



**HAL**  
open science

## Sociology's Stake in Data Science

Philipp Brandt

► **To cite this version:**

Philipp Brandt. Sociology's Stake in Data Science. Sociologica, 2022, 16 (2), pp.149-166.  
10.6092/issn.1971-8853/13434 . hal-03864431

**HAL Id: hal-03864431**

**<https://sciencespo.hal.science/hal-03864431v1>**

Submitted on 21 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Sociology's Stake in Data Science

Philipp Brandt\*

Department of Sociology and Center for the Sociology of Organisations, Sciences Po, Paris (France)

Submitted: August 9, 2021 – Revised version: July 15, 2022

Accepted: July 19, 2022 – Published: October 17, 2022

## Abstract

Data scientists gave sociologists pause when they started disturbing social life and research. This article considers three instances where data science made inroads into the sociology jurisdiction. Instead of calling for a defense, they reveal opportunities for sociological research in the digital age. These opportunities build on the data-analytic thinking that undergirds the discipline's more salient structures and conventions. They recall old sociological intuitions and pragmatist theory that conceptualize the research process in a way that leaves room for novel observations. From this perspective, data science can help integrate sociology around new problems and shared principles and enlarge it by introducing its ideas to different audiences.

**Keywords:** data science; Dewey; computational social science; digital transformation; reflexivity.

## Acknowledgements

The author acknowledges two *Sociologica* reviewers and the editors for detailed comments and suggestions on earlier drafts of this article.

---

\* ✉ [philipp.brandt@sciencespo.fr](mailto:philipp.brandt@sciencespo.fr)

The digital transformation shook everyday life and social scientific research. Sociologists reacted quickly, studying how those changes shaped political opinions and working conditions, and introduced new forms of discrimination (e.g., Bail, 2021; Christin, 2020; Eubanks, 2018; Noble, 2018). Commentators also identified dangers and opportunities for social sciences, often encountering a curious novelty during their assessments: “data scientists,” the digital-era experts interfering with democracy, public services, and other social institutions by using machine learning techniques as well as quantitative methods that sociologists have used for a long time.<sup>1</sup>

What does the appearance of data scientists mean for the discipline of sociology? Reflections on direct encounters have noted some problems and a few promising overlaps. David Ribes (2019) pointed out that data science and science and technology studies both acknowledge the interplay of social and technical dynamics and question boundaries. While these commonalities suggest productive collaborations, Noortje Marres (2017) warned that the results of data science work obscure sociological observations. Matt Salganik (2018) considered new computational ideas, good and bad, for methodological procedures and ethics around digital social data analyses. These accounts had different concerns, but they all showed practical issues data science has raised.

Others started viewing data scientists as a new professional role, if one that is still emerging, heterogeneous, and at times ambiguous (Avnoon, 2021; Börner et al., 2018; Dorschel & Brandt, 2021; Lohr, 2015; Mützel et al., 2018). Besides revealing new conceptual puzzles, this perspective raises questions about data science’s impact on the jurisdiction of data-analytic work, which saw statisticians launch a defense of their discipline early on (Donoho, 2017).<sup>2</sup> I propose a response for sociology that does not keep data science at arm’s length, whether as a supplier of new techniques or a competitor for work. Shared experiences give sociology a stake in data science, which we can leverage for positioning the discipline as the digital transformation alters how we observe the social world.

Sociology’s past has shown that the simple adoption of new techniques quickly backfires. One vivid episode involved Paul F. Lazarsfeld’s radio research project that, like much of data science, analyzed the effects of communication campaigns in collaborations with government offices and private firms (González-Bailón, 2017, p. 52; Katz & Katz, 2016). Peers denounced these activities, leaving the field with memories of a dispute around high-minded theory and entirely practical empiricism (Morrison, 1978), experiences we still process today (Katz & Katz, 2016). The field’s fragmentation has only intensified since then (Abbott, 1998; 2001, ch. 1). Such infighting is not just uncomfortable. If continued, sociology might become a case of its own theories, “dividing” itself for others to “conquer.”<sup>3</sup>

The early divisions partly reflected the technological conditions of their time. Lazarsfeld, his collaborators, and their early successors applied formal techniques from other fields to sociological research problems. They worked as a group of specialists who transmitted their technical expertise from one generation to the next (Abbott, 1998, pp. 166–167). That work required substantial resources. Publications from the 1970s that used computational analyses relied on access to institutional computing facilities (e.g., White et al., 1976). By the 1980s, similar work

1. This working definition of data science synthesizes ideas from the data science community and the academic literature on data science, cited throughout this article. A multi-year research project that involved field observations and quantitative analyses of data science’s emergence has informed my reading of these discussions.
2. I use jurisdiction in the sense of Abbott (1988, p. 20) as “the link between a profession and its work,” which he viewed as the “central phenomenon of professional life.”
3. This remark refers to the old idea of “divide and conquer,” for which sociologists often cite Simmel (1908).

offered readers to inquire about data and software with a request that should include checks to cover costs for duplication and mailing (e.g., Burt, 1987).<sup>4</sup> This early computational research was a slow and secluded affair.

Today, computational research is quick, iterative, and vastly open. Classic datasets ship with R, a widely used open-source software that comes as a free download (R Core Team, 2021).<sup>5</sup> Additional datasets are available online, ranging from long-running survey datasets to more specialized social networks (e.g., Leskovec & Krevl, 2014; NORC, 2021; SIENA, 2022). Academics can request privileged access to popular social media platforms and collect records of major social events as they unfold (Twitter Inc., 2022). Code for analyzing these records and datasets is also available, often in repositories from courses that teach the relevant skills (e.g., Nelson, 2022; Vedres, 2022). New textbooks help students think through the more complex issues around these new resources (e.g., Healy, 2018; Salganik, 2018). And many sociologists who do quantitative research have taken advantage of these technical changes. But although the dynamics have changed, creating new opportunities for new connections in the discipline, the old fault lines often persist.

The diffusion of new technical skills needs reflection to avoid reinforcing sociology's existing divides,<sup>6</sup> an exercise that gets regular endorsements but hasn't become standard practice (e.g., Gouldner, 1970; Romero, 2020). Already a century ago, Weber's (2004) analysis of sociology "as a vocation" warned of the misconception that "science has become a question of simple calculation" (p. 8). Although he insisted on the importance of calculation and specialization in research projects, moving beyond the practical concerns, Weber revealed what he considered "the decisive factor, namely, 'inspiration'" (p. 8). This "inner vocation" is impossible to formalize, which increases the risk for a mature discipline like sociology to look past it. Weber recovered it following a comparison of two academic systems. Today, data science, far from an institutionalized profession, offers a point of reflection on how sociology can take advantage of recent changes for its project of explaining the social world. Weber's ideas indicate a direction that avoids further divisions in the field.

This article considers opportunities for the discipline of sociology in the digital era. While disciplinary conventions and procedures often provide rigor, certainty, and professional identity, recent instances and sociological research itself suggest that they sometimes undermine new ideas and insights. The combination of data science's overlapping concerns and an initial lack of these conventions recalls the "inner vocation for science" (Weber, 2004, p. 7). Data science's exposure to the uncertainty of an emerging set of problems (Hammerbacher, 2009), often an unsettling experience (Ibarra, 1999), points at some challenges. But sociological ideas predict that these experiences can produce new ideas if met collectively (Dewey, 1916, 1939).

---

4. Burt (1987) asked for \$10 for his dataset and \$25 for his software.

5. The most important programming language for data science is a popular point of contention. Depending on the camp, either Python or R take the top spot. But most serious data scientists agree that no single software suffices. At a minimum, this work requires a separate database programming language, like SQL. Most also agree that the specific choice and combination of programming languages depends on the problem and purpose.

6. See for example the dispute between Watts (2014) and Turco & Zuckerman (2017) on sociology's future in the computational age.

## 1 Observations of the Social World as a Jurisdictional Challenge

Data science's rise a decade ago motivated David Donoho, a prominent statistician, to launch a defense of his field (Donoho, 2017). Donoho argued that data science ideas were much older than the popular label suggested. He listed several classical statisticians, described their accomplishments, and concluded that recent ideas for data science were illegitimate because they had practical instead of purely scientific concerns. This section reverses Donoho's strategy. It summarizes three data science instances that made sociological observations outside the discipline of sociology. The next section returns to these examples to identify new opportunities for sociology.

The first of the three instances took place at LinkedIn and quickly became a standard reference in popular and academic discussions of data science's emergence. The second instance, the invention of Google, was not affiliated initially with data science. But more important than the label, the intuitions behind Google's original algorithm reflected patterns that characterize early data science work. The third instance captures a data science response to the kind of pushback data science started receiving soon after it gained prominence. These examples reflect prominent data science applications more than the full range. This selection allows the discussion to depart from typical discussions of the practical consequences or technical underpinnings. It highlights data-analytic ideas in situations with little guidance.

One of the first prominent discussions of data science was about data science's arrival at LinkedIn and how the first data scientist there proposed an idea. He wanted to consider existing relationships between users to suggest a user's friends' connections that were not yet that user's friends as new connections (Davenport & Patil, 2012). This idea may have been new to LinkedIn, but it echoed Simmel's (1908) and Heider's (1958) classic notions of triads and transitivity. That legacy played no role in the data science story, and the data scientist's background in physics, which was central, offered little indication that exposure to the discipline of sociology shaped his work, even if the original idea played some role. The lack of systematic engagement is not only frustrating; it also overlooks the more nuanced findings from recent sociological research. But this and the following examples recall the connection of sociological ideas to detailed observations of a social setting rather than to disciplinary boundaries.

Much of data science invites dismissive views of its preoccupation with arranging vast datasets without deeper questions in mind. But careful data-analytic thinking, the reflection on a problem against the backdrop of a technical operationalization, still matters. Consider Google's beginning as a second data science instance. Today, Google dominates the data economy, often using data science (Noble, 2018). But it started as a student project that its creators prepared for an academic conference (Brin & Page, 1998). Sergey Brin and Larry Page (1998) introduced their idea with the observation that "as of November 1997, only one of the top four commercial search engines finds itself" (p. 108). Brin and Page believed they could build a search engine without such embarrassing limitations by using "The citation (link) graph of the Web." They proposed that "These maps allow rapid calculation of a Web page's 'PageRank,' an objective measure of its citation importance that corresponds well with people's subjective idea of importance" (p. 109). Social scientists started to think carefully about people and the web (e.g., Castells, 1996; Wellman, 1997), and they had pioneered the mathematical ideas underlying the PageRank measure decades before (Bonacich, 1972, 1987).<sup>7</sup> But those ideas did not concern Brin and Page's discussion of what users would find important. Long before Google began encapsulating internet users in a web of algorithms that

7. I thank one anonymous reviewer for pointing out this connection.

shape the browsing experience (Noble, 2018), it used sociological intuitions in an attempt to help users navigate the internet's information web.

Data science has moved beyond purely practical views of data applications. This third instance involves the accommodation rental platform Airbnb, an icon of the digital age that has quickly come to threaten global hotel chains and local residential communities. While the former had signed up for competition in markets, the latter suddenly found themselves in an unequal battle without a cultural or institutional framework. Airbnb kept competing with corporate accommodation providers, but responded to the local damages. They used the same algorithmic infrastructure that has brought the masses into touristic hotspots to divert them elsewhere. An article in *The New Yorker* explained,

Data-analytics software [can] identify parts of the world that are starting to attract interest from visitors, and these destinations are then recommended to other adventurous travellers, through a promotional campaign titled Not Yet Trending. Recent picks include Xiamen, a coastal city in China opposite Taiwan; the Outer Hebrides, in Scotland; and Uzbekistan (Mead, 2019).

This solution echoes deeply sociological intuitions for cumulative advantage mechanisms in global inequalities and even more technical ideas about how new connections, which more likely follow from places not yet trending, produce new ideas (Vedres & Stark, 2010).

The summaries only scratched the surface of the underlying technical work. All major tech companies today also employ data scientists with sociological training, as well as qualitative researchers who introduce reflexive perspectives, if only in response to crucial public pushback.<sup>8</sup> The next section considers some of their findings. But whether or not the occasional sociological idea already snuck into those early data science applications, their development outside the discipline offers a point of reflection.

Traces of sociological ideas across the three vignettes are not surprising if we remember sociology's inception. Like the examples above, what is now "classical theory" originated in an array of empirical observations with limited disciplinary affiliation. In *Science as a Vocation*, Weber (2004) introduced himself as a political economist (p. 1). And he was not just the theorist as which we often remember him today. Besides his thorough historical research, Weber engaged in fierce debates over the interpretation of surveys of workers (Lazarsfeld & Oberschall, 1965). Similarly, Durkheim (1893, 1897) calculated suicide ratios and used indicators that relied on legal texts to analyze the division of labor. W.E.B. Du Bois developed data visualizations that described black populations across the U.S. using techniques that were far from standard practice in the discipline at the time (Battle-Baptiste & Rusert, 2018). Rather than revealing a competitor in data science, which it may be for statistics, these patterns suggest that considering data science can help recall some of our discipline's defining characteristics.

## 2 Discipline and Practice

The data science instances showed sociological intuitions in emergent settings. Sociology started out in similarly emergent settings but has since set up safeguards to manage the uncertainty they bring along. These safeguards include topical debates, a canon, rhetorical

8. Perhaps the most prominent ones are Facebook's "emotional contagion" experiment in 2014 and Cambridge Analytica's role in the 2016 U.S. presidential election. And data scientists recognize less publicly visible instances of poor practice (e.g., O'Neil, 2016, Ch. 1).

conventions, design standards, and output formats. The institutionalization of science has had abundant benefits for knowledge production but can also get in the way of new observations and discoveries (Ben-David, 1971). Data science offers points of reflection for navigating that tension for sociological insights.

## 2.1 Debates

Sociologists seek to advance debates about issues such as class and inequality, education, movements, markets, or the state. But these debates have their own social dynamics, another topic of ongoing debate (e.g., Abbott, 2001; Kuhn, 1970). For example, sociological research has revealed distinct trajectories whereby debates reach consensus (Shwed & Bearman, 2010), shown that they consistently favor canonical contributions (Barabasi & Albert, 2001; Merton, 1968) and that growth comes at the cost of specific ideas (McMahan & McFarland, 2021). All these findings indicate social processes that shape the quest for truth.

Debates remained peripheral to the data science examples, which still engaged in a broader discourse. Brin and Page contributed to concerns with search engine designs, an early data science scandal provoked academic reactions, and Airbnb introduced its new algorithm amid a public debate about the impact of digital services on local communities. These discussions were still in their infancy and lacked the legacies and nuances of most academic debates. Data science instances then indicate that a concern with immediate observations can address problems even outside of close collective guidance.

## 2.2 Classics

We also pay attention to those who came before us, as I did when citing Weber, Du Bois, and others. Arthur Stinchcombe (1982) listed several good reasons for this recognition, ranging from finding hypotheses to signaling a line of thought. These are productive motives, but, in practice, memory is often murky, even if it draws on written records. We overlook and forget half a legacy here (e.g., Mützel & Kressin, 2020, for Simmel), turn complex ideas into catchy punchlines there (see Granovetter's embeddedness view in Krippner et al., 2004), or leave out the empirical foundations of theoretical ideas we like (Ollion & Abbott, 2016). Intellectual traditions are important for continuous knowledge production, but social mechanisms undermine that promise without careful reflection.

Classics featured in the data science examples but were not salient. Brin and Page cited Jon Kleinberg (1999), who, not yet a household name at the time, had provided an authoritative summary of the ideas they worked with. Although the Airbnb and LinkedIn descriptions did not indicate the recognition of ancestors, Thomas Bayes, R. A. Fisher, and other quantitative thinkers make occasional appearances in these discussions. These luminaries don't come with particular hypotheses, however, or specific research directions. The examples still showed familiar patterns in their translation of the empirical situations in front of them into substantive ideas. These intuitions require closer engagement with relevant canonical ideas to lead to meaningful insights, but the initial ignorance sharpened the eyes for new observations.

## 2.3 Rhetoric

Besides recognizing others, whether peers or classics, we present our arguments in distinctive rhetorical styles. Some highlight "puzzles," others find "mechanisms," specify "causal effects," or pursue "thick descriptions" (e.g., Abbott, 2004, pp. 242–248; Geertz, 1973; Hedström &

Swedberg, 1998; Winship & Morgan, 1999). While the underlying ideas are often productive, their rhetorical packaging may limit the explanations that we consider for answering specific research questions. In a telling example, Harrison White (2001) recalled how his stubborn focus on networks kept him from recognizing the institutional processes that explained his case. More broadly, sociologists repeatedly invoked the natural science idea of “laws” at least until the 1960s (Abbott, 1998, pp. 162–163), used culture in as different meanings as categories and hermeneutics (Mohr & Rawlings, 2012), and let contagion and prestige spread statistical significance ideas (Leahey, 2005). The rhetoric one scholar considers crucial for making sense of a social problem the next considers misleading, each potentially forgoing important explanations.

The data science examples showed familiar logics without the familiar rhetorics. Quantitative expertise, old and new, certainly has its own skirmishes about logic and approaches, such as around modeling techniques (e.g., Breiman, 2001) or the utility of data (Mützel et al., 2018). But the instances above largely avoided them. We saw how Brin and Page argued about the mechanisms of information flow, Airbnb recognized the causal effect of its ranking algorithm, and the LinkedIn data scientist relied on a formalist view of social relations. They made those points without using the labels. Data science recalls moments in sociological research when rhetorical conventions take a backseat behind new observations and draw attention to the analytical intuitions that respond to the problems we encounter.

## 2.4 Designs

Different puzzles, debates, and mechanisms favor or require different designs for empirical studies. We have some standard designs, such as surveys, interviews, or participant observations (e.g., Black, 1999; Gray et al., 2007; Pajo, 2017). The rise of digital data and tools challenged our standard approaches initially, but scholars have quickly proposed new rules and procedures (e.g., Bacak & Kennedy, 2019; Marres, 2017; McCormick et al., 2017). Yet, we also know that the research process is less definite than those rules and standards make it look (e.g., Lazer et al., 2021; Martin, 2017). Formal descriptions of that process are essential for intellectual progress, but they necessarily miss details from concrete research situations (Latour, 1987).

Data science uses designs as well. A/B testing is popular among all user-interfacing online services such as the three cases above (Schutt & O’Neil, 2013). But new recommendation principles, like triadic closure, do not always result from this framework. And, as at least partly an engineering project, Brin and Page’s Google prototype involved designing a data infrastructure around the web’s hyperlinks to begin with. Airbnb had to accommodate the complexities of a two-sided market into a data-analytic infrastructure before they could tweak their algorithm to recommend one or the other destination. All these examples involved new dataset designs for making relevant observations. They used analytical principles but without closely observing a set of rules that others derived from their problems and settings. Data science recalls instances when sociologists encounter new research settings that require reflection on a design’s purpose rather than its status in the field.

## 2.5 Work products

With the debate, theory, puzzle, and design in place, we write proposals for implementing them and articles that report the results. But grant applications and publications in top-ranking journals, our most valued output, are subject to social factors outside of our control or effort. Negligible differences in early funding evaluations discourage scholars from obtaining subsequent



funds later in their careers (Bol et al., 2018), different work practices lead to different publication patterns across men and women (Squazzoni et al., 2021), and reputation protects against rejections (Bravo et al., 2018). Current conventions for academic output communicate major findings but miss and skew contributions and accomplishments.

The data science examples used their analyses to create work products other than those with institutional recognition in the discipline. Like sociologists, Brin and Page published their ideas behind Google in an academic paper, and many networks researchers today use the 'PageRank' algorithm. But their main contribution was of course a functioning search engine for the web. Although Airbnb did not start with a research paper, like other big tech companies, it now participates in the open-source community and publishes on questions in search and rankings (Grbovic, 2017), which is easy to imagine, but also trust and algorithmic anxiety (Barbosa et al., 2020; Jhaver et al., 2018). For LinkedIn and Airbnb, the main forms of output were the algorithms that reproduced familiar social experiences in the chaos of the digital transformation. These data science examples point at directions for sociology to consider work products other than publications for sharing ideas and insights (Nelson, 2021).

Sociological practice often differs from its presentation, a phenomenon we know well. If we only listen to ourselves, we risk falling victim to an "attitudinal fallacy" wherein what "people say is often a poor predictor of what they do" (Jerolemack & Khan, 2014, p. 1). The discipline's practice of critical self-reflection provides some protection against that risk (e.g., House, 2019; Weber, 2004). Data science offers opportunities to look at sociology from a different perspective (Krause, 2021). These perspectives give us ideas about practice, which becomes more visible amid limited institutional scaffolding. But isolated instances cannot support a successful scientific project.

### 3 Problem Situations as Solutions

The data science examples showed traces of the ongoing digital transformation and revealed sociological intuitions. I propose turning to John Dewey for interpreting those analogies. Dewey, aside from a recent return to sociological thinking (Waight, 2021; Whitford, 2022), is mainly seen as a philosopher and perhaps a psychologist. But he also wrote about scientific practice, and his ideas about education and learning are useful for sociology to take advantage of the ongoing transformation.

Dewey had a name for the situations in which early sociologists and today's data scientists found themselves. He called them "problem situations," by which he meant situations that are "disturbed, troubled, ambiguous, confused, full of conflicting tendencies, obscure," and so on (Dewey, 1938, p. 105). These situations often result from larger changes, including those of the industrial transformation, when sociology formed, and of the recent digital transformation. In pragmatist thinking, actors influence the local implications of these larger changes, provided they manage to work together across different views because problem situations induce creative conduct (Dewey, 1938, p. 107; Joas, 1992, p. 10; Stark, 2009). These are not purely mental exercises; problem situations require attempted solutions. During those attempts, initial "ends-in-view" turn into "means" toward new ends that have moved into view (Dewey, 1929, p. 119). Actors choose means and ends-in-view in relation to one another during "a tentative trying-out of various courses of action" (Dewey, 1922, p. 202). This micro-level perspective conceptualizes the 'experiences' and 'inspiration' that Weber (2004) saw in scientific work and the accomplishments of the modern data science instances.

The data science activities at LinkedIn, Google, and Airbnb all made sociological observations in ambiguous situations outside the discipline. The physicist had no idea of professional networks and relationships, but adjustments to his technical expertise revealed meaningful features. Brin and Page had a substantial stock of knowledge available to them, but web search was still a new problem in the world and academically. Airbnb turned its data analysis upside down, from a focus on larger numbers to one on outliers, though a specific kind of them, systematic ones. Importantly, it shifted focus amid a backlash against its interference with global travel patterns. These are all clear problem situations. The available material provides no conclusive evidence of Deweyan practices. But more detailed descriptions of early data science work are consistent with such an interpretation (Hammerbacher, 2009), and the observable patterns fit as they used empirical observations without returning all the way to established classics, rhetorics, or other institutional scaffolding.

How can pragmatist theory ensure continuous knowledge production? Debates and paradigms, our primary means of intellectual orientation, are indispensable for continuous knowledge production. And Dewey endorsed taking up ideas of teachers and older ideas in his writing about teaching and learning. But he also stressed that “The basic control [referring to teaching] resides in the nature of the situations in which the young take part” (Dewey, 1916, p. 51). The focus on situations produces systematic results because, in Dewey’s view, members of a group “tend to act with the same controlling ideas, beliefs, and intentions, given similar circumstances” (Dewey, 1916, p. 45). This reasoning would account for Brin and Page and the anonymous data scientists at LinkedIn and Airbnb all proposing successful solutions to data problems that echoed ideas to which they had no extended exposure. In contrast to data science, sociology aims to produce a systematic stock of knowledge. But the patterns from data science suggest that systematic ideas can emerge even when we temporarily step away from the trusted symbols and references that typically align our thinking to expose ourselves to unfamiliar social settings. That is, as long as we continue to talk and listen to each other (Stark, 2009).

Dewey’s theory can appear a bit clumsy, and interpretations have varied (Whitford, 2022). But it has a simple principle, reflexivity, which has already shown its utility for new observations in sociology. Reflexivity is most familiar as the reflection on developments in the discipline and as a methodological concern in qualitative research. But a few recent quantitative projects have pursued a new direction. They started using modern infrastructures for sharing datasets and analytic procedures, the features that have also benefited data scientists, to test whether old results hold up to new scrutiny.

One variant of this exercise involves large numbers of researchers who address the same issue in independent teams, each of which ultimately generates new observations. A telling episode started with a project wherein 29 teams studied the same research question about racial discrimination in sports using the same dataset of penalties during soccer matches (Silberzahn et al., 2018). Despite the common starting point, the teams of researchers produced vastly different results, raising questions about the ability of any single social scientific study to generate robust insights. But a pair of unrelated scholars were skeptical and conducted a reanalysis of the initial study. Their reanalysis showed that more consistent results were possible with a clearer research question (Auspurg & Brüderl, 2021). These iterative reflections refined our thinking about discrimination and about the research process.

The reflections don’t have to stop there. A parallel project produced an even more radical conclusion. This project involved 160 teams that tried to predict life outcomes such as the grade point average of a child or material hardship in a household (Salganik et al., 2020). The

authors found no worrying differences across the results, suggesting that the research question was sufficiently clear. But they noted that the ideas of 160 teams failed to produce meaningfully better predictions of the life outcomes than a simple benchmark analysis. To understand this shortfall, the lead author joined interviewees in the field and learned that the survey missed relevant questions to capture the lived social experiences.<sup>9</sup>

In Dewey's terms, the resolution of one problem situation brought another one into view. This process does not stop with the confines of modeling strategies or datasets, and there is no definite sequence of steps. Reflexivity requires professional judgment that builds on continuous sociological practice.

#### 4 Sociology's Stake in Data Science

When sociologists took up the challenges of the digital transformation, they noticed irritating overlaps with a new group, data scientists, and drew practical conclusions for their research.<sup>10</sup> But data science is more than a set of techniques and technologies. Besides leaving their mark on modern social life, data scientists work on problems sociologists have studied long before. And their ideas reflect sociological intuitions, however rudimentary, even though data science lacks sociology's disciplinary foundation. While others rushed to protect their jurisdiction against data scientists, I have argued that sociology has a stake in data science that we can leverage for adjusting to the digital era. Weber's focus on the "inner vocation" for science highlighted data-analytic experiences as the building blocks of research. Dewey's pragmatism offered a framework for conceptualizing those experiences, which nevertheless remain impossible to formalize. Data science offers no solution for that problem, but the overlaps provided a point of reflection on making sociological observations that advance the discipline.

Recent programmatic articles have argued that quantitative sociology needs to break out of the deductivist agenda from previous decades and adopt more exploratory perspectives (Evans & Foster, 2019; McFarland et al., 2016). Many have already proposed new directions for sociology that integrate formal and qualitative perspectives (e.g., Nelson, 2020; Wagner-Pacifici et al., 2015). Dewey's pragmatist colleague Charles S. Peirce supplied a promising epistemological framework for this transition, abduction, which uses existing theory to guide surprising discoveries (Brandt & Timmermans, 2021; Goldberg, 2015). Outlets like *Socius*, *Sociologica*, and *Sociological Science* endorse article formats that deviate from conventions in ways that accommodate these strategies. But, to take one more lesson from data science's emergence (Hammerbacher, 2009), these new frameworks and outlets only work if we pursue this agenda as a community.

Such a collective project is not tied to specialized skills or interests. We can find guidance across technical and substantive specializations in a pragmatist view of learning, reasoning, and conduct that we more typically observe among our subjects. This focus on practice, what problems we choose and how we study them, allows us to reconsider disciplinary conventions as the discipline adjusts to the digital transformation. The computational social science movement has cultivated the technical skills for keeping up. But early successes have come with warning signs. The new skills and modern data collection and storage approaches are so resource-

9. Salganik shared the insights from the interviews in a conference presentation in 2019.

10. When data science made a name for itself as "the sexiest job of the twenty-first century" (Davenport & Patil, 2012), friends and colleagues asked me whether they, thanks their quantitative expertise, could call themselves data scientists. Some wanted to tease me for my curious research topic (Brandt, 2016); others were genuinely interested. Both concerns say something about data science's effect on the discipline.

intensive that they require large labs or research groups, which quickly undermine the problem situations that data science has illustrated.

This final paragraph would be a good place for summarizing new standards that guide research in the digital era. But positioning sociology in the digital age is an ongoing process that must involve the whole discipline and motivate appreciation of unfamiliar situations that lead to new observations. Those in Ph.D. programs, with endowed chairs, or still undergraduates, in sociology or other disciplines, and with or without data science-like skills, who are curious about new problems and directions must find each other at conferences, in review processes, committees, and other situations, jointly reflecting on research practice against the backdrop of social scientific problems at hand. If they do, the discipline can not only expand into new problems and perspectives. It can also speak to those who did not have the privilege of comprehensive sociological training but feel drawn to making sociological observations. There is too much to do to let jurisdictional skirmishes undermine work on substantive problems.

## References

- Abbott, A. (1988). *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, IL: University of Chicago Press.
- Abbott, A. (1998). The Causal Devolution. *Sociological Methods & Research*, 27(2), 148–181. <https://doi.org/10.1177%2F0049124198027002002>
- Abbott, A. (2001). *Chaos of Disciplines*. Chicago, IL: University of Chicago Press.
- Abbott, A. (2004). *Methods of Discovery Heuristics for the Social Sciences*. Manhattan, NY: WW Norton & Company.
- Auspurg, K., & Brüderl, J. (2021). Has the Credibility of the Social Sciences been Credibly Destroyed? Reanalyzing the “Many Analysts, One Data Set” Project. *Socius*, 7, 1–14. <https://doi.org/10.1177%2F23780231211024421>
- Avnoon, N. (2021). Data Scientists’ Identity Work: Omnivorous Symbolic Boundaries in Skills Acquisition. *Work, Employment and Society*, 35(2), 332–349. <https://doi.org/10.1177%2F0950017020977306>
- Bail, C. (2021). *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. Princeton, NJ: Princeton University Press.
- Bačák, V., & Kennedy, E.H. (2019). Principled machine learning using the super learner: An application to predicting prison violence. *Sociological Methods & Research*, 48(3), 698–721. <https://doi.org/10.1177%2F0049124117747301>
- Barabási, A-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509–512. <https://doi.org/10.1126/science.286.5439.509>
- Barbosa, N.M., Sun, E., Antin, J., & Parigi, P. (2020, April). Designing for Trust: A Behavioral Framework for Sharing Economy Platforms. In Y. Huang, I. King, T-Y Liu, & M van Steen (Eds.), *Proceedings of The Web Conference 2020* (pp. 2133–2143). New York, NY: ACM. <https://doi.org/10.1145/3366423.3380279>

- Battle-Baptiste, W., & Rusert, B. (Eds.). (2018). *WEB Du Bois's Data Portraits: Visualizing Black America*. San Francisco, CA: Chronicle Books.
- Ben-David, J. (1971). *The Scientist's Role in Society: A Comparative Study*. Hoboken, NJ: Prentice-Hall.
- Black, T. R. (1999). *Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics*. London: Sage.
- Bol, T., de Vaan, M., & van de Rijdt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences of the United States of America*, 115(19), 4887–4890. <https://doi.org/10.1073/pnas.1719557115>
- Bonacich, P. (1972). Factoring and Weighting Approaches to Status Scores and Clique Identification. *Journal of Mathematical Sociology*, 2(1), 113–120. <https://doi.org/10.1080/0022250X.1972.9989806>
- Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92(5), 1170–1182. <http://dx.doi.org/10.1086/228631>
- Börner, K., Scriver, O., Gallant, M., Ma, S., Liu, X., Chewning, K., Wu, L., & Evans, J. A. (2018). Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proceedings of the National Academy of Sciences of the United States of America*, 115(50), 12630–12637. <https://doi.org/10.1073/pnas.1804247115>
- Brandt, P. (2016). *The Emergence of the Data Science Profession* [Doctoral dissertation, Columbia University]. <https://doi.org/10.7916/D8BK1CKJ>
- Brandt, P., & Timmermans, S. (2021). Abductive Logic of Inquiry for Quantitative Research in the Digital Age. *Sociological Science*, 8, 191–210. <http://dx.doi.org/10.15195/v8.a10>
- Bravo, G., Farjam, M., Moreno, F.G., Birukou, A., & Squazzoni, F. (2018). Hidden Connections: Network Effects on Editorial Decisions in Four Computer Science Journals. *Journal of Informetrics*, 12(1), 101–112. <https://doi.org/10.1016/j.joi.2017.12.002>
- Breiman, L. (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>
- Brin, S., & Page, L. (1998). The Anatomy of a Large-scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Burt, R.S. (1987). Social Contagion and Innovation: Cohesion versus Structural Equivalence. *American Journal of Sociology*, 92(6), 1287–1335. <https://doi.org/10.1086/228667>
- Castells, M. (1996). *The Rise of the Network Society: The Information Age: Economy, Society, and Culture* (Vol. 1). Hoboken, NJ: Wiley.
- Christin, A. (2020). *Metrics at Work: Journalism and the Contested Meaning of Algorithms*. Princeton, NJ: Princeton University Press.
- Davenport, T.H., & Patil, D.J. (2012). Data Scientist: The Sexiest Job Of the 21st Century. *Harvard Business Review*, 90(10), 70–76.

- Dewey, J. (1916). *Democracy and Education: An Introduction to the Philosophy of Education*. New York, NY: Macmillan
- Dewey, J. (1922). *Human Nature and Conduct*. New York, NY: The Modern Library.
- Dewey, J. (1929). *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. New York, NY: Minton, Balch.
- Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York, NY: Henry Holt and Co.
- Dewey, J. (1939). *Theory of Valuation*. Chicago, IL: University of Chicago Press.
- Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(7), 745–766. <https://doi.org/10.1080/10618600.2017.1384734>
- Dorschel, R., & Brandt, P. (2021). Professionalization via Ambiguity. The Discursive Construction of Data Scientists in Higher Education and the Labor Market. *Zeitschrift für Soziologie*, 50(3-4), 193–210. <https://doi.org/10.1515/zfs0z-2021-0014>
- Durkheim, É. (1893). *De la division du travail social. Étude sur l'organisation des sociétés supérieures*. Paris: Félix Alcan.
- Durkheim, É. (1897). *Le suicide. Étude de sociologie*. Paris: Félix Alcan.
- Eubanks, V. (2018). *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.
- Evans, J., & Foster, J.G. (2019). Computation and the Sociological Imagination. *Contexts*, 18(4), 10–15. <https://doi.org/10.1177/1536504219883850>
- Geertz, C. (1973). Thick Description: Toward an Interpretive Theory of Culture. In *The Interpretation of Cultures* (pp. 310–323). New York, NY: Basic Books.
- Goldberg, A. (2015). In Defense of Forensic Social Science. *Big Data & Society*, 2(2), pp. 1–3. <https://doi.org/10.1177/2053951715601145>
- González-Bailón, S. (2017). *Decoding the Social World: Data Science and the Unintended Consequences of Communication*. Cambridge, MA: MIT Press.
- Gouldner, A.W. (1970). *The Coming Crisis of Western Sociology*. Portsmouth, NH: Heinemann.
- Gray, P.S., Williamson, J.B., Karp, D.A., & Dalphin, J.R. (2007). *The research imagination: An introduction to qualitative and quantitative methods*. Cambridge: Cambridge University Press.
- Grbovic, M. (2017). Search ranking and personalization at Airbnb. *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 339–340. <https://doi.org/10.1145/3109859.3109920>
- Hammerbacher, J. (2009). Information platforms and the rise of the data scientist. In T. Segaran & J. Hammerbacher (Eds.), *Beautiful Data: The Stories Behind Elegant Data Solutions* (pp. 73–84). Sebastopol, CA: O'Reilly Media.

- Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton, NJ: Princeton University Press.
- Hedström, P., & Swedberg, R. (1998). Social mechanisms: An introductory essay. In P. Hedström & R. Swedberg (Eds.) *Social Mechanisms: An Analytical Approach to Social Theory* (pp. 1–31). Cambridge: Cambridge University Press.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. London: Psychology Press.
- House, J.S. (2019). The Culminating crisis of American Sociology and its Role in Social Science and Public Policy: An Autobiographical, Multimethod, Reflexive Perspective. *Annual Review of Sociology*, 45, 1–26. <https://doi.org/10.1146/annurev-soc-073117-041052>
- Ibarra, H. (1999). Provisional selves: Experimenting with Image and Identity in Professional Adaptation. *Administrative Science Quarterly*, 44(4), 764–791. <https://doi.org/10.2307/2667055>
- Jerolmack, C., & Khan, S. (2014). Talk is Cheap: Ethnography and the Attitudinal Fallacy. *Sociological Methods & Research*, 43(2), 178–209. <https://doi.org/10.1177/0049124114523396>
- Jhaver, S., Karpfen, Y., & Antin, J. (2018). Algorithmic Anxiety and Coping Strategies of Airbnb Hosts. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3173574.3173995>
- Joas, H. (1992). *Die Kreativität des Handelns*. Frankfurt am Main: Suhrkamp.
- Katz, E., & Katz, R. (2016). Revisiting the Origin of the Administrative versus Critical Research Debate. *Journal of Information Policy*, 6(1), 4–12. <https://doi.org/10.5325/jinfopoli.6.2016.0004>
- Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632. <https://doi.org/10.1145/324133.324140>
- Krause, M. (2021). On Sociological Reflexivity. *Sociological Theory*, 39(1), 3–18. <https://doi.org/10.1177/0735275121995213>
- Krippner, G., Granovetter, M., Block, F., Biggart, N., Beamish, T., Hsing, Y., Hart, G., Arrighi, G., Mendell, M., Hall, J., Burawoy, M., Vogel, S., & O’Riain, S. (2004). Polanyi symposium: a conversation on embeddedness. *Socio-Economic Review*, 2(1), 109–135. <https://doi.org/10.1093/soceco/2.1.109>
- Kuhn, T.S. (1970). *The Structure of Scientific Revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.
- Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.
- Lazarsfeld, P.F., & Oberschall, A.R. (1965). Max Weber and Empirical Social Research. *American Sociological Review*, 30(2), 185–199. <https://doi.org/10.2307/2091563>
- Lazer, D., Hargittai, E., Freelon, D., González-Bailón, S., Munger, K., Ognyanova, K., & Radford, J. (2021). Meaningful Measures of Human Society in the Twenty-first Century. *Nature*, 595(7866), 189–196. <https://doi.org/10.1038/s41586-021-03660-7>

- Leahey, E. (2005). Alphas and Asterisks: The Development of Statistical Significance Testing Standards in Sociology. *Social Forces*, 84(1), 1–24. <https://doi.org/10.1353/sof.2005.0108>
- Leskovec, J., & Krevl, A. (2014). SNAP Datasets. Stanford University. <https://snap.stanford.edu/data/>
- Lohr, S. (2015). *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*. New York, NY: Harper Collins.
- Marres, N. (2017). *Digital sociology: The Reinvention of Social Research*. Hoboken, NJ: Wiley.
- Martin, J.L. (2017). *Thinking Through Methods: A Social Science Primer*. Chicago, IL: University of Chicago Press.
- McCormick, T.H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E.S. (2017). Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing. *Sociological Methods & Research*, 46(3), 390–421. <https://doi.org/10.1177/0049124115605339>
- McFarland, D.A., Lewis, K., & Goldberg, A. (2016). Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist*, 47(1), 12–35. <https://doi.org/10.1007/s12108-015-9291-8>
- McMahan, P., & McFarland, D.A. (2021). Creative Destruction: The Structural Consequences of Scientific Curation. *American Sociological Review*, 86(2), 341–376. <https://doi.org/10.1177/0003122421996323>
- Mead, R. (2019). The Airbnb Invasion of Barcelona. *The New Yorker*, 22 April. <https://www.newyorker.com/magazine/2019/04/29/the-airbnb-invasion-of-barcelona>
- Merton, R.K. (1968). The Matthew Effect in Science. *Science*, 159(3810), 56–63. <https://doi.org/10.1126/science.159.3810.56>
- Mohr, J.W., & Rawlings, C. (2012). Four Ways to Measure Culture: Social Science, Hermeneutics, and the Cultural Turn. In J.C. Alexanders, R.N. Jacobs, & P. Smith (Eds.), *The Oxford Handbook of Cultural Sociology* (pp. 70–113). Oxford: Oxford University Press.
- Mützel, S., & Kressin, L. (2020). From Simmel to Relational Sociology. In S. Abrutyn & O. Lizardo (Eds.), *The Handbook of Classical Sociological Theory* (pp. 217–238). New York, NJ: Springer.
- Mützel, S., Saner, P., & Unternährer, M. (2018). Schöne Daten! Konstruktion und Verarbeitung von digitalen Daten. In D. Houben, & B. Prietl (Eds.), *Datengesellschaft* (pp. 111–132). Berlin: Verlag.
- Nelson, L.K. (2021, August). Early Career Faculty Spotlight. *ASA Methodology Section Newsletter*, 8.
- Nelson, L.K. (2020). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*, 49(1), 3–42. <https://doi.org/10.1177/0049124117729703>
- Nelson, L.K. (2022). Laura Nelson. GitHub. <https://github.com/lknelson>
- Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: New York University Press.



- NORC. (2021) Get the Data. NORC at the University of Chicago. <https://gss.norc.org/get-the-data>
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Crown Books.
- Ollion, E., & Abbott, A. (2016). French Connections: The Reception of French Sociologists in the USA (1970-2012). *Archives européennes de sociologie*, 57(2), 331–372. <https://doi.org/10.1017/S0003975616000126>
- Pajo, B. (2017). *Introduction to Research Methods: A Hands-on Approach*. London: Sage.
- R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ribes, D. (2019). STS, Meet Data Science, Once Again. *Science, Technology, & Human Values*, 44(3), 514–539. <https://doi.org/10.1177%2F0162243918798899>
- Romero, M. (2020). Sociology Engaged in Social Justice. *American Sociological Review*, 85(1), 1–30. <https://doi.org/10.1177%2F0003122419893677>
- Salganik, M.J. (2018). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.
- 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., Filip-pova, A., Gilroy, C., Goode, B.J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., Morgan, A.C., ..., & McLanahan, S. (2020). Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 117(15), 8398–8403. <https://doi.org/10.1073/pnas.1915006117>
- Schutt, R., & O'Neil, C. (2013). *Doing Data Science*. Sebastopol, CA: O'Reilly Media, Inc.
- Shwed, U., & Bearman, P.S. (2010). The Temporal Structure of Scientific Consensus Formation. *American Sociological Review*, 75(6), 817–840. <https://doi.org/10.1177/0003122410388488>
- SIENA. (2022). Data sets for use with Siena. Oxford University. [https://www.stats.ox.ac.uk/~snijders/siena/siena\\_datasets.htm](https://www.stats.ox.ac.uk/~snijders/siena/siena_datasets.htm)
- 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., Fong, N., ..., & 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>
- Simmel, G. (1908). *Soziologie. Untersuchungen über die Formen der Vergesellschaftung*. Berlin: Duncker & Humblot.
- Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., Willis, M., Birukou, A., Don-dio, P., & Grimaldo, F. (2021). Peer Review and Gender Bias: A Study on 145 Scholarly Journals. *Science Advances*, 7(2). <https://doi.org/10.1126/sciadv.abdo299>

- Stark, D. (2009). *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton, NJ: Princeton University Press.
- Stinchcombe, A.L. (1982). Should Sociologists Forget Their Mothers and Fathers? *American Sociologist*, 17(1), 2–11. <https://www.jstor.org/stable/27702490>
- Turco, C.J., & Zuckerman, E.W. (2017). Verstehen for Sociology: Comment on Watts. *American Journal of Sociology*, 122(4), 1272–1291. <https://doi.org/10.1086/690762>
- Twitter, Inc. (2022). *Twitter API Academic Research Access*. Twitter. <https://developer.twitter.com/en/products/twitter-api/academic-research>
- Vedres, B. (2022). Balazs Vedres. CEU. [http://www.personal.ceu.hu/staff/Balazs\\_Vedres/](http://www.personal.ceu.hu/staff/Balazs_Vedres/)
- Vedres, B., & Stark, D. (2010). Structural Folds: Generative Disruption in Overlapping Groups. *American Journal of Sociology*, 115(4), 1150–1190. <https://doi.org/10.1086/649497>
- Wagner-Pacifici, R., Mohr, J.W., & Breiger, R.L. (2015). Ontologies, Methodologies, and New uses of Big Data in the Social and Cultural Sciences. *Big Data & Society*, 2(2), 1–11. <https://doi.org/10.1177/2053951715613810>
- Waight, H. (2021). Recovering John Dewey's Lost Vision for Social Science in Contemporary American Sociology. *The American Sociologist*, 52, 420–448. <https://doi.org/10.1007/s12108-021-09482-4>
- Watts, D.J. (2014). Common Sense and Sociological Explanations. *American Journal of Sociology*, 120(2), 313–351. <https://doi.org/10.1086/678271>
- Weber, M. (2004). Science as a Vocation. In D.S. Owen, T.B. Strong, & R. Livingstone (Eds.), *The Vocation Lectures* (R. Livingstone, Trans.) (pp. 1–31). Indianapolis, IN: Hackett. (Original work published 1919)
- Wellman, B. (1997). An Electronic Group is Virtually a Social Network. In S. Jiesler (Ed.), *Culture of the Internet* (pp. 179–205). Mahwah, NJ: Lawrence Erlbaum.
- White, H.C. (2001). Interview with Harrison White: 4-16-01 by Alair MacLean and Andy Olds. Theory@Madison. <https://www.ssc.wisc.edu/theoryatmadison/papers/ivwWhite.pdf>
- White, H.C., Boorman, S.A., & Breiger, R.L. (1976). Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions. *American Journal of Sociology*, 81(4), 730–780. <https://www.jstor.org/stable/2777596>
- Whitford, J. (2022). Disambiguating Dewey; or Why Pragmatist Action Theory Neither Needs Nor Asks Paradigmatic Privilege. In N. Gross, I.A. Reed, & C. Winship (Eds.), *The New Pragmatist Sociology: Inquiry, Agency, and Democracy*. New York, NY: Columbia University Press.
- Winship, C., & Morgan, S.L. (1999). The Estimation of Causal Effects from Observational Data. *Annual Review of Sociology*, 25(1), 659–706. <https://doi.org/10.1146/annurev.soc.25.1.659>

**Philipp Brandt** – Department of Sociology and Center for the Sociology of Organisations, Sciences Po, Paris (France)

✉ philipp.brandt@sciencespo.fr; 🔗 <https://www.sciencespo.fr/cso/en/researcher/Philipp%20Brandt/8424.html>

Philipp Brandt is an Assistant Professor of sociology at Sciences Po (France) and a researcher at the Centre de Sociologie des Organisations (CSO). His research analyzes the emergence of the data science profession, immigrant career trajectories, and economic policy implementation using computational techniques, interviews, and field observations. He also writes about methodology.