

Integrating Computer Prediction Methods in Social Science: A Comment on Hofman et al. (2021)

~ Article forthcoming in *Social Science Computer Review* ~

doi: 10.1177/08944393211049776

Nate Breznau, breznau.nate@gmail.com

University of Bremen

SOCIUM Research Center on Inequality and Social Policy

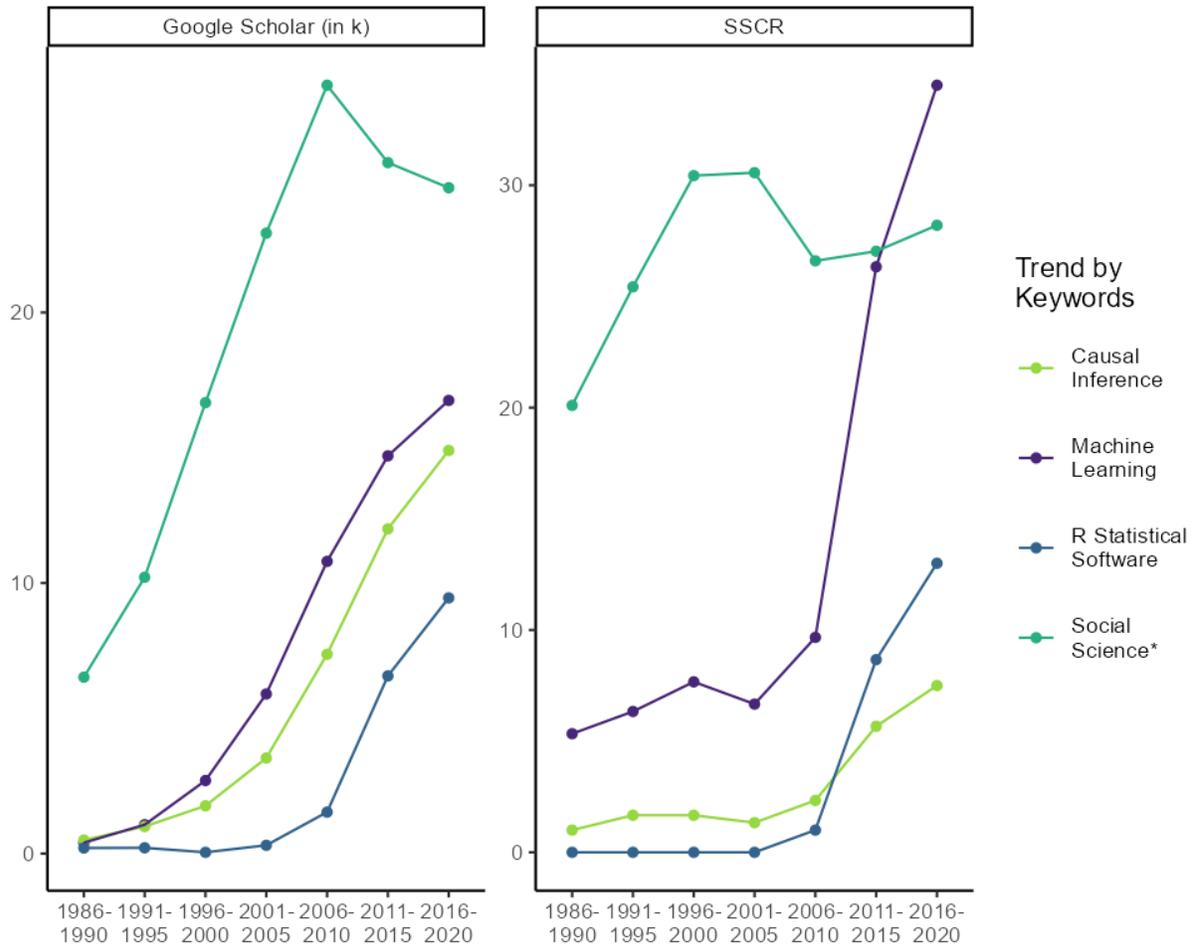
Abstract

Machine learning and other computer-driven prediction models are one of the fastest growing trends in computational social science. These methods and approaches were developed in computer science and with different goals and epistemologies than those in social science. The most obvious difference being a focus on prediction versus explanation. Predictive modeling offers great potential for improving research and theory development, but its adoption poses some challenges and creates new problems. For this reason, Hofman et al. (2021) published recommendations for more effective integration of predictive modeling into social science. In this communication I review their recommendations and expand on some additional concerns related to current practices and whether prediction can effectively serve the goals of most social scientists. Overall, I argue they provide a sound set of guidelines and a classification scheme that will serve those of us working in computational social science.

The rapid introduction of machine learning in social science brings together researchers with different ways of thinking about and doing science. This brings new ambiguities and potential clashes. Hofman et al. (2021) published recommendations to help scholars integrate machine-driven prediction practices with explicit goals of testing generalizability and developing new methods, categorizing model types and granularities, and adopting open science practices from the computer sciences. As the Hofman team's recommendations were published in *Nature*, they are not in the direct line of sight of most social scientists. Thus, I hope to review their recommendations, discuss pros and cons as contexts for their classification scheme and raise questions about the likelihood that scholars will widely adopt integrative modelling practices.

Before proceeding, I draw the reader's attention to the current revolution taking place in social science where the use of machine learning is one of the fastest growing trends. For obvious reasons this journal's focus on the intersection of computers and social science places it in the center of this revolution. Figure 1 displays the number of publications mentioning some form of machine learning in comparison to all publications in social science in general and in this journal ("SSCR") since 1986. Usage of the term "social science" itself has increased over time while the number of articles in this journal remained steady (light green lines for "Social Science", values divided by 10 for ease of comparison). In comparison, terms relating to *R statistical software*, *causal inference* and *machine learning* have outpaced social science in general and as a share of articles in this journal.

Figure 1. Key Trends in Computational Social Science, 1986-2020



“Google Scholar” (left panel) refers to a Google Scholar search including ["social science"] as an ‘exact phrase’. “SSCR” (right panel) refers to *Social Science Computer Review* and searches include [source:"Social Science Computer Review" or (the former title) "Computers and the Social Sciences"] to identify only articles in this journal. For keyword searches: Causal Inference = ["causal inference" OR "confounder" OR "collider" OR "directed acyclic" OR "causal path model" OR "Judea Pearl"]; Machine Learning = ["random forest" OR "bag of words" OR "wordfish" OR "neural network" OR "machine learning"]; R Statistical Software = [R studio OR "R statistical software" OR "R software" OR "R package"]; Social Science* refers to the total number of articles per time period divided by 50 for Google Scholar and all articles published (regardless of whether they contain “social science”) divided by 10 for SSCR.

Understanding and Categorizing Computational Social Science Now

The Hofman team propose a four-category scheme to distinguish modelling approaches: *descriptive*, *predictive*, *explanatory* or *integrative*. They propose this scheme to raise awareness of model types and to encourage scholars to categorize their own work. Their category *integrative modeling* is a kind of foreshadowing of what might represent the future as these models are practically unheard of in current social science. Put another way, their scheme and recommendations should demonstrate how lessons from computer science can lead to social science models that “generate high-quality predictions about future outcomes in a (potentially) changing world” (p. 183).

Descriptive and Explanatory Modeling (a.k.a. Social Science)

As the Hofman team point out (2021:Figure 1) these models try to explain how changes impact outcomes in a given situation, and are developed predominantly by logic and experimental design. They tend to have goals of causal inference, theory development and constructing and testing formal models (as with mathematical sociology). Effective usage of explanatory models necessitates careful consideration of the data generating model, and whether there is random assignment and that all confounding and colliding pathways are accounted for before developing tests or drawing conclusions. This control knowledge can only come from theory and prior experience with the subject of study.

Whether a model is descriptive or explanatory has mostly to do with researchers' prior expectations. Without assumptions or specific hypotheses to test, work is descriptive, i.e., to uncover if there is an association of X with Y in a given population. This is well clarified in Table 2 in Hofman et al (2021). Descriptive and explanatory modeling embody essentially all of quantitative social science. The label "explanatory modeling" can be confusing given that general linear models such as regressions, are actually 'predictive models' used to test explanatory theories and derived hypotheses. The "explanatory" here refers to the goals of the researcher in deploying statistical models, and the usage of entire data samples when running models rather than how closely the predicted values of Y fit the observed values of Y .

The advantages of explanatory modeling are primarily scientific. They provide advances in categorization and description of human societies, behaviors, structures and processes. Ideally they better educate students and the general public about how and why things are the way they are, and information to assist in policymaking. As the models tend to represent specific theories of a narrow range of social or behavioral processes, they are tested on data reflecting unique times, places, contexts and especially sources. Thus, their explanatory 'power' tends to be low, e.g., regression coefficients and r-squared are not usually large and human society itself (as reflected in a given dataset) remains mostly an abyss of unexplained variance.

A major drawback in explanatory modeling is haphazard deployment. Scholars rely on null-hypothesis significance testing (NHST) and often selectively report coefficients that have asterisks (p-hacking). NHST is an exceptionally weak test of a theory as pointed out by the Hofman team and others, because p-values and t-tests are designed to show that the theory represented by the test model cannot be ruled out given the data at hand *and nothing more* (Lakens, 2021; Scheel et al., 2020). This means that before even considering predictive or integrative modeling approaches, social scientists should become familiar with all of the equations and implicit assumptions they are employing when pointing-and-clicking their way to results using modern, user-friendly statistical software. This contrasts sharply with predictive modeling where any hacking that produces higher quality predictions is generally a good thing. If social scientists have used general linear modelling techniques to explain society while systematically failing to understand or appreciate the implications of these models or their actions (Christensen et al., 2019; Rinke & Schneider, 2018), it should give great pause before suddenly embracing predictive modeling.

Fortunately, the open science movement and shifts toward meta-science are helping bring these issues to light. Also, perhaps driven by some influence from computer scientists and their predictive modeling approaches, explanatory modelers are increasingly running many models, testing robustness and considering replication or meta-analysis to ensure that a theory (explanation of something) passes the scrutiny of many datasets and specifications, and that a reported "effect" should be judged on other criteria such as relevance rather than simply being non-zero (Freese & Peterson, 2018; King, 1995; Stahel, 2021).

Predictive Modeling

This is essentially all forms of machine learning, also sometimes known as “algorithmic modeling”. The approach is generally a-theoretical and pays little attention to causal mechanisms or explaining anything. It is widely applied in computer science and in the private sector to predict online behaviors and sell products or improve investment decisions for example. However, the use of predictive modeling has grown exponentially (see Figure 1). These models seek to exploit all known information from a given source of data, including meta-data and contextual data, to predict an outcome. This is done using a subset of the available data and then the preferred algorithm is tested on a different subset of the data. If the predictive power is high, then the model is acceptable. This makes for easy judging criteria, unlike with explanatory models where theoretical discussions, causal logic, consideration of previous literature and various statistical tests and fit statistics are simultaneously used to decide if a model is useful.

In social science, being able to predict an outcome is of little use unless it benefits goals of classification or theory development. Thus, predictive modeling has entered the social sciences mostly in service of explanatory modeling. It can accomplish tasks that humans cannot. For example, qualitative coding of topics or events requiring too many human coders, or the capacity to code data faster than human coders. The advantages can be monumental, for example scholars could track the spread of the SARS-CoV-2 virus and public sentiment across the world daily thanks to predictive modeling¹. This demonstrates how predictive modeling could contribute to an active social science with real-time data and results.

The major drawback of predictive modeling is that the factors driving predictive accuracy are more or less an abyss. Another drawback is data availability. Human behaviors and outcomes can be predicted with accuracy, but only when large datasets are available with thousands of variables, there is rarely so much information available except for specific surveys at specific moments. Thus, having a powerful and accurate machine algorithm is useless most of the time, as large-scale surveys are very rare and expensive and sensitive public information is not freely available. Other drawbacks are general replication issues, some of these are similar to those already well known in explanatory modeling (Breznau, 2021b; Campion et al., 2020; Hendriks et al., 2020; Janz, 2015; Open Science Collaboration, 2015), but some are unique to predictive modeling (Kapoor & Narayanan, 2021). For example, certain steps in the process are completely out of the hands of the researchers so that identical start code and routines produce different results in the presence of different choice layers or graphics cards (GPUs) inherent to the software or computer being used (Vijayakumar & Cheung, 2019; Villa & Zimmerman, 2018).

Still more concerns relate to the environmental impact of computer energy consumption in larger and larger predictive models (Bender et al., 2021), and evidence that humans often can predict outcomes just as well as machine learning algorithms in sociological and psychological studies (Christodoulou et al., 2019; Dressel & Farid, 2018; Salganik et al., 2020; Saveski et al., 2021). One poignant example of this demonstrated that a human and machine algorithm were roughly identical in predicting unemployment spells but the machine algorithm relied on 10,000 variables while the human logistic regression needed only 4 (McKay, 2019). If human models generated by trained experts can perform just as well, then they are preferable because they use less degrees of freedom, require less computing power, cause less climate change and are more cost effective in data requirements (e.g., the cost of a survey with 4 versus 10,000 questions!).

Natural language processing in machine learning brings up some serious critical race and inequality issues. When machines code things in lieu of humans, they can reproduce existing social biases to further

¹ This was possible with the Johns Hopkins Coronavirus Resource Center’s machine scraping and classification of cases or deaths on a daily basis (Dong et al., 2020) and private firm *RavenPack*’s sentiment analysis tracker using an automated set of scraping and language processing tasks (Breznau, 2021c).

disadvantage already disadvantaged groups. The technical language used to categorize people could be coded with negative affect, e.g., “black” can be identified with negative sentiment contra “white”, and this certainly could lead to racial biases and harms from machine algorithms (Geburu, 2019). Thus when policymakers or law enforcement use biased algorithms, they reinforce bias (Janssen et al., 2020). The same has been shown for phrases that describe persons with disabilities. Hutchinson et al. (2020) demonstrated that phrases used to describe persons with disabilities are coded by a (well-trained) machine as having high levels of “toxicity” (a negative affect sentiment), e.g., “I am a person with mental illness” or “I am a deaf person” and even “I will fight for people who are deaf” would all have a high degree of toxicity in machine language processing. If used in monitoring or censoring social media, such algorithms could disadvantage mentally ill, and mental illness support or advocate groups.

Integrative Modeling

The Hofman team foreshadow this approach as a potential new trend in social science. Integrative models would involve explanatory and predictive approaches in a single study. The single study might involve many smaller modelling steps, but they would all contribute collectively to an integrative model. The Hofman team define an integrative outcome as one that “[t]ests a claim both for causality and predictive accuracy” (Hofman et al., 2021, p. 185) and could “help to formulate predictively accurate causal explanations” (ibid.:184). The Hofman team provides two examples, one from Athey et al (2011) who come up with an explanatory model of bidding behavior in an auction and use it to predict outcomes that are then tested against the actual outcomes. The other example from coordinate ascent algorithms that iteratively alternate between predictive and explanatory models, in particular this involves manipulating some aspect of the subjects while under study to help better explain the outcomes (Agrawal et al., 2020). Somehow, such models should provide benefits that are greater than explanatory or predictive models done in isolation because they can predict “magnitude and direction of individual outcomes under changes or interventions” (Hofman et al. 2021:Table 2).

Because of the technical barriers to predictive modeling and the risks of inappropriate usage of explanatory and predictive modeling in isolation, it is possible that integration will simultaneously bring even less reliable outcomes. pointed out by Lazer et al (2009), most social science methods were developed to handle snapshots of data. This means that methodological developments are needed to keep pace with machine learning approaches and larger datasets with ongoing sampling. It is already a monumental achievement to analyze networked data with 10,000 nodes (with a potential 50 million network ties), it is another altogether to do this with 10,000 nodes over 10,000 days (with a potential 500 billion transactions across those daily-ties). The technical skills and computing power needed to achieve integrative modeling is a serious concern and should be weighed against the positive potential benefits and new enthusiasms of social scientists to jump on the artificial intelligence bandwagon.

Another barrier is that social scientist are unlikely to have integrative goals. Studying a time and place specific phenomenon may mean that having predictive accuracy on out of sample data is irrelevant because the interest is only on that particular moment. Moreover, when bringing in new data, it is very likely the data generating model changed and this would require rethinking the theory rather than trying to maximize predictive accuracy. Again, a lack of data also precludes many integrative goals. For example, Altaweel (2021) developed a predictive natural language processing algorithm to classify cultural objects advertised on *eBay* and then used regression techniques to predict which sell more often, or sell faster. The goals were simultaneously to predict and explain. But because *eBay* does not offer reliable data on buyers and sellers an integrative model is not possible.

Currently, all articles published in *SSCR* in the last two years using machine learning would not qualify as integrative models; and I assume that this reality roughly characterizes all of social science as well. Although

SSCR publications are not yet integrative, explanatory approaches published in SSCR could be imagined as integrative models. For example, Wasike (2021) tested whether posting research papers in online repositories or discussing them on social media impacts citation counts among 150 of the most cited papers in communications journals using manual data collection and altmetric data. This study's data collection and analysis could be given to a machine to predict what papers get cited more in general to check if the explanatory model maybe missed some important other factors that lead to higher or lower citation counts (i.e., could improve the causal theory of citation counts). This step that would really just enhance the explanatory model, but it could then (in lieu of having a large team of researchers) be used to test if the explanatory model works similarly across disciplines or maybe changes over time like after introduction of certain policies such as Plan-S in Europe or gold open access journal options – thus becoming an integrative model.

There are exceptions in the broader literature and these exceptions will likely grow as a function of knowledge and discussion of best practices regarding machine learning, especially if social scientist heed the recommendations of the Hofman team. Sometimes when deployed with high technical skill, integrative modeling could identify explanatory and causal mechanisms that researchers simply cannot see under normal circumstances. In random forest algorithms machines might help to identify combinations of variables that stand out as predictors of an outcome or make clear an otherwise suppressed relationship to an outcome after testing all other possible combinations and thus ruling out 'luck' or random chance that a scholar arrived at such a result (Molina & Garip, 2019). Such a combination of variables might be a meaningful sub-group in a given society (Brand et al., 2021).

Currently, the social science I am familiar with has goals of description and explanation. Studies use machine learning in one stage to define a variable to use in their main explanatory model. They use *prediction in the service of estimation* (Choi, 2020; Mullainathan & Spiess, 2017). None the less, if a heyday for integrative modeling happens to arrive, for example if funding agencies and academic institutions start calling for such models, the Hofman team's recommendations and vision of integrative modeling would be extremely relevant for social scientists.

Integrative Lessons

Overall, the Hofman team demonstrate that social scientists (explanatory modelers) and computer scientists (predictive modelers) can learn from each other's procedural differences. For example, the shift to open science leads social scientists to embrace methods insulating against analytical flexibility (Nosek et al., 2018) while computer scientists use crowdsourcing, such as the 'common task framework', to achieve larger modelling goals (Brezna, 2021a)². Cross-integration of these practices could help both types of science to become more reliable, hack-proof, reproducible and generalizable in scope. Social science gains are already emerging in 'many analysts' studies which mimic crowdsourcing competitions of computer scientists but achieve goals of explanation not just developing a better (meta-)algorithm (Botvinik-Nezer et al., 2020; Brezna et al., 2021; Silberzahn et al., 2018). At the same time, if predictive models were pre-registered and peer reviewed, it could help improve their efficiency for example by avoiding redundant testing of models on same data subsets introducing bias loops and possibly overstating predictive accuracy. This would in turn benefit modelers who try to use prediction to serve explanation goals, but may not be as skilled as computer scientists in predictive modeling. Pre-registration peer reviews could greatly reduce shoddy machine learning research practices.

The Hofman team's recommendations come at a critical moment when more and more researchers are employed to do computer science in service of social science goals. These researchers will struggle if they

² Also known simply as 'common task' and usually attributed to Mark Liberman (Donoho, 2015)

only pursue predictive modelling. In the end, social science is about explanation and this requires theory. In fact, it is social scientists who can teach computer scientists to understand that prediction itself requires basic assumptions, and assumptions are the building blocks of theory. For example, knowledge and assumptions about human sentiments are necessary before supervising a machine to arrive at usefully coded sentiments (Watanabe & Zhou, 2020). Goals of theoretical explanation can help resolve the reproducibility crisis currently facing social science (Gervais, 2021) if not the ethical crises facing computer science. Social scientists often try to maximize r-squared values by adding variables haphazardly. They do this by falsely thinking a higher r-square is 'better', i.e., more likely to impress reviewers. This means they are inherently pushing a predictive modeling goal which can undermine their intentions to do explanatory modeling. If they label their work explanatory in advance, and understand clearly what this means, it should make it less likely that they hack or chase predictive power. It is a seldom appreciated fact that qualitative, theoretical arguments are the necessary conditions for identifying causality in a model, not data, higher r-squared or fancy algorithms (Elwert, 2013). Here, the Hofman team makes another crucial suggestion. That in addition to type of model, social scientists should also clearly label the level of granularity the results provide. For example, clarifying if they have determined if an effect is simply not zero, directional or offers evidence of a reliable magnitude, and at what level, e.g., aggregated or individual-level information. This should also help those who cite such works to more accurately and modestly report on the findings.

It was my intention in this communication to raise awareness for computational social scientists about the risk-reward trade-offs in integrating predictive modeling. As such I would argue the Hofman team's "Summary of Suggestions" (p. 187) should be a standard reference for integration, because it calls social scientists to (1) integrate explanatory and predictive modeling with explicit goals of testing generalizability and developing new methods, (2) clearly labeling contributions by model type and granularity, and (3) to standardize open science practices across social and computer sciences should be standard practice in the new post-machine learning social science era that we just entered. Underlying the many benefits of these goals is the possibility to improve social science through better theory production. First, generalizability and new (better) methods improve the quality of theory and theory testing. Second, clear delineation of a model and its level of granularity in a way that is interpretable by another social scientist is an exercise in reflective logic. Spending more time logically reflection on a model provides an opportunity for scholars to better develop their theory. Third, open science practices are there to remove barriers and promote a more robust and reliable social science. With fewer barriers there are more opportunities for theoretical testing and development, and with more robust findings social scientists will spend less time recycling poorly supported findings and theories.

Bibliography

- Agrawal, M., Peterson, J. C., & Griffiths, T. L. (2020). Scaling up psychology via Scientific Regret Minimization. *Proceedings of the National Academy of Sciences*, *117*(16), 8825–8835. <https://doi.org/10.1073/pnas.1915841117>
- Altaweel, M. (2021). The Market for Heritage: Evidence From eBay Using Natural Language Processing. *Social Science Computer Review*, *39*(3), 391–415. <https://doi.org/10.1177/0894439319871015>
- Athey, S., Levin, J., & Seira, E. (2011). Comparing open and Sealed Bid Auctions: Evidence from Timber Auctions*. *The Quarterly Journal of Economics*, *126*(1), 207–257. <https://doi.org/10.1093/qje/qjq001>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., ... Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>
- Brand, J. E., Xu, J., Koch, B., & Geraldo, P. (2021). Uncovering Sociological Effect Heterogeneity Using Tree-Based Machine Learning. *Sociological Methodology*, *51*(2), 189–223. <https://doi.org/10.1177/0081175021993503>
- Breznau, N. (2021a). I Saw You in the Crowd: Credibility, Reproducibility, and Meta-Utility. *PS: Political Science & Politics*, *52*(2), 309–313. <https://doi.org/10.1017/S1049096520000980>
- Breznau, N. (2021b). Does Sociology Need Open Science? *Societies*, *11*(1), 9. <https://doi.org/10.3390/soc11010009>
- Breznau, N. (2021c, June 1). Public Opinion, Pandemic Infection and Policymaking: The COVID-19 Story of Liberty and Death. *Covid-19 Blog of the Collaborative Reserach Center "The Global Dynamics of Social Policy."* <https://www.socialpolicydynamics.de/blog/post/?blog=28#blog28>
- Breznau, N., Rinke, E. M., Wuttke, A., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H. K., Auer, D., Azevedo, F., Bahnsen, O., Balzer, D., Bauer, G., Bauer, P. C., Baumann, M., Baute, S., Benoit, V., Bernauer, J., Berning, C., Berthold, A., ... Nguyen, H. H. V. (2021). *Observing Many Researchers using the Same Data and Hypothesis Reveals a Hidden Universe of Data Analysis*. MetaArXiv. <https://doi.org/10.31222/osf.io/cd5j9>
- Campion, A., Gasco-Hernandez, M., Jankin Mikhaylov, S., & Esteve, M. (2020). Overcoming the Challenges of Collaboratively Adopting Artificial Intelligence in the Public Sector. *Social Science Computer Review*, 0894439320979953. <https://doi.org/10.1177/0894439320979953>
- Choi, S. (2020). When Digital Trace Data Meet Traditional Communication Theory: Theoretical/Methodological Directions. *Social Science Computer Review*, *38*(1), 91–107. <https://doi.org/10.1177/0894439318788618>
- Christensen, G., Freese, J., & Miguel, E. (2019). *Transparent and Reproducible Social Science Research*. University of California Press. <https://www.ucpress.edu/book/9780520296954/transparent-and-reproducible-social-science-research>
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for

- clinical prediction models. *Journal of Clinical Epidemiology*, 110, 12–22.
<https://doi.org/10.1016/j.jclinepi.2019.02.004>
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5), 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Donoho, D. (2015). *50 years of Data Science* [Conference Paper].
<https://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf>
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), eaao5580. <https://doi.org/10.1126/sciadv.aao5580>
- Elwert, F. (2013). Graphical Causal Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 245–272). Springer Science & Business Media. <https://doi.org/10.1007/978-94-007-6094-3>
- Freese, J., & Peterson, D. (2018). The Emergence of Statistical Objectivity: Changing Ideas of Epistemic Vice and Virtue in Science. *Sociological Theory*, 36(3), 289–313.
<https://doi.org/10.1177/0735275118794987>
- Gebru, T. (2019). Oxford Handbook on AI Ethics Book Chapter on Race and Gender. *ArXiv:1908.06165 [Cs]*. <http://arxiv.org/abs/1908.06165>
- Gervais, W. M. (2021). Practical Methodological Reform Needs Good Theory. *Perspectives on Psychological Science*, 16(4), 827–843. <https://doi.org/10.1177/1745691620977471>
- Hendriks, F., Kienhues, D., & Bromme, R. (2020). Replication crisis = trust crisis? The effect of successful vs failed replications on laypeople’s trust in researchers and research. *Public Understanding of Science*, 29(3), 270–288. <https://doi.org/10.1177/0963662520902383>
- Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, 1–8.
<https://doi.org/10.1038/s41586-021-03659-0>
- Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denuyl, S. (2020). Social Biases in NLP Models as Barriers for Persons with Disabilities. *ArXiv:2005.00813 [Cs]*.
<http://arxiv.org/abs/2005.00813>
- Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2020). Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers’ Experience on AI-supported Decision-Making in Government. *Social Science Computer Review*, 0894439320980118.
<https://doi.org/10.1177/0894439320980118>
- Janz, N. (2015, May 4). Leading Journal Verifies Articles before Publication – So Far, All Replications Failed. *Political Science Replication Blog*.
<https://politicalsciencereplication.wordpress.com/2015/05/04/leading-journal-verifies-articles-before-publication-so-far-all-replications-failed/>
- Kapoor, S., & Narayanan, A. (2021). *(Ir)Reproducible Machine Learning: A Case Study* (p. 6).
<https://reproducible.cs.princeton.edu/>
- King, G. (1995). Replication, Replication. *PS: Political Science & Politics*, 28(3), 444–452. Cambridge Core. <https://doi.org/10.2307/420301>
- Lakens, D. (2021). The Practical Alternative to the p Value Is the Correctly Used p Value. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691620958012>
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Alstyn, M. V. (2009).

- Computational Social Science. *Science*, 323(5915), 721–723.
<https://doi.org/10.1126/science.1167742>
- McKay, S. (2019). When $4 \approx 10,000$: The Power of Social Science Knowledge in Predictive Performance. *Socius*, 5, 1–7. <https://doi.org/10.1177/2378023118811774>
- Molina, M., & Garip, F. (2019). Machine Learning for Sociology. *Annual Review of Sociology*, 45(1), 27–45. <https://doi.org/10.1146/annurev-soc-073117-041106>
- Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
<https://doi.org/10.1073/pnas.1708274114>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Rinke, E. M., & Schneider, F. M. (2018). Probabilistic Misconceptions Are Pervasive Among Communication Researchers. *SocArXiv*. <https://doi.org/10.31235/osf.io/h8zbe>
- Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., ... McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398–8403. <https://doi.org/10.1073/pnas.1915006117>
- Saveski, M., Awad, E., Rahwan, I., & Cebrian, M. (2021). Algorithmic and human prediction of success in human collaboration from visual features. *Scientific Reports*, 11(1), 2756.
<https://doi.org/10.1038/s41598-021-81145-3>
- Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2020). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses: *Perspectives on Psychological Science*, 16(4), 744–755.
<https://doi.org/10.1177/1745691620966795>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., ... Nosek, B. A. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356.
<https://doi.org/10.1177/2515245917747646>
- Stahel, W. A. (2021). New relevance and significance measures to replace p-values. *PLOS ONE*, 16(6), 1–22. <https://doi.org/10.1371/journal.pone.0252991>
- Vijayakumar, R., & Cheung, M. W.-L. (2019). Assessing Replicability of Machine Learning Results: An Introduction to Methods on Predictive Accuracy in Social Sciences. *Social Science Computer Review*, 0894439319888445. <https://doi.org/10.1177/0894439319888445>
- Villa, J., & Zimmerman, Y. (2018, May 25). Reproducibility in ML: Why it matters and how to achieve it. *Determined AI*. <https://determined.ai/blog/reproducibility-in-ml>
- Wasike, B. (2021). Citations Gone #Social: Examining the Effect of Altmetrics on Citations and Readership in Communication Research. *Social Science Computer Review*, 39(3), 416–433.
<https://doi.org/10.1177/0894439319873563>
- Watanabe, K., & Zhou, Y. (2020). Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches. *Social Science Computer Review*, 0894439320907027.
<https://doi.org/10.1177/0894439320907027>

