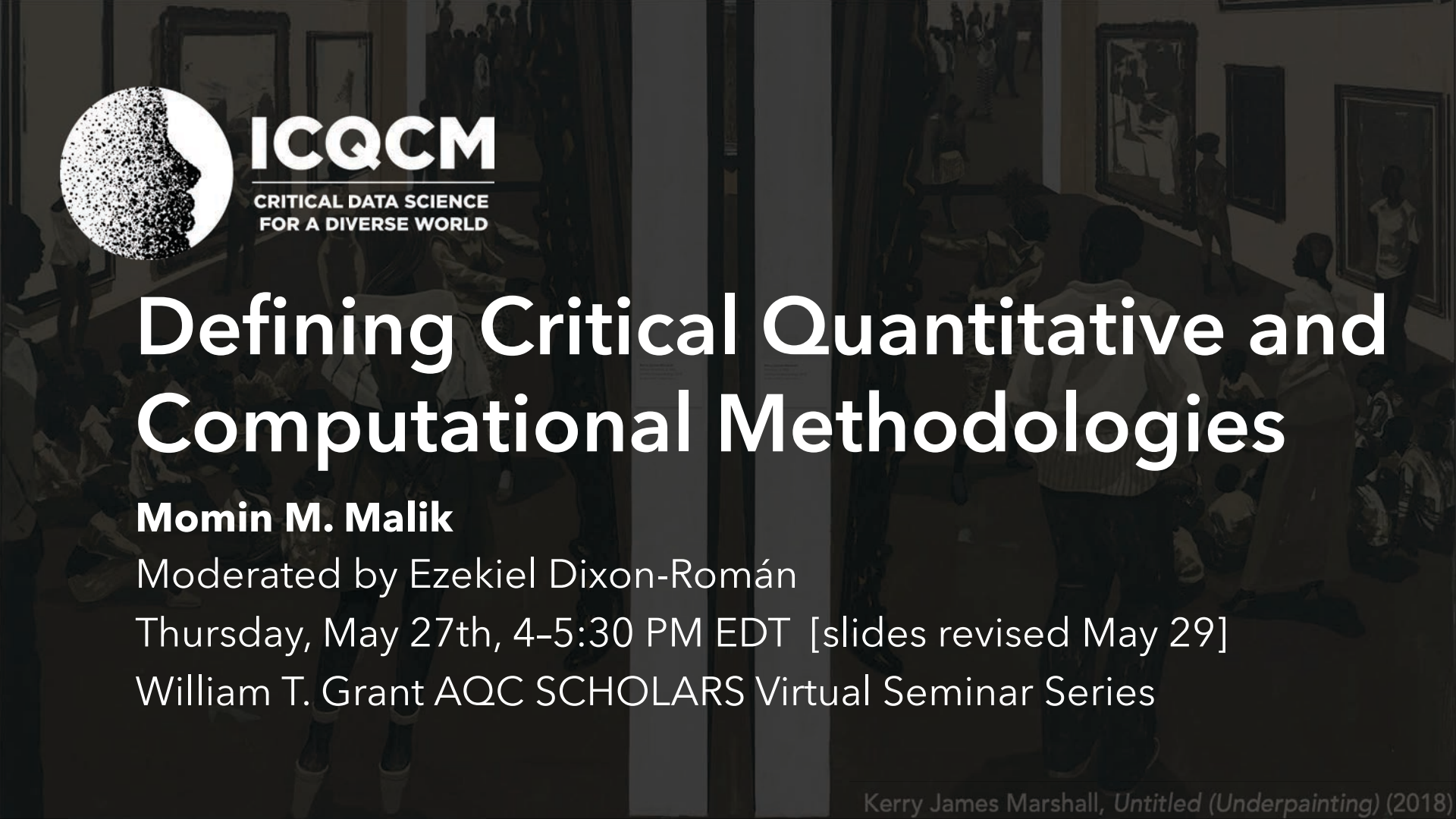


Defining Critical Quantitative and Computational Methodologies

Momin M. Malik

Moderated by Ezekiel Dixon-Román

Thursday, May 27th, 4–5:30 PM EDT [slides revised May 29]

William T. Grant AQC SCHOLARS Virtual Seminar Series

Kerry James Marshall, *Untitled (Underpainting)* (2018)

Since who we are matters…

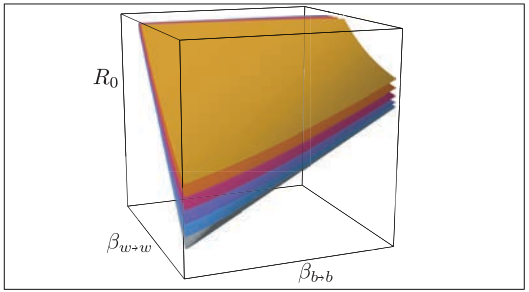
About me:

- 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: Oxford Internet Institute; University of Oxford
- PhD: Carnegie Mellon University, School of Computer Science
- Post-doc: Berkman Klein Center for Internet & Society at Harvard University

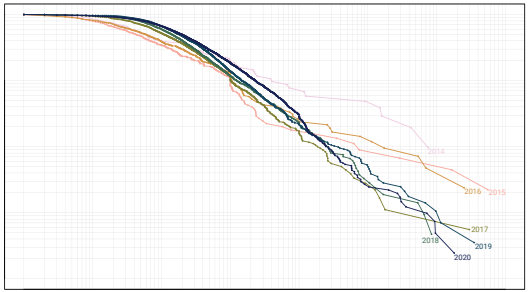
My grounding: science studies

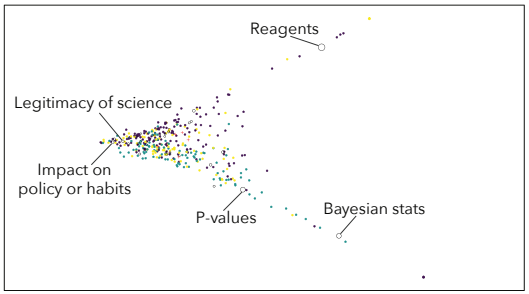
- "The declared aim of modern science is to establish a strictly detached, objective knowledge. Any falling short of this ideal is accepted only as a temporary imperfection, which we must aim at eliminating. **But suppose that tacit thought forms an indispensable part of all knowledge, then the ideal 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"

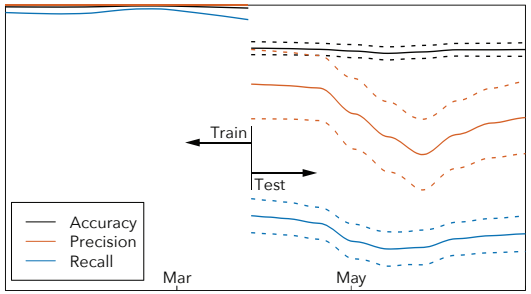
Training: Stats, ML, general modeling

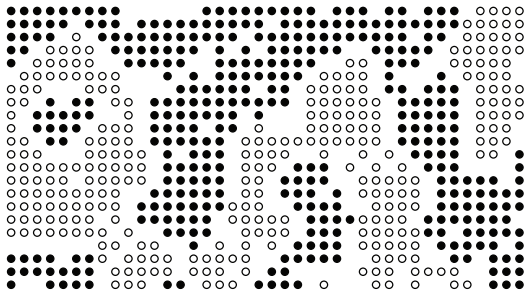


$$
\begin{aligned}
\mathrm{Err}(\hat{\mu}) &= \tfrac{1}{n}\mathbb{E}_f\|Y^* - \widehat{Y}\|_2^2 \\
&= \tfrac{1}{n}\left[\mathbb{E}_f\|Y^*\|_2^2 + \mathbb{E}_f\|\widehat{Y}\|_2^2 - 2\mathbb{E}_f(Y^{*T}\widehat{Y})\right] \\
&= \tfrac{1}{n}\left[\mathbb{E}_f\|Y^*\|_2^2 + \mathbb{E}_f\|\widehat{Y}\|_2^2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



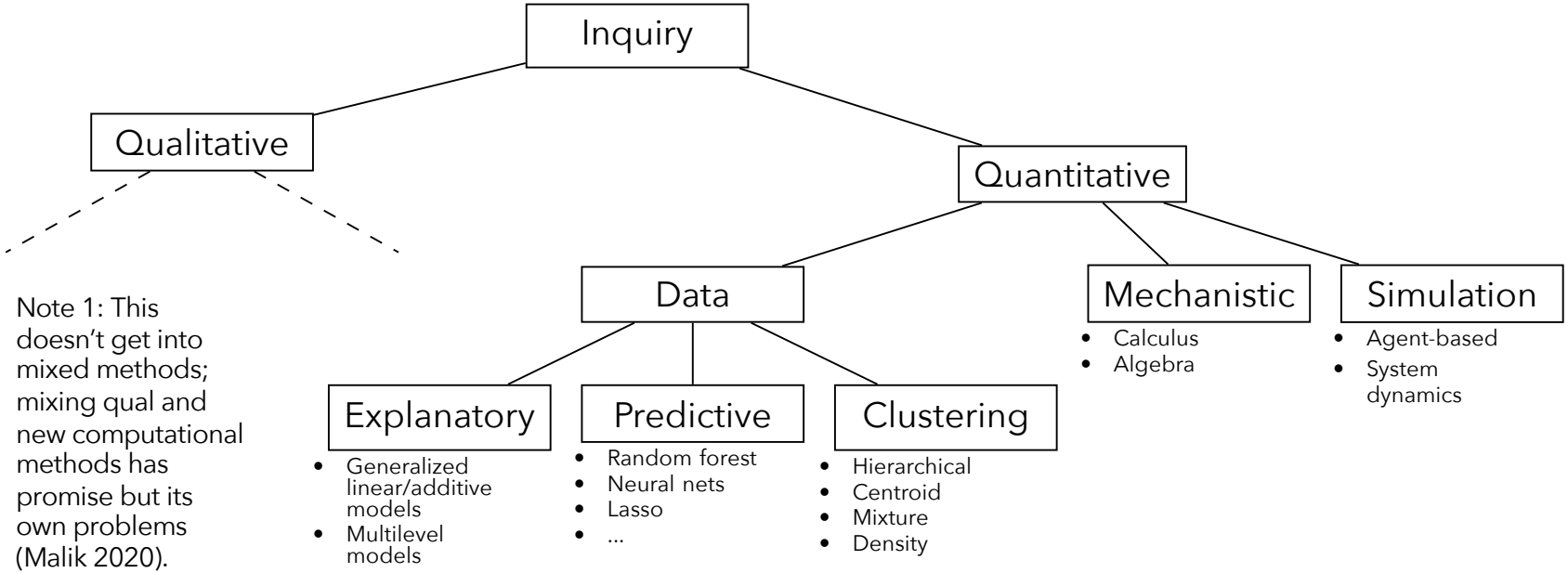






- Co-organizer, Workshop on Critical Data Science, ICWSM-2019 (Mayer and Malik 2019)
- Communities: ICWSM, Sunbelt[, Tapia]
- Disclaimer: I model **social** systems. What I offer may have limited relevance to chem, bio

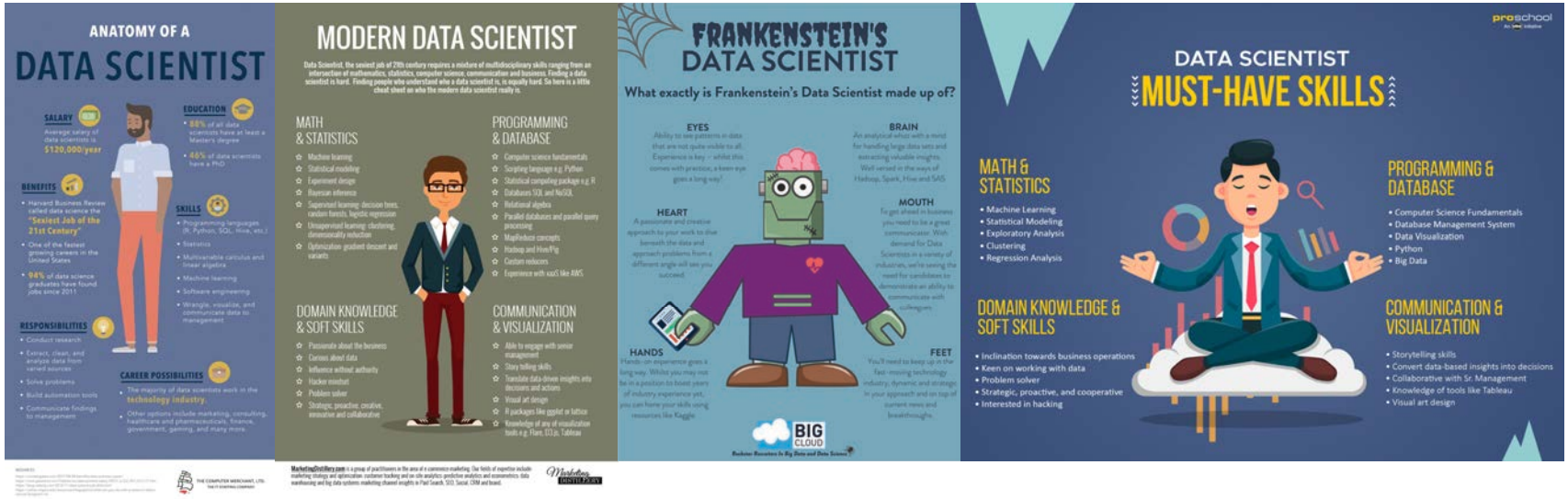
Defining quantitative methodologies



Note 1: This doesn't get into mixed methods; mixing qual and new computational methods has promise but its own problems (Malik 2020).

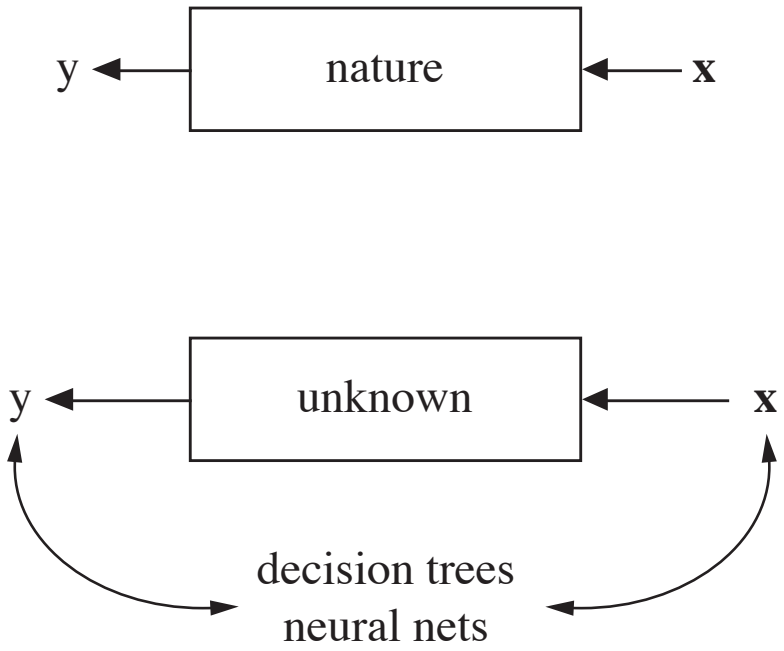
Note 2: I am not separately defining computational methods, since statistical modeling is now inseparable from computation (see Efron and Hastie 2016), machine learning/datamining/"predictive" modeling came out of computational capacity (see Breiman 2001; Friedman 1997), and simulation is *almost* by nature computational (see Pfeffer and Malik 2017). Even mechanistic modeling can be aided by Mathematica. I don't know all the literature theorizing this convergence, but for a great philosophical work on computation in natural sciences around simulation, see Winsberg (2010).

Defining "data science"



- Applied statistics and applied machine learning, mostly in business
- I say: fine, let business have it

Defining machine learning



My definition: 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.

Defining "critical"

- Fay (1987): "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

Related to *constructivism*

- Social construction of [scientific] knowledge: the idea that knowledge and structure is at most *constrained* by nature, not determined by it
  - Applied to science: hated by many scientists; subject of 1990s "Science Wars" (good retrospective: Labinger and Collins 2001)
- All [scientific] knowledge comes to be through a social process, and could have been different
- Sometimes, the "hidden/extraordinary sphere" is 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 categories
- Critique and constructivism can be in conflict, if critique is not *reflexive* (the critique can understand that it is itself constructed)

Contrast with realism, 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" (Payne and Payne 2004), insisting on the possibility and superiority of universal, objective, neutral knowledge, but differ in their metaphysical commitment

Who should get to define critical QCM?

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

My proposed definition

- Critical 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. E.g., this definition does not require action, or a specific ethical stance, or a specific theory of power or change

Further details of my conception

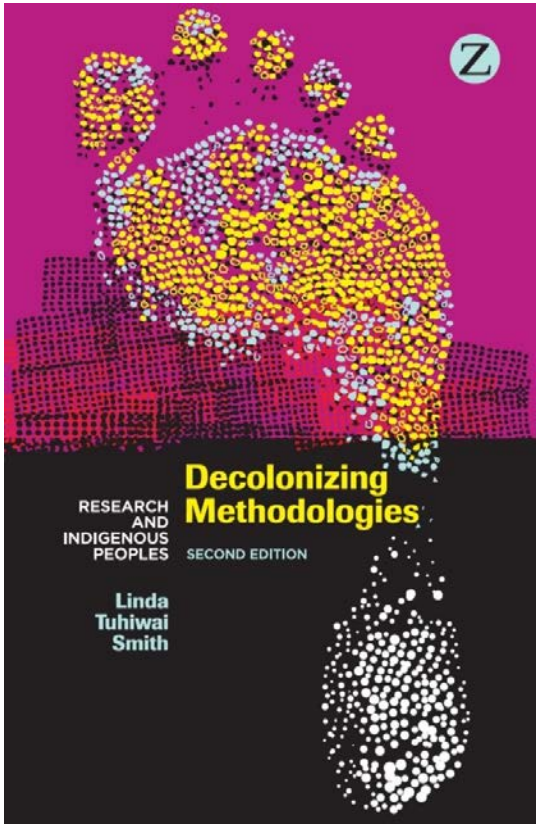
- Incorporating ready-made quant methods into a critical approach is okay
  - "Minimal" critical QCM: quant demonstrations of disparities *that links to theory* about the source of those disparities (e.g., white supremacy, dehumanization)
- But more intellectually interesting for me is integrating the logic of modeling with the logic of critical theory at a fundamental level
- Much harder—requiring dual training—but a rich intellectual project
  - On the other hand, maybe useless practically, and the "minimal" version of critical QCM is most useful

Why critical QCM?

- Strategic quantification/strategic modeling (after Spivak's "strategic essentialism") to demonstrate inequality?
  - Rhetorical use: convince power-brokers?
- "Counterhegemonic modeling" (Richardson 2020): modeling ironically to reveal the absurdity of modeling?
- 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

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 delegitimating 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, Tuhiwai Smith 2012 →)



Problem: Quant is realist or positivist

- Statistics: most principled and well-developed of quantitative methods in social research, in that it has a realist 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; see Jones 2018)
- Even the mainstream of Bayesian statistics is arguably realist (subjective beliefs are how we get to objective reality; Gelman and Hennig 2017)
- While simulation modeling can imagine alternative worlds, and in that depart from realism/positivism, I remain skeptical about its overall value (Pfeffer and Malik 2017; Malik 2020)

Social inquiry: Requires aligning methodology with basic beliefs

Issue	Positivism	Postpositivism	Critical theory et al.	Constructivism	Participatory
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
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
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 and Lincoln 2005)

"Understanding a person…"

(slide from Barbara Kiviat)

"

	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

The problem: quant can only understand people as cases, which is incompatible with critical perspectives

"

"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

Problem: 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)

Problem: "Thinning" flattens meanings

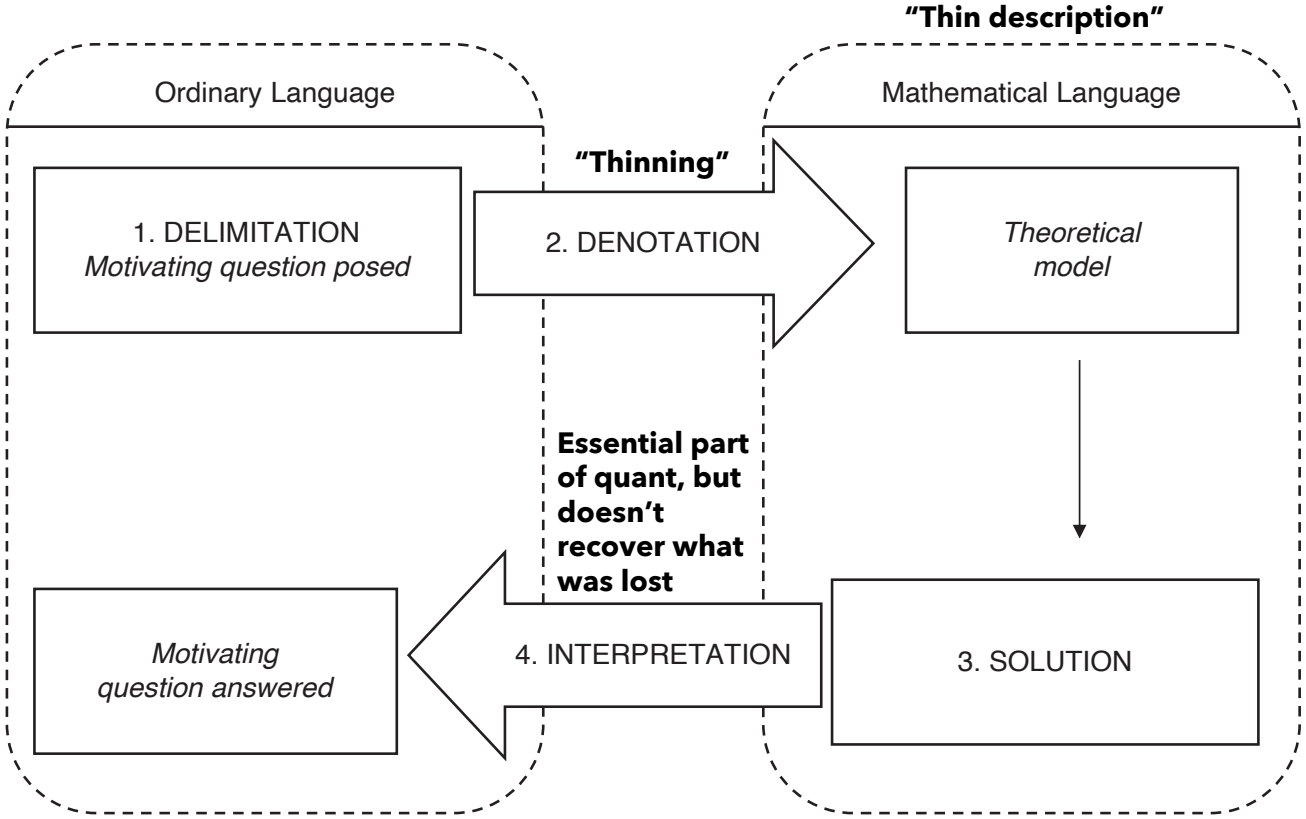
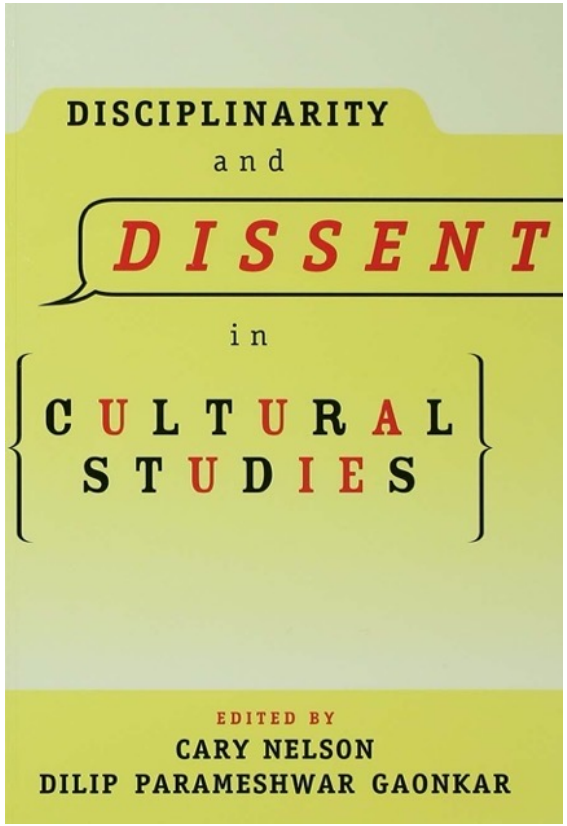


Diagram from Spiegler (2015); science (and, I would say, modeling) as "thin" description from Porter (2012)

Counterpoint: Maybe incompatibility is constructed



"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)

"…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)

Ethical problems with critical QCM

- Danger of co-option: using qual + theory to "improve" quant, rather than challenging hierarchies of knowledge (where quant is above qual) and overturning power relations
- 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 2019)

Ethical problem: Quant as distraction

"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)

Whose knowledge do we value?

"The starving fellah, (or the jobless inner city N.H.I., the global New Poor or *les dannés*), 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)

Examples of critical QCM?

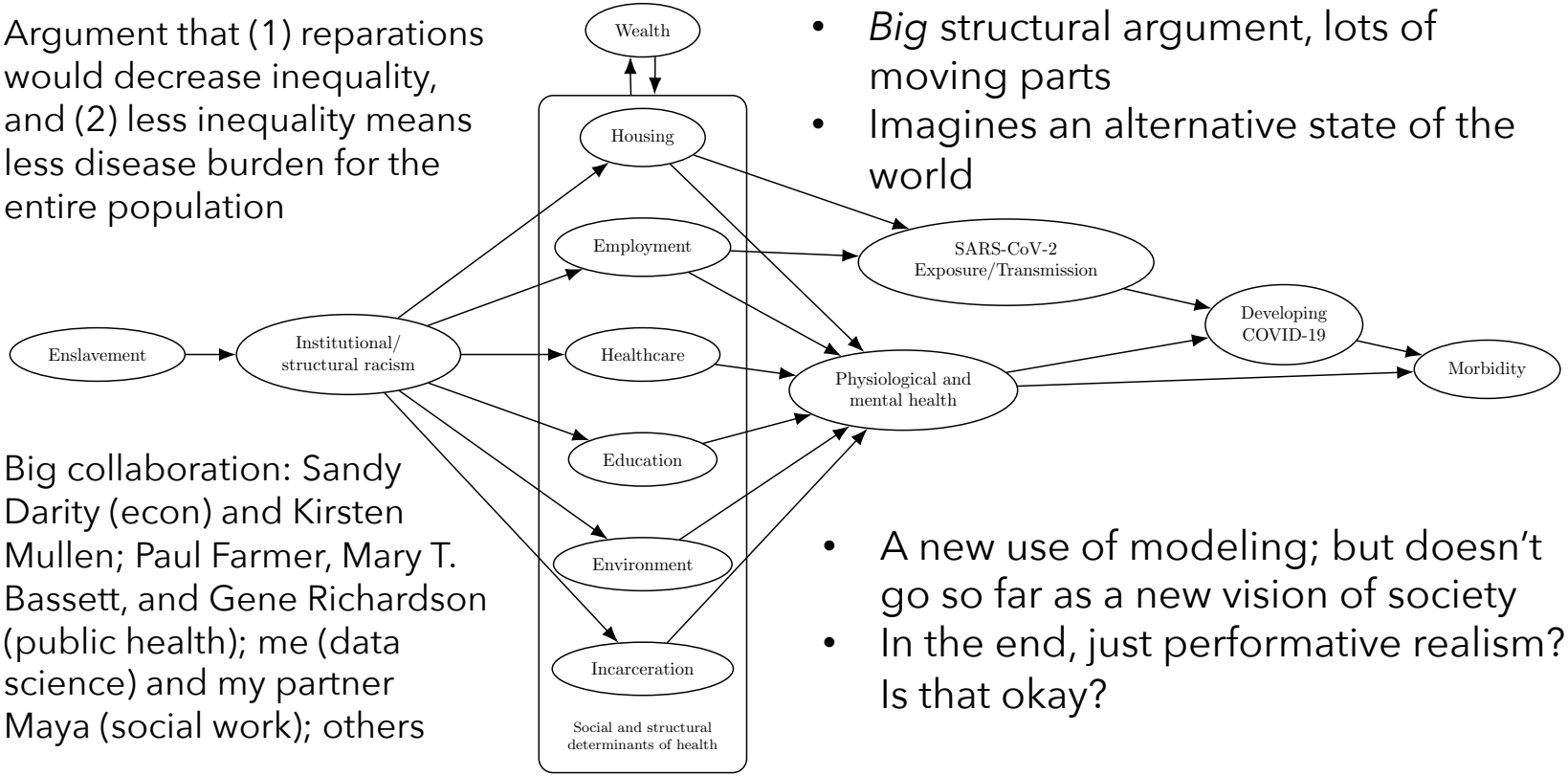
- Gender Shades (Buolamwini and Gebru 2018). Advanced quant (the trained models were neural nets, although the measurement of difference was straightforward), critical (connects to critical race theory), and effected change. That and Joy and Timnit's other work (together and individually) are a slam dunk example for what we might want to see; but are there any others?
- Wallace and Wallace, 1999, "Emerging Infections and Nested Martingales: The Entrainment of Affluent Populations into the Disease Ecology of Marginalization." Neat, but do things like this actually accomplish anything?
- Work of David Freedman? (posthumous collection: Freedman 2010)
- Non-social data example: Virginia Tech Professor Marc Edwards, with Flint water crisis? White chemist using science to "prove" Black people's suffering
- Gelman et al. 2007, "An Analysis of the New York City Police Department's 'Stop-and-Frisk' Policy in the Context of Claims of Racial Bias." Did *this* actually convince anybody? If so, who, and why weren't they convinced by descriptions of experiences? Why did the NYPD give *this* group their data?

Examples of critical QCM?

- My attempts:
  - Malik and Pfeffer, 2016, "Platform Effects in Social Media Data": uses causal inference to show how Facebook's People You May Know creates the phenomenon (triadic closure) it assumes. Quantitatively shows "performativity" (Healy 2015), but is the 'weak' sense of performativity
  - Malik, 2020, "A Hierarchy of Limitations in Machine Learning." I'm tremendously proud of it, but it's ultimately just a review
  - 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" (see next slide)

Reparations and COVID-19 paper

Argument that (1) reparations would decrease inequality, and (2) less inequality means less disease burden for the entire population



- *Big* structural argument, lots of moving parts
- Imagines an alternative state of the world

Big collaboration: Sandy Darity (econ) and Kirsten Mullen; Paul Farmer, Mary T. Bassett, and Gene Richardson (public health); me (data science) and my partner Maya (social work); others

- A new use of modeling; but doesn't go so far as a new vision of society
- In the end, just performative realism? Is that okay?

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. Or do at the same time
    - I like Matt Jones and Chris Wiggins' course at Columbia, *Data: Past Present and Future,* for juxtaposing quantitative and critical training but want to see whole curricula, courses of study

References (1 of 2)

Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6 (2): 169–186. https://doi.org/10.2307/202114

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity.

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science* 16 (3): 199–231. https://doi.org/10.1214/ss/1009213726

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Proceedings of Machine Learning Research 81, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

Dixon-Román, Ezekiel J. 2016. "Diffractive Possibilities: Cultural Studies and Quantification." *Transforming Anthropology* 24 (2): 157–167. https://doi.org/10.1111/traa.12074

Efron, Bradley, and Trevor Hastie. 2016. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press. https://web.stanford.edu/~hastie/CASI/

Fay, Brian. 1987. *Critical Social Science: Liberation and its Limits*. Cornell University Press.

Freedman, David A. 2010. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Cambridge University Press.

Friedman, Jerome H. 1997. "Data Mining and Statistics: What's the Connection?" *Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics*.

Gelman, Andrew, and Christian Hennig. 2017. "Beyond Subjective and Objective in Statistics." Journal of the Royal Statistical Society, Series A 180 (4): 967–1033. https://doi.org/10.1111/rssa.12276

Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 2007. "An Analysis of the New York City Police Department's 'Stop-and-Frisk' Policy in the Context of Claims of Racial Bias." Journal of the American Statistical Association 102 (479): 813–823. https://doi.org/10.1198/016214506000001040

Guba, Egon G., and Yvonna S. Lincoln. 2005. "Paradigmatic Controversies, Contradictions, and Emerging Confluences." In *The SAGE Handbook of Qualitative Research*, edited by Norman K. Denzin and Yvonna S. Lincoln, 191–215. 5th ed. SAGE.

Healy, Kieran. 2015. "The Performativity of Networks." *European Journal of Sociology* 56 (2): 175–205. https://doi.org/10.1017/S0003975615000107

Jones, Matthew L. 2018. "How We Became Instrumentalists (Again): Data Positivism since World War II." *Historical Studies in the Natural Sciences* 48 (5): 673–684. https://doi.org/10.1525/hsns.2018.48.5.673

Kleinman, Arthur, and Joan Kleinman. 1991. "Suffering and its Professional Transformation: Towards an Ethnography of Interpersonal Experience." *Culture, Medicine and Psychiatry* 15 (3): 275-301. https://doi.org/10.1007/BF00046540

Labinger, Jay A., and Harry Collins, eds. 2001. *The One Culture?: A Conversation about Science*. University of Chicago Press.

Malik, Momin M. 2020. "A Hierarchy of Limitations in Machine Learning." https://www.arxiv.org/abs/2002.05193

References (2 of 2)

Malik, Momin M., and Jürgen Pfeffer. 2016. "Identifying Platform Effects in Social Media Data." In *Proceedings of the Tenth International AAAI Conference on Web and Social Media* (ICWSM-16), 241–249. AAAI Press. http://mominmalik.com/malik_chapter2.pdf

Mayer, Katja, and Momin M. Malik. 2019. *Critical Data Scientists at Work: Summary report of the ICWSM-2019 Workshop on Critical Data Science*. https://critical-data-science.github.io/

Morrison, Toni. 1975. A humanist view, 30 May. Black Studies Center public dialogue, Pt. 2. Portland State University, Oregon Public Speakers Collection. https://mackenzian.com/blog/2014/07/07/transcript-morrison-1975/.

Nelson, Cary, and Gaonkar, Dilip Parameshwar. 1996. "Cultural Studies and the Politics of Disciplinarity: An Introduction." In *Disciplinarity and dissent in Cultural Studies*, edited by Cary Nelson and Dilip Parameshwar Gaonkar, 1–22. Routledge.

Patton, Michael Quinn. 2015. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. 4th ed. SAGE.

Payne, Geoff, and Judy Payne. 2004. "Positivism and Realism." In *Key Concepts in Social Research*, 171–174. SAGE, 2004. https://doi.org/10.4135/9781849209397

Pfeffer, Jürgen, and Momin M. Malik. 2017. "Simulating the Dynamics of Socio-Economic Systems." In Networked Governance: New Research Perspectives , edited by Betina Hollstein, Wenzel Matiaske, and Kai-Uwe Schnapp, 143–161. Springer. https://doi.org/10.1007/978-3-319-50386-8_9

Polanyi, Michael. 1966. *The Tacit Dimension*. Doubleday.

Porter, Theodore M. 2012. "Thin Description: Surface and Depth in Science and Science Studies." *Osiris* 27 (1): 209–226. https://doi.org/10.1086/667828

Richardson, Eugene T., Momin M. Malik, William A. Darity, Jr., A. Kirsten Mullen, Michelle E. Morse, Maya Malik, Adia Benton, Mary T. Bassett, Paul E. Farmer, Lee Worden, and James Holland Jones. 2021. "Reparations for Black American Descendants of Persons Enslaved in the U.S. and their Potential Impact on SARS-CoV-2 Transmission." Social Science & Medicine 276: 113741. https://doi.org/10.1016/j.socscimed.2021.113741

Richardson, Eugene T. 2020. *Epidemic Illusions: On the Coloniality of Global Public Health*. MIT Press.

Sismondo, Sergio. 2011. "Questioning Functionalism in the Sociology of Science." In *An Introduction to Science and Technology Studies*, 23–35. 2nd ed. Wiley-Blackwell.

Smith, Linda Tuhiwai. 2012. *Decolonizing Methodologies: Research and Indigenous Peoples*. 2nd ed. London & Zed Books.

Spiegler, Peter. 2015. *Behind the Model: A Constructive Critique of Economic Modeling*. Cambridge University Press.

Wallace, Rodrick, and Deborah N. Wallace. 1999. "Emerging Infections and Nested Martingales: The Entrainment of Affluent Populations into the Disease Ecology of Marginalization." *Environment and Planning A: Economy and Space* 31: 1787–1803. https://doi.org/10.1068/a311787

Winsberg, Eric. 2010. *Science in the Age of Computer Simulation*. University of Chicago Press.

Wynter, Sylvia. 1994. "'No Humans Involved:' An Open Letter to My Colleagues." In *Forum N. H. I: Knowledge for the 21st Century* 1 (1), 42–71. Institute N. H. I. https://libcom.org/files/Wynter5.pdf