The Center for Growth and Opportunity at Utah State University

Public Interest Comment for the National Telecommunications and Information Administration (NTIA) on the Intersection of Privacy, Equity, and Civil Rights

Author:
Will Rinehart[a]

Docket ID: NTIA-2023-0001-0001
Submitted: March 6, 2023

The Center for Growth and Opportunity at Utah State University is a research center dedicated to producing ideas that transform lives. We explore the interactions between key institutions—business, government, and civil society—to improve opportunity, broad-based economic growth, and individual well-being. The Center occasionally conducts independent analyses addressing government rulemakings and proposals. This comment is designed to assist the agency as it explores these issues. The views expressed in this comment are those of the author(s) and do not necessarily reflect the views of The Center for Growth and Opportunity at Utah State University or the views of Utah State University.

a Will Rinehart, Senior Research Fellow, The Center for Growth and Opportunity at Utah State University

The Center for Growth and Opportunity at Utah State University

The National Telecommunications and Information Administration asks a series of crucial questions in this proceeding. It seeks to understand how commercial data collection and use, especially through artificial intelligence (AI) methods, might adversely affect underserved or marginalized communities through disparate impacts.[1] Importantly, it wants to understand how "specific data collection and use practices potentially create or reinforce discriminatory obstacles for marginalized groups regarding access to key opportunities, such as employment, housing, education, healthcare, and access to credit."

What the NTIA seeks to tackle is a wicked problem in Rittel and Webber's classic definition.[2] The following comments argue for a twist on that theme. Wicked problems, which plague public policy and planning are distinct from natural problems because "natural problems are definable and separable and may have solutions that are findable [while] the problems of governmental planning and especially those of social or policy planning are ill-defined." But the case of fairness in AI shows that they are over-defined. The reason why "social problems are never solved," they "are only resolved-over and over again" is because there are many possible solutions.

When the NTIA issues its final report, it should resist the tendency to reduce wicked problems into natural ones. Rather, the agency should recognize, as one report described it, the existence of a hidden universe of uncertainty about AI models.

The following comments are intended to address this problem holistically:

- The first section explains how data-generating processes can create legibility but never solve the problem of illegibility.
- The second section explains what is meant by bias, breaks down the problems in model selection, and walks through the problem of defining fairness.
- The third section explores why people have a distaste for the kind of moral calculations made by machines and why we should focus on impact.

Legibility and illegibility

Nowhere in this proceeding is the most important first question that everyone working with data must ask: *What are the data-generating processes (DGPs)?*

Data-generating processes are the processes that cause data to occur as they do. For companies and digital platforms, data generation comes as a result of their business. If Google didn't exist, there would be no search graph data. If Facebook didn't exist, there wouldn't be social graph data. If Walmart didn't exist, there be no scanner data.[3] Data generation comes from a specific method or technology. As such, the method produces the result.

Business data is generated by and calibrated to a purpose or end goal. DGPs might create metrics like clicks and time spent on the webpage, or they might produce metrics such as key performance indicators (KPIs) like revenue growth, revenue per client, and profit margin. Each of these data

---

1 "Privacy, Equity, and Civil Rights Request for Comment," Federal Register, January 20, 2023, https://www.federalregister.gov/documents/2023/01/20/2023-01088/privacy-equity-and-civil-rights-request-for-comment. To be filed at "Regulations.Gov," https://www.regulations.gov/document/NTIA-2023-0001-0001.

2 Horst W. J. Rittel and Melvin M. Webber, "Dilemmas in a General Theory of Planning," *Policy Sciences* 4, no. 2 (June 1, 1973): 155–69, https://doi.org/10.1007/BF01405730.

3 Walmart Supply Chain, "RFID Technology," *Walmart Supply Chain* (blog), October 1, 2013, https://walmartsupplychain.weebly.com/rfid-technology.html.

have a different goal, they might aim to increase sales or brand awareness, but they are all in service of the business itself.

Goodhart's Law is a famous warning in metrics that "when a measure becomes a target, it ceases to be a good measure." In the classical understanding of the Law, when a measure is used to reward performance, incentives align to manipulate the measure. But the more subtle reading of the Law is that every measurement aimed at a target means a myopia of other targets. As Kenneth Burke noted, "A way of seeing is also a way of not seeing—a focus upon object A involves a neglect of object B."

DGPs create legibility and illegibility. Every act of measurement implies some shadow that isn't measured.

The term legibility comes from James C. Scott, a political theorist whose work has focused on early state formation. As he defined it, legibility references,

> a state's attempt to make society legible, to arrange the population in ways that simplified the classic state functions of taxation, conscription, and prevention of rebellion. Having begun to think in these terms, I began to see legibility as a central problem in statecraft. The premodern state was, in many crucial respects, partially blind; it knew precious little about its subjects, their wealth, their landholdings and yields, their location, their very identity. It lacked anything like a detailed "map" of its terrain and its people. It lacked, for the most part, a measure, a metric, that would allow it to "translate" what it knew into a common standard necessary for a synoptic view.[4]

Scott used the term *legibility* in a narrow sense to describe government, but it should be understood more broadly as the ability of a metric to capture some underlying feature of the world to grant insight into it. Businesses collect and produce metrics like KPIs to make their internal operations legible. But it is never complete. There are often aspects of a business that remain unquantified and unable to be seen. *Illegibility* remains.

This proceeding seemingly skips over the problems of illegibility. Indeed, there is no mention of data deserts or lack of data in the AI democracy blueprint or in this proceeding.[5] Oftentimes groups of people remain illegible because they are outside of the scope of the DGP. Critically then, some populations don't have enough data collected on them.[6]

People can be excluded from data generating processes because of unequal access to broadband services, access to technology, or disparities in digital literacy. Decisions may overlook the unique needs of members of communities when these communities are not represented (or underrepresented) in the data.

Even for companies that collect troves of data, illegibility exists. Because of the scrutiny it has faced, Meta's problems with illegibility are particularly well-known:

- As *The Verge* reported, employees at Meta discovered by March 2021 that "our ability to detect vaccine-hesitant comments is bad in English, and basically non-existent else-

4 James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, New Haven: Yale University Press, 2020.

5 *Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource*, National Artificial Intelligence Research Resource Task Force, January 2023, https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf.

6 "Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms," Health Informatics | JAMA Network, September 2020, https://jamanetwork.com/journals/jama/fullarticle/2770833.

where."[7] Another employee chimed in, noting that comments "are almost a complete blind spot for us in terms of enforcement and transparency right now," even though they make up a "significant portion of misinfo on FB."

- Meta employees also wrote that, "We currently don't understand whether Group comments are a serious problem. It's clear to us that the 'good post, bad comment' problem is a big deal, but it's not necessarily as clear that comments on posts are additive to the harm."[8]
- Reporting from CNN on leaked internal documents on January 6th, "[A]t the time it was very difficult to know whether what we were seeing was a coordinated effort to delegitimize the election, or whether it was protected free expression by users who were afraid and confused and deserved our empathy. But hindsight being 20:20 makes it all the more important to look back to learn what we can about the growth of the election delegitimatizing movements that grew, spread conspiracy, and helped incite the Capitol insurrection."[9]
- "Facebook was counting on its artificial-intelligence system to make the platform work properly by enforcing its rules against violent or hate-filled speech. But its own documents show the system can't tell the difference between car crashes and cockfights."[10]

Just as important, even if a group is understood to be legible, we shouldn't assume that the DGP is accurately capturing the variable of interest.

For example, research has found that Facebook Like data can be used to accurately predict highly sensitive personal attributes like sexual orientation, ethnicity, religious view, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, gender, and, most important for this discussion, political opinions.[11] In practice, however, it is unclear if Facebook has this level of predictability.

Facebook splits people into groups for advertising purposes in what they call affinities. When Pew surveyed users in 2019 and asked about how well these categories actually track their preferences, only 13 percent said that they are very accurate descriptions.[12] Another 46 percent of users thought the categories were somewhat accurate. On the negative side of the ledger, 27 percent of users "feel it does not represent them accurately" and another 11 percent of users weren't assigned categories at all. In other words, over a third of all users are effectively in error.

Twitter miscounted user profile counts for years, overcounting by up to 1.9 million users each quarter.[13] The error was due to Twitter inadvertently counting multiple accounts as active when

---

7 Russell Brandom, Alex Heath, and Adi Robertson, "Eight Things We Learned from the Facebook Papers," *The Verge*, October 25, 2021, https://www.theverge.com/22740969/facebook-files-papers-frances-haugen-whistleblower-civic-integrity.

8 Brandom, Heath, and Robertson, "Eight Things We Learned."

9 Donie O'Sullivan, Tara Subramaniam, and Clare Duffy, "Not Stopping 'Stop the Steal:' Facebook Papers Paint Damning Picture of Company's Role in Insurrection," CNN, October 24, 2021, https://www.cnn.com/2021/10/22/business/january-6-insurrection-facebook-papers/index.html.

10 Deepa Seetharaman, Jeff Horwitz, and Justin Scheck "Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts." *Wall Street Journal*, October 17, 2021, https://t.co/aDv0yJ0Glp

11 Michal Kosinski, David Stillwell, and Thore Graepel, "Private Traits and Attributes Are Predictable from Digital Records of Human Behavior," *Proceedings of the National Academy of Sciences* 110, no. 15 (April 9, 2013): 5802–5, https://doi.org/10.1073/pnas.1218772110.

12 Sara Atske, "Facebook Algorithms and Personal Data," *Pew Research Center: Internet, Science & Tech* (blog), January 16, 2019, https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/.

13 Jacob Kastrenakes, "Twitter Miscounted Its Daily Users for Three Years Straight," *The Verge*, April 28, 2022, https://www.theverge.com/2022/4/28/23046170/twitter-miscounted-daily-users-three-years-q1-2022-earnings.

they were all tied to a single user, even if they weren't all in use. These incorrect usage numbers were given for Q1 2019 through Q4 2021.

This proceeding needs to recognize that a lack of good data is an important part of understanding and overcoming bias.

Bias and model selection

The proceedings seem to presuppose that digital bias originates with human bias found in the algorithm designer or in the data collection process, whether intentional or unintentional. While this can certainly be the source of bias in some instances, it is not universally true, and creating regulations that focus only on eliminating human bias in systems could actually have the effect of making things less fair for marginalized or disadvantaged communities.

Bias occurs in statistical models as a result of differences in distribution. Efforts to eliminate bias in large populations should be seen more as trade-offs between bias and variance, not a problem that can be solved outright.

Consider this example:

Let's say we have a group of people and estimated their height via a statistical model. For simplicity's sake, we know the mean height is 5'10", but our model produced an estimate that said everyone was 6'. The estimate would be biased by 2 inches. To statisticians, economists, and data scientists, bias has a very specific meaning. Bias is the property of an estimate that describes how far it is from the true value of a population.

In the real world, however, we often cannot know the true estimate of a population. We don't know what the mean truly is. And so, most classifiers trade-off between bias and another quality, variance. Variance describes the variability of the prediction, the spread of the estimates.

Going back to the example above, suppose that instead of just one estimate of height, we calculated four and this time, we got 5'8", 5'10", 6', and 6'2". In this round of estimates, our mean comes to 5'11", which is closer to the 5'10" mean. However, the variance would be higher because we got a range of different estimates that weren't the correct mean. In the real world, bias is often traded for variance.

Indeed, this trade-off is really a subclass of a larger problem that is of central focus in data science, econometrics, and statistics. As Pedro Domingos noted:

> You should be skeptical of claims that a particular technique "solves" the overfitting problem. It's easy to avoid overfitting (variance) by falling into the opposite error of underfitting (bias). Simultaneously avoiding both requires learning a perfect classifier, and short of knowing it in advance there is no single technique that will always do best (no free lunch).[14]

This gets even more complicated when two populations coexist.

When two populations have different feature distributions, the classifier will fit the larger population because they contribute more to the average error. Minority populations can benefit or suffer depending on the nature of the distribution difference. This is not based on explicit human

14 Pedro Domingos, "A Few Useful Things to Know about Machine Learning," *Communications of the ACM* 55, no. 10 (October 1, 2012): 78–87, https://doi.org/10.1145/2347736.2347755.

bias, either on the part of the algorithm designer or on the part of the data collection process, and it is worse if we force the algorithm to be group-blind artificially. So, it is possible that regulations intended to promote fairness can actually make things less fair and less accurate by prohibiting decision makers from considering sensitive attributes.

Julia Angwin and her team at ProPublica stumbled upon this fact when they dove deeper into a commonly used post-trial sentencing tool known as COMPAS (Correctional Offender Management Profiling for Alternative Sanctions).[15] Instead of predicting behavior before a trial takes place, COMPAS purports to predict a defendant's risk of committing another crime in the sentencing phase after a defendant has been found guilty. As Angwin's team discovered, the risk system was biased against African American defendants, who were more likely to be incorrectly labeled as higher risk than they actually were. At the same time, white defendants were labeled as lower risk than what was actually the case.

The task of the COMPAS tool is to estimate the degree to which people possess a likeliness for future risk. In this sense, the algorithm aims for calibration, one of at least three distinct ways we might understand fairness. Aiming for fairness through calibration means that people were correctly identified as having some probability of committing an act. Indeed, as subsequent research has found, the number of people who committed crimes were correctly distributed within each group.[16] In other words, the algorithm did correctly identify a set of people as having a probability of committing a crime.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan explain the nature of Angwin's criticism in "Inherent Trade-Offs in the Fair Determination of Risk Scores."[17] The kind of fairness that Angwin aligns with might be understood as a balance for the positive class. To violate this kind of fairness notion, people would be later identified as being part of the class, yet they were initially predicted as having a lower probability by the algorithm. For example, as the ProPublica study found, white defendants that did commit crimes in the future were assigned lower risk scores. This would be a violation of balance for the positive class.

Similarly, fairness could be understood as balancing for a negative class. To violate this kind of fairness notion, people that would be later identified as not being part of the class would be predicted initially as having a higher probability of being part of it by the algorithm. Both of these conditions try to capture the idea that groups should have equal false negative and false positive rates.

After formalizing these three conditions for fairness, Kleinberg, Mullainathan, and Raghavan proved that it isn't possible to satisfy all constraints simultaneously except in special cases. These results hold regardless of how the risk assignment is computed, since "it is simply a fact about risk estimates when the base rates differ between two groups."

Some views of fairness might simply be incompatible with each other. Balancing one kind of notion of fairness is likely to come at the expense of another. Meanwhile, most people tend to hold steadfast in demanding fairness. Across a range of laboratory studies, cross-cultural research, and experiments with babies and young children, humans seem to favor fair distributions, not

---

15 Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

16 Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores," Research Gate, September 19, 2016, https://www.researchgate.net/publication/308327297_Inherent_Trade-Offs_in_the_Fair_Determination_of_Risk_Scores.

17 Kleinberg, Mullainathan, and Raghavan, "Inherent Trade-Offs."

equal ones. Just as important, when fairness and equality clash, people prefer fair inequality over unfair equality.[18]

To add to the confusion, there is no guarantee that there will be a convergence toward one specific estimate.

Research conducted in the past couple of years only confirms this point. In one study, 73 independent research teams used identical cross-country survey data to see if more immigration reduces support for government services.[19] As the authors concluded, "Instead of convergence, teams' results varied greatly, ranging from large negative to large positive effects of immigration on social policy support. The choices made by the research teams in designing their statistical tests explain very little of this variation; a hidden universe of uncertainty remains."

Other studies confirm the flexibility that exists in data interpretation. Comparisons of independent analysts using the same dataset also found varying results on the effects of gender and professional status on verbosity during group meetings,[20] whether soccer referees are more likely to give red cards to dark-skin-toned players than to light-skin-toned players,[21] and how the same MRIs were analyzed by different teams of scientists.[22]

Rightly, this proceeding is worried that, "Even when digital advertisers do not intend to use discriminatory targeting criteria, the datasets they use may reflect current or historic inequities and the algorithms they use may unintentionally replicate those biases or others—such as untargeted ads for certain types of jobs being delivered disproportionately to men or women." Despite having neutral targeting parameters, ads for employment and housing opportunities can be delivered with a significant skew along gender and racial lines.[23]

But the problem is even deeper than the NTIA suggests. Teams working to reduce unfairness might not converge on the same model or parameters, leaving disadvantaged or marginalized groups in the middle of competing efforts on their behalf. This sounds like more of a good thing, but as already mentioned, these efforts could backfire, leaving groups even more vulnerable than they were before the intervention.

The adoption of AI in the real world

People court failure in predictable ways. Human beings routinely take shortcuts because of limited time, personnel, or resources that create poor conditions for decision making. AI based systems could help to reverse some of our worse tendencies, but they have to be implemented in the real world. And in the real world, new systems will not be adopted without friction. These facts about the world suggest that the NTIA should be focused on outcomes and real-world impact.

---

18 Christina Starmans, Mark Sheskin, and Paul Bloom, "Why People Prefer Unequal Societies," *Nature Human Behaviour* 1, no. 4 (April 7, 2017): 1–7, https://doi.org/10.1038/s41562-017-0082.

19 Nate Breznau, Eike Mark Rinke, Alexander Wuttke, and Tomasz Żółtak, "Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty" PNAS," October 28, 2022, https://www.pnas.org/doi/10.1073/pnas.2203150119.

20 Martin Schweinsberg et al., "Same Data, Different Conclusions: Radical Dispersion in Empirical Results When Independent Analysts Operationalize and Test the Same Hypothesis," *Organizational Behavior and Human Decision Processes* 165 (July 1, 2021): 228–49, https://doi.org/10.1016/j.obhdp.2021.02.003.

21 R. Silberzahn et al., "Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results," *Advances in Methods and Practices in Psychological Science* 1, no. 3 (September 1, 2018): 337–56, https://doi.org/10.1177/2515245917747646.

22 Rotem Botvinik-Nezer et al., "Variability in the Analysis of a Single Neuroimaging Dataset by Many Teams," *Nature* 582, no. 7810 (June 2020): 84–88, https://doi.org/10.1038/s41586-020-2314-9.

23 Muhammad Ali et al., "Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes," *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (November 7, 2019): 1–30, https://dl.acm.org/doi/pdf/10.1145/3359301.

The implementation of pretrial risk assessment instruments highlights the potential variability when algorithms get deployed. Pretrial risk assessment instruments help to guide judges when they decide what is going to happen to a defendant before a trial. Will they be put on bail and what will be the cost? The most popular of these instruments is known as the Public Safety Assessment (PSA), which was developed by the Laura and John Arnold Foundation and has been adopted in over 30 jurisdictions in the last five years.

Judges demonstrate all kinds of bias. When the Louisiana State University football team loses in an upset, for example, judges in Louisiana add 1,296 days of punishment to juvenile defendants, as well as 136 extra days of jail time.[24] Anything that puts judges in bad mood tends to lead to harsher sentences. A study of 207,000 immigration cases found that "a 10° F degree increase in case-day temperature reduces decisions favorable to the applicant by 6.55%."[25] It was hoped that pretrial risk assessments would mitigate these problems.

The adoption of the PSA across regions helps to demonstrate just how disparate implementation can be. In New Jersey, the adoption of the PSA seems to have correlated with a dramatic decline in the pretrial detention rate.[26] In Lucas County, Ohio, the pretrial detention rate increased after the PSA was put into place.[27] In Chicago, judges seem to be simply ignoring the PSA.[28] Indeed, there seems to be little agreement on what the high risk classification corresponds to in the PSA, as re-arrest can be as low as 10 percent or as high as 42 percent depending on how the PSA is integrated in a region.[29]

And in the most comprehensive study of its kind, law professor Megan Stevenson at George Mason University looked at Kentucky after it has implemented the PSA and found significant changes in bail setting practice, but only a small increase in pretrial release.[30] Over time, these changes eroded as judges returned to their previous habits.

Although it was focused on pretrial risk assessments, Stevenson's call for a broader understanding of these tools applies to the entirety of algorithm research:

> Risk assessment in practice is different from risk assessment in the abstract, and its impacts depend on context and details of implementation. If indeed risk assessment is capable of producing large benefits, it will take research and experimentation to learn how to achieve them. Such a process would be evidence-based criminal justice at its best: not a flocking towards methods that bear the glossy veneer of science, but a careful and iterative evaluation of what works and what does not.

---

24 Ozkan Eren and Naci Mocan, "Emotional Judges and Unlucky Juveniles," *American Economic Journal: Applied Economics* 10, no. 3 (July 2018): 171–205, https://doi.org/10.1257/app.20160390.

25 Anthony Heyes and Soodeh Saberian, "Temperature and Decisions: Evidence from 207,000 Court Cases," *American Economic Journal: Applied Economics* 11, no. 2 (April 2019): 238–65, https://doi.org/10.1257/app.

26 Jon Schuppe, "Post Bail," NBC News Specials, August 22, 2017, https://www.nbcnews.com/specials/bail-reform/.

27 Notice of Filing of Copy of Presentation Assessing Impact of Public Safety Assessment, *Jones v. Wittenberg*, No. C70-388 (Dist. Ct. N. Dist. of Ohio, Western Div.), https://thecrimereport.org/wp-content/uploads/2017/08/Lucas-County-court-filing.pdf

28 Frank Main, "Cook County Judges Not Following Bail Recommendations: Study," *Chicago Sun-Times*, July 3, 2016, https://chicago.suntimes.com/news/cook-county-judges-not-following-bail-recommendations-study-find/.

29 Mayson, Sandra Gabriel, "Dangerous Defendants," Yale Law Journal 490 (2018), http://dx.doi.org/10.2139/ssrn.2826600.

30 Stevenson, Megan, "Assessing Risk Assessment in Action," Minnesota Law Review 303 (2018), http://dx.doi.org/10.2139/ssrn.3016088.

By and large, however, people have a distaste for the kind of moral calculations made by machines.[31] Even when the consequences are good, people tend to mark certain actions as right and wrong according to moral intuition, not according to their consequences.[32] In philosophy, economics, and data science, however, calculating consequences is the de facto way to make moral decisions. But to most people, just focusing on consequences is ethically unsatisfying.

Take credit scores, for example. Even though this algorithm is specifically blind to race and other protected categories, it garners fierce criticism because it reproduces already existing tendencies. "Rather than leveling the playing field, credit scores serve as gatekeepers to wealth-building for communities already facing the highest barriers," says Common Future.[33]

While it is imperfect, the wide adoption of credit scores has been important in pushing loan decisions toward nondiscriminatory practices.[34] In a study that predates the build-up in housing credit, the implementation of sophisticated risk models was found to be connected to the expansion of home ownership in minority communities, helping it to grow from 34 percent to 47 percent between 1983 and 2001.[35] Gender, race, religion, nationality, and marital status were implicit factors in decision making before credit scores.[36] It would be a stretch to claim that the previous system of judgmental lending was *more legitimate*.

Effectively, only one study exists that shows the causal impact of credit scoring on households' loan pricing. Bank, Segev, and Shaton (2022) were able to track loans in Israel before and after the implementation of a formal credit score to calculate the true impact of credit scores on loans.[37] What they found is that the credit score led to a decrease in loan prices for households with strong relationship banking. When banks held a monopoly on potential borrowers' credit history, they charged higher interest rates, all else being equal. The suggestion here is that scores have helped everyone, not just marginalized communities, to have better access to credit.

When the NTIA produces its report, it should ensure that it focuses on impact and real harm, not potential harm.

Indeed, the most high-profile case that purported to show AI harm never got around to discussing the impact. Sparked by *National Fair Housing Alliance v. Facebook, Inc.*, the Department of Justice (DOJ) eventually got Meta to agree to change their ad services to settle complaints over ad discrimination on its platform.[38] The original complaint by the National Fair Housing Alliance

---

31 1. Fiery Cushman, Liane Young, and Marc Hauser, "The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm," *Psychological Science* 17, no. 12 (December 1, 2006): 1082–89, https://doi.org/10.1111/j.1467-9280.2006.01834.x. 2. Daniel M. Bartels and David A. Pizarro, "The Mismeasure of Morals: Antisocial Personality Traits Predict Utilitarian Responses to Moral Dilemmas," *Cognition* 121, no. 1 (October 1, 2011): 154–61, https://doi.org/10.1016/j.cognition.2011.05.010.

32 Joshua D. Greene et al., "An FMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293, no. 5537 (September 14, 2001): 2105–8, https://doi.org/10.1126/science.1062872.

33 Common Future, "Why Credit Scores Are Racist," *Common Future* (blog), July 28, 2021, https://medium.com/commonfuture/why-credit-scores-are-racist-da109fcfb300.

34 Hollis Fishelson-Holstine, "The Role of Credit Scoring in Increasing Homeownership for Underserved Populations," Joint Center for Housing Studies Working Paper Series, Harvard University, February 2004, http://www.jchs.harvard.edu/sites/jchs.harvard.edu/files/babc_04-12.pdf.

35 Michael Turner, *The Fair Credit Reporting Act: Access, Efficiency & Opportunity The Economic Importance Of Fair Credit Reauthorization*, Information Policy Institute, June 2003, http://www.perc.net/wp-content/uploads/2013/09/fcra_report_exec_sum.pdf.

36 Lauer, Josh. *Creditworthy a History of Consumer Surveillance and Financial Identity in America*. New York, N.Y: Columbia University Press, 2017.

37 Tali Bank, Nimrod Segev, and Maya Shaton, "Relationship Banking and Credit Scores: Evidence from a Natural Experiment," Working Paper, January 10, 2022, https://www.dropbox.com/s/avkqs5nd89hltcg/Relationship%20Banking%20and%20Credit%20Scores_10012022.pdf?dl=0.

38 "Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising," June 21, 2022, https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known.

alleged that Facebook's classification of its users and its ad targeting tools permitted landlords, developers, and housing service providers to limit the audience for their ads based on sex, religion, familial status, and national origin in violation of the FHA.

As a result of the settlement, Meta dropped the "Special Ad Audience" tool for its housing ads, which utilized a discriminatory algorithm, according to the complaint. In the process of complying with the DOJ, the company got rid of thousands of ad categories, including the much-maligned "African American multicultural affinity."[39]

But there was never a formal study or finding that the Facebook tool had the effect of being discriminatory in practice. It could have been that these tags were used in a positive manner to target Black Americans, rather than to limit their choices.

Besides, it was well known by advertisers that the targeting criteria had their problems. As the Pew data cited above only confirms, these affinities weren't all that accurate. Indeed, the author of this comment, who is white, was categorized under "African American multicultural affinity."

Follow-up research of the changes made by Meta in the wake of the DOJ pressure seem to confirm it had little effect. As one report explained it, "Merely removing demographic features from a real-world algorithmic system's inputs can fail to prevent biased outputs."[40] In the end, the authors of the report suggested other approaches to mitigating discriminatory effects. But oftentimes, those other approaches have negative consequences.

Equal-exposure is one method that purports to deal with fairness. In this scenario, advertising platforms might artificially raise the bid of an economic-opportunity advertiser to purchase female impressions (or give them away for free). According to one analysis, advertising platforms may earn more profit if equal-exposure fairness is enforced because it intensifies the competition between advertisers.[41] When certain groups get buying power, they often have the effect of increasing the advertising platform's profit. As such, advertisers might have an interest to adopt equal-exposure fairness.

In other words, remedies might not actually solve the problems they intend to reverse. They might just entrench current business interests. The focus for the NTIA should be on real impacts.

Conclusion

NTIA's goal to better understand discriminatory disparities in the digital economy is both worthy and messy. It is a wicked problem that cannot be easily defined or solved. The report that NTIA creates as a result of these proceedings should acknowledge the complexity of these issues. NTIA should present policymakers with options for gaining a comprehensive understanding of the issues before moving cautiously forward with regulations, recognizing that any action taken by policymakers may improve or exacerbate the problem.

---

39 Andrew Hutchinson, "Facebook Removes Over 1,000 Ad Targeting Options Due to Low Usage," Social Media Today, August 12,2020, https://www.socialmediatoday.com/news/facebook-removes-over-1000-ad-targeting-options-due-to-low-usage/583406/.

40 Piotr Sapiezynski et al., "Algorithms That 'Don't See Color': Measuring Biases in Lookalike and Special Ad Audiences," in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '22 (New York, NY, USA: Association for Computing Machinery, 2022), 609–16, https://doi.org/10.1145/3514094.3534135.

41 Di Yuan, Manmohan Aseri, and Tridas Mukhopadhyay, "Is Fair Advertising Good for Platforms?," SSRN Scholarly Paper (Rochester, NY, August 20, 2021), https://doi.org/10.2139/ssrn.3913739.