If the data do not speak for themselves, how ought we to speak for the data?

Ian Van Buskirk[1], **Brian Zaharatos**[2], **Aaron Clauset**[1, 3, 4], **Daniel B. Larremore**[1, 3]

[1] Department of Computer Science, University of Colorado Boulder
[2] Department of Applied Mathematics, University of Colorado Boulder
[3] BioFrontiers Institute, University of Colorado Boulder
[4] Santa Fe Institute
ian@colorado.edu, brian.zaharatos@colorado.edu, aaron.clauset@colorado.edu, daniel.larremore@colorado.edu

The falling barrier between scientific scholarship and politics represents a looming challenge for our field. A growing number of authors, spanning statistics (Clayton 2020), data science (D'ignazio and Klein 2020), and computational social science (Hu 2021), urge us to give up the pretense of objectivity, impartiality, and neutrality in our work, and instead "embrace a political orientation" (Green 2021). Indeed, technologies are not neutral: they encode particular sets of values and as such shape the world and the people in it (Miller 2021). These authors recognize, correctly, that when mathematical or computational methods are oriented toward social questions, the answers necessarily involve politics. But, the most difficult subsequent questions have been left open. Toward what politics should we orient ourselves, as individuals? How can or should a community of scholars adjudicate when different views emerge? And given that data do not (and should not) speak for themselves, how should we fill the silence?

Answering these questions will require "a set of evaluative standards that transcends the competing interests of those who advocate rival answers to a question" (Anderson 1995). In their absence, simply heeding the call for political action may, in practice, look more like the expression of personal preference than principled engagement—a perception that would undermine the credibility of scholarship in general. To avoid this, we need higher level values which are separate from the politics of particular social and technological issues. In 1980, Winner argued that technologies, being inherently political, may "be compatible with" some kinds of outcomes, but may "require" others (Winner 1980). In the remainder of this essay, we propose an adaptation of Winner's views as a way to consider the positive and negative impacts of technologies, and address the looming challenges of a politically oriented computational social science.

Our adaptation of Winner's framework asks three questions of a technology: **Which practices does a technology merely enable? Which practices does it encourage? Which does it make inevitable?** Answering these questions requires that we parse and assess the potential impacts of a technology from a more impartial position—not good and bad, but rather possible, probable, and inevitable—and may therefore bring us one step closer to "a set of evaluative standards that transcends the competing interests of those who advocate rival answers to a question" (Anderson 1995).

This three-factor framework can be applied to technologies of pressing interest today, including large language models, algorithms for policing and bail, and recommender systems in social media. However, our own engagement with this framework arose from grappling with *name-based gender classification*—the process of assigning individuals gendered labels based on their names alone—a technology we have both used and developed in our own work.

Name-based gender classification (NBGC) is a common method both inside and outside the academy, with a variety of free and paid services available. Outside of commercial applications such as targeted advertising, most scholarship engages briefly with the practice and politics of NGBC in the Methods or Discussion sections (West et al. 2013; Dworkin et al. 2020), typically by noting some discomfort with the practice while nevertheless using it. At a high level, these discomforts are typically that (i) NBGC clusters individuals into identically-labeled groups, (ii) typically provides man/-woman binary labels or a 0-to-1 scalar on a man/woman axis, (iii) may misclassify (misgender) some and (iv) may fail to classify (and thus exclude) others. Based on personal conversations and correspondence, the pattern of disclaiming one or more of these discomforts and then proceeding seems to stem mostly from uncertainties with how to engage the politics of NBGC.

Our initial interest in NBGC was simply as users of the technology in studying gender inequalities, illustrating the technology's ability to **enable**. Indeed, NBGC has enabled the study of gender inequalities in diverse areas, from academic hiring and retention (Wapman et al. 2022), to citation practices (Dworkin et al. 2020) and authorship dynamics (West et al. 2013). It has also enabled commercial enterprises to gender-target advertisements, algorithmically deny or approve loans, and customize email salutations. In this context, while one might view scholarly applications as implicitly positive and commercial applications as implicitly negative, we note that there may be both positive and negative valences to each of these uses, yet it is nevertheless clear that NBGC has enabled all of them.

Which practices does NBGC **encourage**? Here, things get murkier still. First, NBGC shapes the way people think, both about gender and about how we study it. All NBGC methods of which we are aware operate within a binary framework. One might argue that this framework is conceptually distinct from other accounts of gender such as gender identity, expression, or perception—that NBGC methods capture a structural dimension of gender. Indeed, name-gender associations reflect how one's social position is derivative of real (or perceived) sex-related characteristics (Haslanger 2000), and academic users of NBGC may indeed reason that they hope to measure how others perceive one's gender, not one's gender identity per se (e.g., in studies of how names affect callback rates for otherwise identically qualified job applicants). Nevertheless, it is a real risk that NBGC reaffirms and encourages a gender binary in a harmful way. Further, NBGC can bias studies towards the quantitative and the aggregate, papering over meaningful distinctions the way technology too often does (Dreyfus and Kelly 2011).

On the other hand, NBGC encourages scholars to study social inequalities and other important social phenomena. Indeed, this was a major reason that, in the development of our own method for NBGC, based on cultural consensus theory, we set the goal of creating a method that was not only competitive with paid NBGC services, but was free, open-data, and open-source (Van Buskirk, Clauset, and Larremore 2022), and therefore more likely to encourage use by other academics whose values favor transparency and whose budgets are limited.

Technologies also wield their political influence by encouraging not only practices but entire systems to set up around them (Winner 1980). For instance, once one decides that a technology or method ought to exist, one must confront how to make the technology available, an especially familiar process to those developing computational tools. For instance, our consideration of transparent NBGC methods and data led us to release a Python package (Van Buskirk 2022b) along with datasets that capture name-gender associations at multiple levels of granularity (Van Buskirk 2022a). However, perhaps we should go further: build an api, design a website, or have cute animations (a la gender-api.com). Instead, perhaps we should restrict access to our work by requiring scientists to apply for a license and evaluating their use-cases. Or, we could require interested parties to send us their data so we can privately use our method and avoid the need to release our name-gender association data at all. Navigating this decision is about more than just who gets to use our method. It is about the kind of power structures we want to see around the technologies we use in computational social science and related fields. Indeed, many of the options above (both to further open and to restrict our work) have been suggested to us during the peer review process, reflecting an engagement with the practices and cultures that a technology encourages.

Finally, like all technologies, NBGC makes some outcomes **inevitable**. First and foremost, the practice of associating names and genders will misgender some individuals. If one attempts to sidestep this issue by declining to classify those names with a weakly gendered cultural consensus, then one excludes those individuals from analysis entirely—an inescapable dilemma. We note that this inevitability generalizes to any imperfect classifier, i.e., any system, code, or process with nonzero misclassification rates. Ironically then, the presence of misclassification errors allows us to argue in absolutes, not about the possibilities or probabilities of technologies, but inevitabilities.

Even after consideration of the possible, the probable, and the inevitable, researchers must be accountable to their own evaluations and decisions. Our view is that the benefits of NBGC technology outweigh the harms, but not everyone agrees. In engaging other views, what arguments are at our disposal? Must we rely only on our personal beliefs, or could we collectively appeal to a more general set of values? To make clear how difficult this task may be, consider that it is not (even) enough for two discussants to share a consequentialist framework, agree on the potential harms and benefits, and accept a common accounting of the severities and probabilities of each: if one person believes that the *average* of the expected consequences is most important, while the other person embraces a Rawlesian *maximin* principle—that the right decision is that which maximizes the minimum outcome, e.g., the outcome for the most vulnerable or disempowered—they may arrive at very different conclusions. How can we collectively reason which ethics, theory of justice, and as a result politics, to accept?

We believe that this three-factor framework, considering the possible, the probable, and the inevitable, provides a useful approach to move beyond a good-vs-bad framing of the technologies and methods of computational social science, data science, and statistics. Nevertheless, even equipped with this approach, we still need a clear set of values to contextualize our discussions, and here, we offer no solutions at present other than open and good-faith discussion, and accountability to one's conclusions.Without such accountability, we lack a principled way to navigate differences of view and leave too much to be settled by power and the politics of those who have it.

References

Anderson, E. 1995. Knowledge, human interests, and objectivity in feminist epistemology. *Philosophical topics*, 23(2): 27–58.

Clayton, A. 2020. How Eugenics Shaped Statistics.

D'ignazio, C.; and Klein, L. F. 2020. *Data feminism*. MIT press.

Dreyfus, H.; and Kelly, S. D. 2011. All things shining: Reading the Western classics to find meaning in a secular age. *Adam Academy Journal of Social Sciences*, 1(2).

Dworkin, J. D.; Linn, K. A.; Teich, E. G.; Zurn, P.; Shinohara, R. T.; and Bassett, D. S. 2020. The extent and drivers of gender imbalance in neuroscience reference lists. *Nature neuroscience*, 23(8): 918–926.

Green, B. 2021. Data science as political action: grounding data science in a politics of justice. *Journal of Social Computing*, 2(3): 249–265.

Haslanger, S. 2000. Gender and race : (What) are they? (What) do we want them to be? *Noûs*, 34(1): 31–55.

Hu, L. 2021. Race, Policing, and the Limits of Social Science.

Miller, B. 2021. Is technology value-neutral? *Science, Technology, & Human Values*, 46(1): 53–80.

Van Buskirk, I. 2022a. Name-Based Gender Classification. https://osf.io/tz38q.

Van Buskirk, I. 2022b. nomquamgender. https://github.com/ianvanbuskirk/nomquamgender.

Van Buskirk, I.; Clauset, A.; and Larremore, D. B. 2022. An Open-Source Cultural Consensus Approach to Name-Based Gender Classification. *arXiv preprint arXiv:2208.01714*.

Wapman, K. H.; Zhang, S.; Clauset, A.; and Larremore, D. B. 2022. Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature*, 610(7930): 120–127.

West, J. D.; Jacquet, J.; King, M. M.; Correll, S. J.; and Bergstrom, C. T. 2013. The role of gender in scholarly authorship. *PloS one*, 8(7): e66212.

Winner, L. 1980. Do artifacts have politics? *Daedalus*, 121–136.