Y}) \mathbb{E}_f (\hat{Y})^T - 2 \mu^T \mathbb{E}_f (\hat{Y}) \right] \\
 &= \frac{1}{n} \left[\text{tr} \Sigma + \|\mu - \mathbb{E}(\hat{Y})\|_2^2 + \text{tr} \text{Var}_f(\hat{Y}) - 2 \text{tr} \text{Cov}_f(Y^*, \hat{Y}) \right] \\
 &= \text{irreducible error} + \text{bias}^2 + \text{variance} - \text{optimism}
 \end{aligned}$$



- Co-organizer, Workshop on Critical Data Science, ICWSM-2019 (Munich, Germany)
- Communities: ICWSM, Sunbelt, [Tapia]
- Disclaimer: I only model **social** systems. May have limited

Defining quantitative methodologies

Who I am

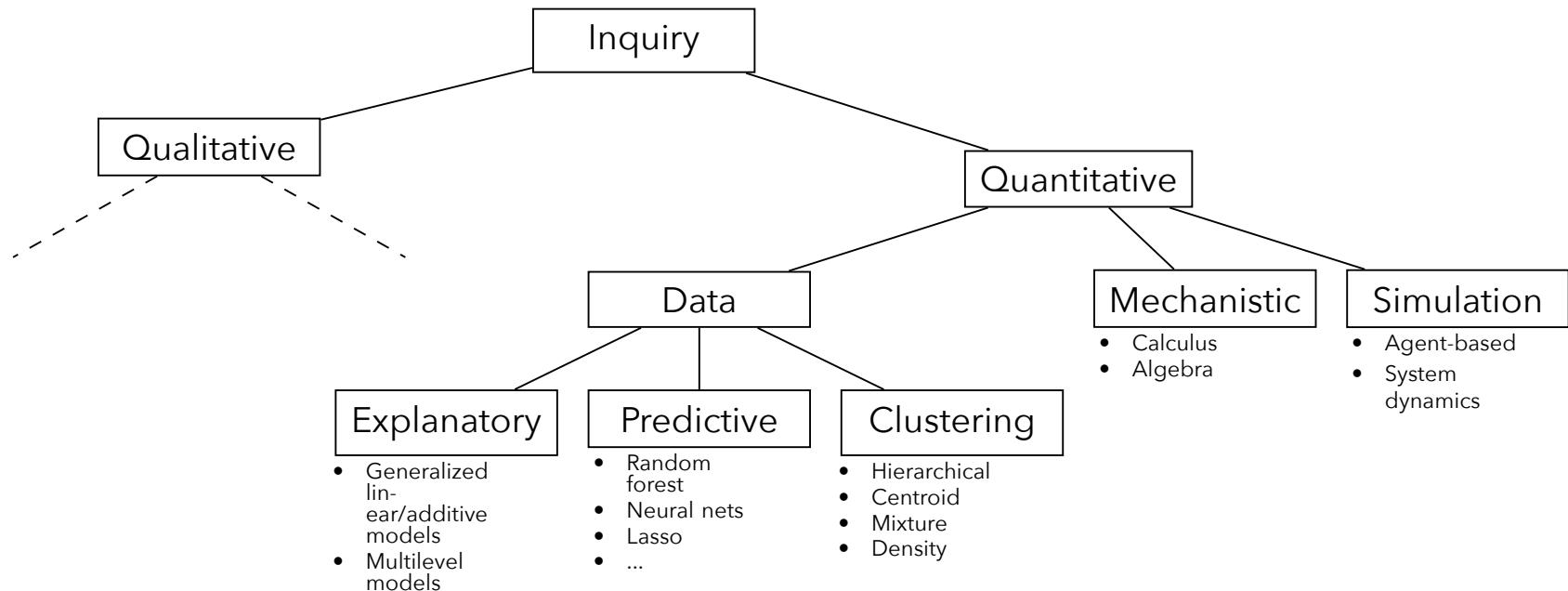
Defining
quantitative
methods

Defining
"critical"

Defining
QCM

Problems for
QCM

Open
questions





Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Defining “data science”



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an interaction of mathematics, statistics, computer sciences, communication and business sense. Finding a Data Scientist is hard. Finding people who understand what a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variational
- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Databases SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with tools like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solving
- Statistical, creative, intellectual and collaborative
- Able to engage with senior management
- Story telling skills
- Turn data driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Hive, USQL, tableau



FRANKENSTEIN'S DATA SCIENTIST

What exactly is Frankenstein's Data Scientist made up of?

EYES

Ability to see patterns in data that are not quite visible to all. Experience is key - whilst this comes with practice, a keen eye goes a long way!

HEART

An analytical mind with a need for handling large data sets and extracting valuable insights. Well versed in the ways of Hadoop, Spark, Hive and SAS

HANDS

Hands-on experience goes a long way. Whilst you may not be in a position to boast years of industrial experience yet, you can hone your skills using resources like Kaggle.

FEET

You'll need to keep up in the fast-moving technology industry, dynamic and strategic in your approach and top of current trends and breakthroughs.



BIG CLOUD
Provider Reviews In Big Data and Data Science

DATA SCIENTIST MUST-HAVE SKILLS

MATH & STATISTICS

- Machine Learning
- Statistical Modeling
- Exploratory Analysis
- Clustering
- Regression Analysis

DOMAIN KNOWLEDGE & SOFT SKILLS

- Inclination towards business operations
- Keen on working with data
- Problem solver
- Strategic, proactive, and cooperative
- Interested in hacking



PROGRAMMING & DATABASE

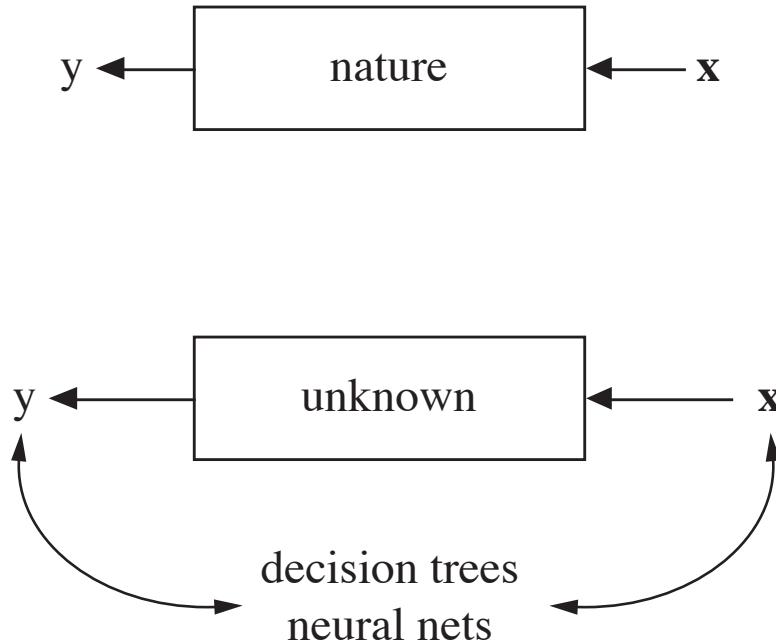
- Computer Science Fundamentals
- Database Management System
- Data Visualization
- Python
- Big Data

COMMUNICATION & VISUALIZATION

- Storytelling skills
- Convert data-based insights into decisions
- Collaborative with Sr. Management
- Knowledge of tools like Tableau
- Visual art design

- Applied statistics and applied machine learning, mostly in business

Defining machine learning



- An instrumental use of statistical correlations to *mimic* the output of a target process, rather than understand the *relationship* between inputs and outputs.
Involves finding expressions that maximize correlation.

Breiman 2001. See also Jones 2018.



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Defining “critical”

- Fay (1986): “humanist variant of estrangement theory”
 - Estrangement theory: most people live in a manifest/ordinary sphere that keeps them trapped from what is best in life, which exists in a hidden/extraordinary sphere
 - Humanist variant: locates the hidden/extraordinary sphere in the social plane, not religious or spiritual
- Components:
 - Theory of false consciousness
 - Theory of crisis
 - Theory of education
 - Theory of transformative action
- Examples: Critical race theory, Afropessimism, feminist theory, Marxism



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Related to constructivism

- Social construction of [scientific] knowledge: the idea that knowledge and structure is at most *constrained* by nature, not determined by it
- All scientific knowledge comes to be through a social process, and could have been different
- Sometimes, knowledge of
 - the constructed nature of categories (e.g., race, gender, class)
 - the historical process of that construction
 - the political process of organizing society along those constructed categoriesis the “hidden/extraordinary sphere”.
- Critique and constructivism can be in conflict, if critique is not *reflexive* (the critique can understand that it is itself constructed)



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Contrast with realism and positivism

- Realism: there is an underlying “true” reality that exists prior to and independent of our conception of it
- Positivism: the only meaningful knowledge is that which can be empirically verified and demonstrated
- Both are “scientistic”, insisting on the possibility and superiority of universal, objective, neutral knowledge, but differ in their metaphysical commitment



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Must quant be realist or positivist?

- Statistics: most principled and well-developed of quantitative methods in social research, in that it has a theory of how the world is (although one based on mathematical convenience rather than conviction)
- (Machine learning inherits its foundations from statistics, but takes an instrumental approach that is positivist)
- Even the mainstream of Bayesian statistics is realist (there is an objective reality; subjective beliefs just help us get there faster)

Problem especially in social science

Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

	Issue	Positivism	Postpositivism	Critical theory et al.	Constructivism	Participatory
Defining quantitative methods	Ontology	Naïve realism—"real" reality but apprehensible	Critical realism—"real" reality but only imperfectly and probabilistically apprehensible	Historical realism—virtual reality shaped by social, political, cultural, economic, ethnic, and gender values; crystallized over time	Relativism—local and specific co-constructed realities	Participative reality—subjective-objective reality, cocreated by mind and given cosmos
Defining "critical"	Epistemology	Dualist/objectivist; findings true	Modified dualist/objectivist; critical tradition/community; findings probable true	Transactional/subjectivist; value-mediated findings	Transactional/subjectivist; co-created findings	Critical subjectivity in participatory transaction with cosmos; extended epistemology of experimental, propositional, and practical knowing; cocreated findings
Problems for QCM	Methodology	Experimental/manipulative; verification of hypotheses; chiefly quantitative methods	Modified experimental/manipulative; critical multiplism; falsification of hypotheses; may include qualitative methods	Dialogic/dialectical	Hermeneutical/dialectical	Political participation in collaborative action inquiry; primacy of the practical; use of language grounded in shared experimental context

"Basic beliefs (metaphysics) of alternative inquiry paradigms" (Guba & Lincoln 2005)



"Understanding a person..." (from Barbara Kiviat)

Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

	As a case	In narrative
Context/ circumstance	Stripped away	Key
Mental states	Absent (for the most part)	Crucial; constitutive
Relevant features	Determined in advance	Emergent
Orientation to time	Atemporal	Chronological
Ordering of features	Unimportant	Meaningful
Other actors	Invisible	Often present
Causal logic	Mathematical	Theoretical
To boost predictive validity	Add cases	Know person better

"Bowker and Star 2000; Bruner 1986; Desrosières 1998; Espeland 1998; Espeland and Stevens 1998, 2008; Fourcade and Healy 2017; Hacking 1990; Porter 1994, 1995; Ricouer 1998; White 1980, 1984". I would add: Patton 2005; Abbott 1988



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Who should get to define QCM?

- Disclaimer: probably not me
- Desiderata:
 - Lived experience
 - Critical theorist
 - Advanced technical training with strong foundations
- Collective
- ...but I'll offer my thoughts



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

My proposed definition

- QCM: *The use of quantification and mathematical modeling (e.g., mechanistic, statistical, "algorithmic", simulation) within a critical and constructivist framework that understands quantification and modeling as social, situated, contingent, and "productive" (often towards harm), not natural, universal, inevitable, or neutral.*
- I am offering something more limited, rather than comprehensive and aspirational. This does not require action, or a specific ethical stance, or a specific theory of power or change



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Further details of my conception

- Incorporating ready-made quant methods into a critical approach is okay
- But more interesting is integrating the logic of modeling with the logic of critical theory at a fundamental level
- Much harder—requiring dual training—but a rich intellectual project



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Why QCM?

- Strategic quantification/strategic modeling ("Counterhegemonic modeling") to demonstrate inequality?
- Rhetorical use: convert power-brokers?
- Alternatively: just because quantification is currently associated with power does not mean it is essentially so. Qualitative inquiry can be just as or more oppressive, it just isn't currently in power



Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions

Why *not* QCM?

- Danger of co-option: using qual + theory to “improve” quant, rather than challenging hierarchies of knowledge and overturning power relations and
- Why give credence to tools of oppression?
- Does it even work?
 - Modeling is an opportunistically used tool of power, not the source of power
 - Who really listens to evidence and modeling? E.g., quantitative work of Ida B. Wells-Barnett and W. E. B. DuBois (Benjamin 2018)

Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

**Problems
for QCM**

Open questions

Why *not* QCM?



"The function, the very serious function of racism is distraction. It keeps you from doing your work. It keeps you explaining, over and over again, your reason for being. Somebody says you have no language and you spend twenty years proving that you do. Somebody says your head isn't shaped properly so you have scientists working on the fact that it is. Somebody says you have no art, so you dredge that up. Somebody says you have no kingdoms, so you dredge that up. None of this is necessary. There will always be one more thing." (Morrison 1975)

Who I am

Defining
quantitative
methods

Defining
"critical"

Defining
QCM

**Problems
for QCM**

Open
questions



Why not QCM?

"The starving fellah, (or the jobless inner city N.H.I., the global New Poor or *les damnés*), Fanon pointed out, does not have to *inquire into the truth*. He *is*, they are, the Truth. It is we who constitute this "Truth." We must now undo their narratively condemned status." (Wynter 1994)



Who I am

Defining
quantitative
methods

Defining
“critical”

Defining
QCM

Problems
for QCM

Open
questions

Examples of QCM?

- Wallace & Wallace, 1999, "Emerging infections and nested martingales: the entrainment of affluent populations into the disease ecology of marginalization"?
- Work of David Freedman?
- Virginia Tech Professor Marc Edwards, with Flint water crisis?
- My attempts:
 - Malik and Pfeffer, 2016, "Platform effects in social media data"
 - Malik, 2020, "A hierarchy of limitations in machine learning"
 - Richardson, Malik, Darity, Mullen, Morse, Malik, Benton, Bassett, Farmer, Worden, Jones, 2021, "Reparations for Black American descendants of persons enslaved in the U.S. and their potential impact on SARS-CoV-2 transmission"

Reparations and COVID-19 paper

Who I am

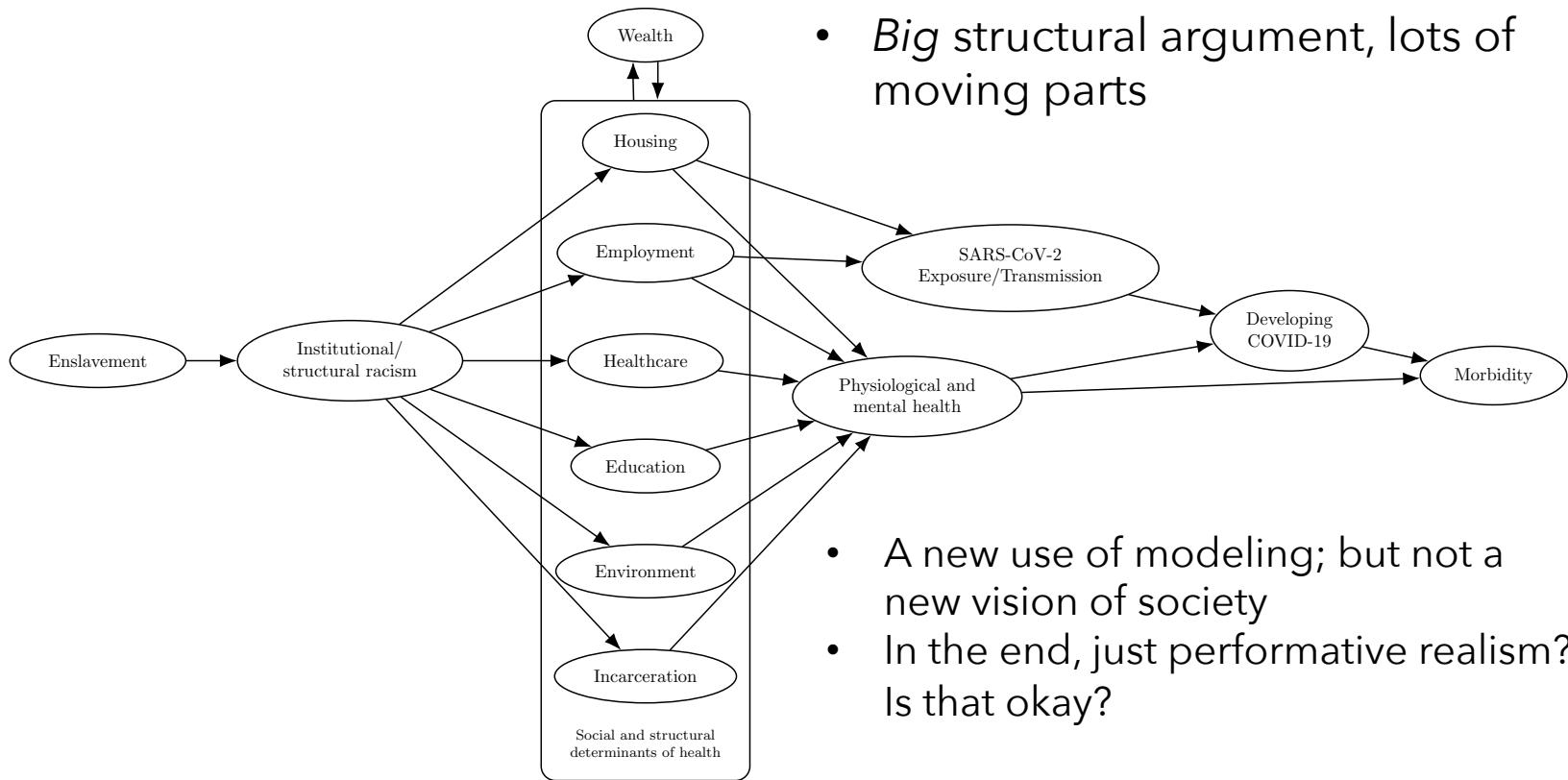
Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

Open questions





Who I am

Defining quantitative methods

Defining "critical"

Defining QCM

Problems for QCM

**Open
questions**

Open questions

- What should we try to accomplish?
- What should be included?
- How should we institutionalize training?
 - We can't expect people to get training in multiple entirely different ways of thinking
 - We should find ways to first induct people into critical ways of thinking, and then find a palatable way of teaching quant methods *after*, rather than have to undo positivism/realism



ICQCM



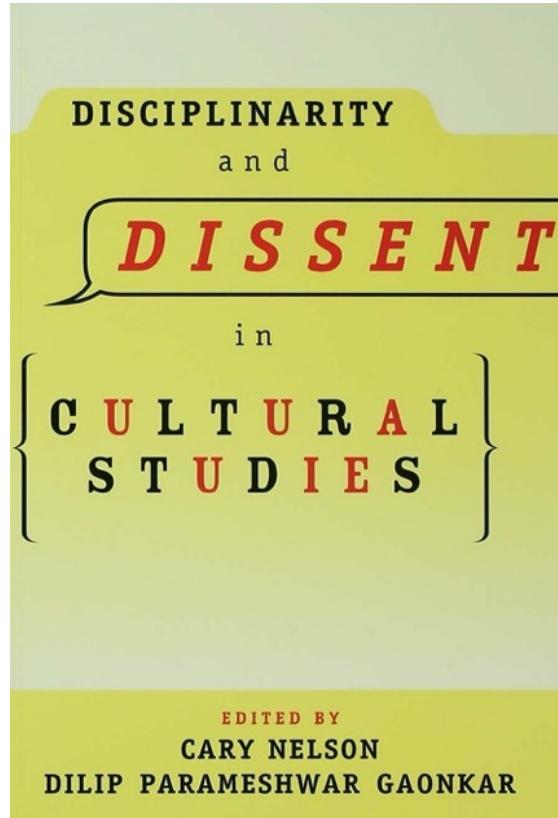
ICQCM

Backup slides

Symmetry?

- "...cultural studies have maintained a hermeneutics of suspicion toward the methods of quantification. But, to what extent does this suspicion toward quantitative inquiry compromise the deconstructive project of cultural studies by falling into the trap of the quantitative/qualitative and, related, nature/culture binaries?" (Dixon-Román 2016)
- "Why, in a thick world, do economists stride with heads high through the corridors of power, while cultural historians pass along their possibly profound insights to one another? ...we are dealing here not merely with a choice between alternative intellectual strategies of interpretation and analysis, but with practical ways of knowing the world and of shaping action." (Porter 2012)

Does *intolerance* cut both ways?

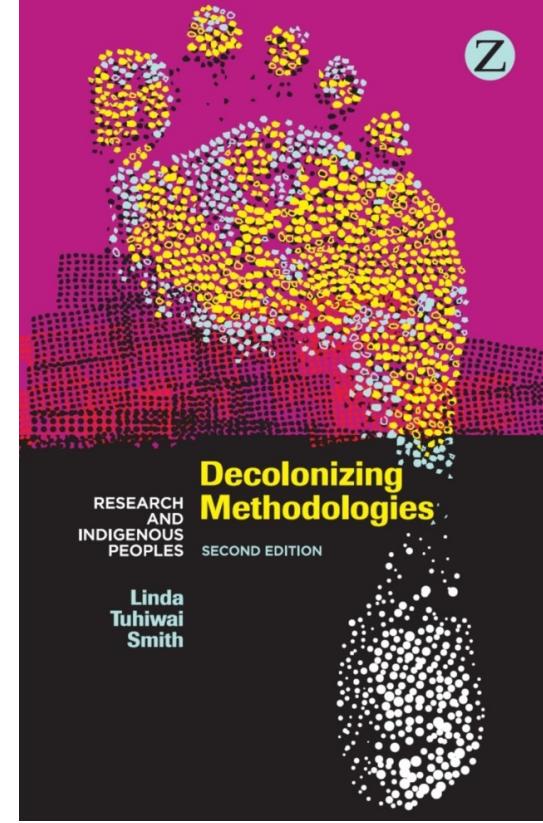


"But that is how disciplines police their boundaries, by training their members to internalize them, naturalize them, and then fancy themselves free as birds. In some sectors of the culture... the unthinkable and undoable could be named and cast out when necessary. One of those sectors was the academy, and its disciplines would discipline unruliness whenever it arose."

(Nelson and Gaonkar 1996)

Qual not intrinsically better

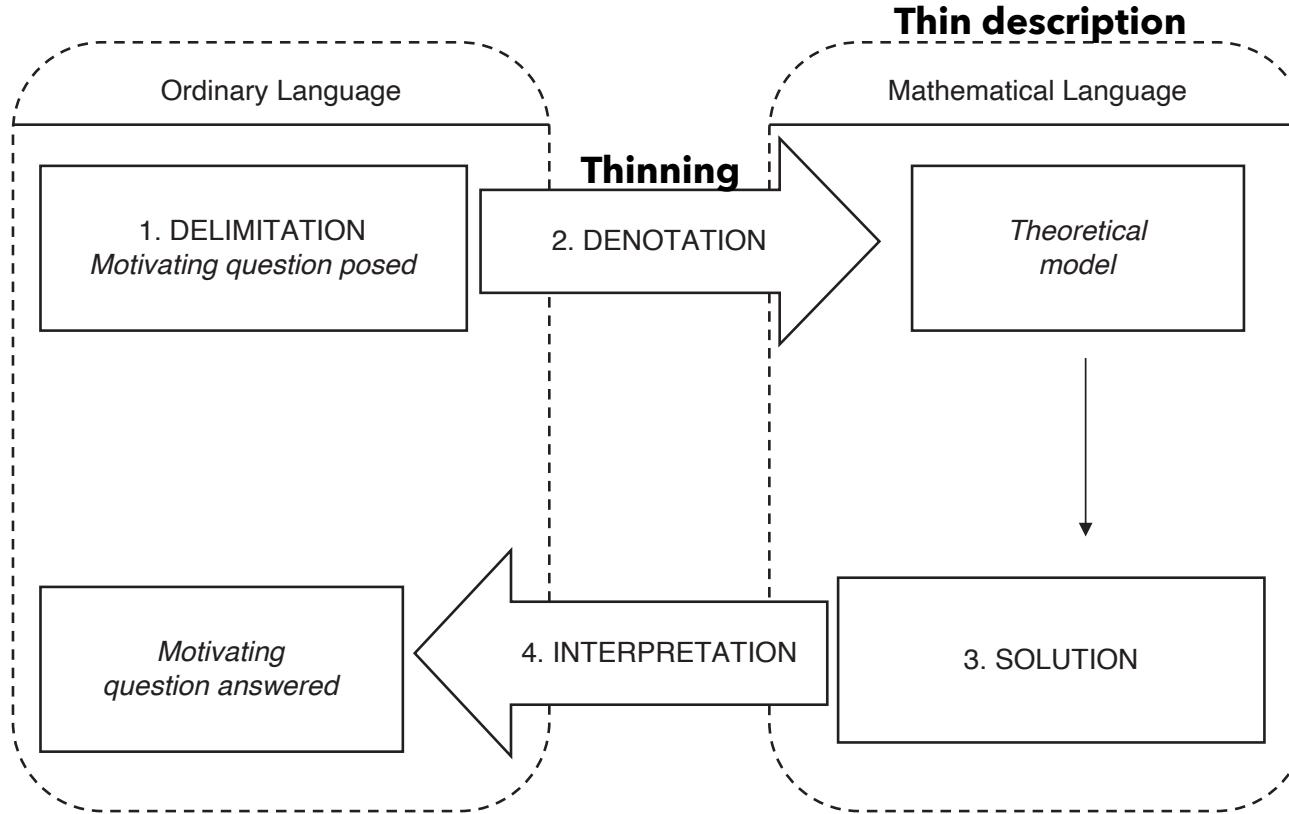
"we are suggesting that anthropological analyses (of pain and passion and power), when they are experience-distant, are at risk of delegitimizing their subject matter's human conditions. The anthropologist thereby constitutes a false subject; she can engage in a professional discourse every bit as dehumanizing as that of colleagues who unreflectively draw upon the tropes of biomedicine or behaviorism to create their subject matter." (Kleinman and Kleinman 1991; also, Smith 2012 →)



Core incompatibility

"...it is striking how absolutely these assumptions (of linear models) contradict those of the major theoretical traditions of sociology. Symbolic interactionism rejects the assumption of fixed entities and makes the meaning of a given occurrence depend on its location... Both the Marxian and Weberian traditions deny explicitly that a given property of a social actor has one and only one set of causal implications... all approach social causality in terms of stories, rather than in terms of variable attributes." (Abbott 1988)

Any thinning flattens meanings



Sticking points

- How can we dissolve such a powerful binary?
- Related: the actor/analyst (emic/etic) binary (Jardine 2004; Collins 2008)
 - (lateral anthropology [Candea 2018, 2019; Gad and Jensen 2016] flattens how actor and analyst relate, but still has this binary)