



HAL
open science

Abductive Logic of Inquiry for Quantitative Research in the Digital Age

Philipp Brandt, Stefan Timmermans

► **To cite this version:**

Philipp Brandt, Stefan Timmermans. Abductive Logic of Inquiry for Quantitative Research in the Digital Age. Sociological Science, 2021, 8, pp.191 - 210. 10.15195/v8.a10 . halshs-03266002

HAL Id: halshs-03266002

<https://shs.hal.science/halshs-03266002>

Submitted on 21 Jun 2021

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.



Abductive Logic of Inquiry for Quantitative Research in the Digital Age

Philipp Brandt,^{a, b} Stefan Timmermans^c

a) Sciences Po; b) Center for the Sociology of Organizations, c) University of California, Los Angeles

Abstract: We propose an abductive logic of scientific inference for quantitative research. The advent of computational sociology has exposed the limitations of a deductive logic of inquiry for quantitative researchers due to a lack of traditional sociological variables and an abundance of unfamiliar variables and data formats, complicating hypothesis testing. In response, some researchers have embraced inductive inference, but inductive analysis without theoretical guidance risks producing atheoretical findings. An abductive logic of inquiry rests on developing new theoretical insights based on surprising research results in light of existing theories. In computational sociology, such surprising findings can be cultivated by taking advantage of the analytical potential of scaled-up data and developing flexible analytical and visualization procedures. We illustrate these tactics with a surprising finding in a study of the labor supply decisions of New York City yellow cab drivers.

Keywords: computational social science; abduction; large-scale data; scientific inference; induction; deduction

QUANTITATIVE research in the social sciences generally aspires to work deductively. Most research articles develop a research question based on a limitation or causal explanation in one or more existing theories, derives theoretically informed hypotheses, and tests the predictions with respect to a limited number of observational variables linked together by measures of association. This formal mode of scientific inference has worked well for surveys, experimental research, and administrative data sets (e.g., census) that come with comprehensive documentation and limit, by design, the kinds of research questions to ask. Even if the actual research process is not as clean as presented (e.g., the controversy about p-hacking [Simmons, Nelson, and Simonsohn 2011]), most quantitative textbooks favor a deductive model when discussing “the scientific method” (e.g., Black 1999; Gray et al. 2007; Pajo 2017).

As the data sources and analytical techniques available to quantitative researchers change, the need to revisit issues around exploration, explanation, and prediction has become clear (Hofman, Sharma, and Watts 2017; Salganik 2018; Shmueli 2010). Digital technologies provide access to unfamiliar data sets that include behavioral traces (e.g., point of sales, geolocation data, social media scrapings, CCTV recordings), machine-readable texts, and code and data repositories (Lazer et al. 2009; Lazer and Radford 2017). These “found” data produced without research goals in mind don’t look like the familiar surveys: they not only require new technical skills and computing capacities to manage their scale and format (e.g., Bail 2017; Heiberger and Riebling 2016) and require critical evaluation of their content (Salganik 2018:16), but they also demand a different logic of scientific inquiry (Evans and Foster 2019; McFarland, Lewis, and Goldberg 2016; Wagner-Pacifici,

Citation: Brandt, Philipp, and Stefan Timmermans. 2021. “Abductive Logic of Inquiry for Quantitative Research in the Digital Age.” *Sociological Science* 8: 191-210.

Received: March 20, 2021

Accepted: April 20, 2021

Published: June 17, 2021

Editor(s): Jesper Sørensen, Mario Small

DOI: 10.15195/v8.a10

Copyright: © 2021 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited.

Mohr, and Breiger 2015). Such data sets do not lend themselves easily to a deductive hypothesis testing approach because some “obvious” hypotheses based on existing theories are not testable due to missing variables, whereas many others could be tested due to an abundance of new variables. Often, the findings are also surprising, not accounted for by existing theories, and require the formulation of new theories.

In this article, we advance philosopher Charles Peirce’s notion of abduction as an alternative logic of scientific inference suitable to quantitative computational inquiry (Peirce 1934). Peirce saw abduction as the engine of theory *discovery*: compared with deduction and induction, it is both the most creative and the most tenuous of scientific logics. When researchers notice surprising or anomalous findings in light of existing theories, they need to come up with, as Peirce put it, a hypothesis on probation. Abduction refers to this creative leap, the formulation of a tentative hypothesis that may result in the revision of an existing theory or development of a new theory. Abduction begins with an unexpected, anomalous, or surprising observation in light of existing theories and then systematizes the discovery of inferences that may account for the initial observation. Peirce (1934:5.181) offered the following syllogism to explain abduction:

The surprising fact, C, is observed;
But if A were true, C would be a matter of course,
Hence, there is reason to suspect that A is true.

The key is that the hypothesis consists of a new connection beyond what is observed, rendering the surprising fact a “matter of course.” These hypotheses are then firmed up (or refuted) by deductively and inductively motivated research. Peirce proposed three criteria for picking promising hypothesis: observability, verifiability, and economic cost of pursuing the hypothesis (Fann 1970).

Because abductive inference starts with surprising observations unaccounted for by existing theories and develops new theoretical insights to reconcile these findings, the first prerequisite for abductive inferencing is an in-depth familiarity with a broad variety of pertinent theories. Findings are only surprising and anomalous when they are unaccounted for in the expectations and predictions set by existing theories. Armed with theoretical insights, researchers can also deliberately cultivate abductive inferences through methodological strategies—the topic of this article.¹ In qualitative research, where abductive analysis has emerged as an alternative to inductive grounded theory approaches (Tavory and Timmermans 2014; Timmermans and Tavory 2012), ethnographers and interviewers have developed new data gathering and coding approaches to open up surprising findings. Quantitative researchers can similarly improve their chances for anomalous observations while playing to the strengths of “big” data and deliberately exploring new forms of data analysis and visualization (Edelmann et al. 2020; Healy 2018; McFarland et al. 2016).

Although some quantitative researchers have hinted that abduction could be useful for computational sociology (e.g., Goldberg 2015; Van Maanen, Sørensen, and Mitchell 2007; Wagner-Pacifi et al. 2015), our contribution is to render the abductive logic of theory discovery effective for this mode of research. By taking advantage of novel data sets and designing studies to cultivate surprising findings, abduction makes discovering innovative and creative theoretical insights in quantitative research more likely. In the next section, we show that abductive reasoning is

promising when past researchers have scaled up data and used new data visualization and manipulation techniques. Then, we demonstrate the benefits of abductive reasoning in computational sociology with a study of New York Taxi drivers.

From Induction-Deduction to Abduction: Two Methodological Tactics

When new data and analytical tools of the digital transformation became available, social scientists saw new opportunities for studying long-standing theoretical questions (Evans and Foster 2011; Lazer et al. 2009; Watts 2007). Edelman et al. (2020) summarized seven areas that have been particularly active since then, including social networks, collective behavior and political sociology, and cultural sociology. Theory links computational and mainstream sociology, pointing at new opportunities to test old ideas, viewing the digital as a distinct social setting, and advancing the interpretation of computational patterns.

In order to capitalize on the theoretical potential of computational sociology, researchers proposed inductive as alternative to deductive scientific inquiry (e.g., Evans and Foster 2019; Nelson 2020). Nelson (2020) appropriated Glaser and Strauss's (1967) grounded theory approach for text analysis as a computational implementation of content analysis. She proposed three steps: finding, understanding, and testing patterns. Theory emerged, consistent with its long history in sociology due to the legacy of grounded theory (Charmaz 2014), inductively in Nelson's approach, which contrasted with competing proposals that see a "theoretical hunch" at the beginning of this kind of research (McFarland et al. 2016). McFarland et al. (2016) proposed that computational sociology adopts a "forensic social science" model, by which they meant a "middle ground that is both inductive and theory-oriented" (P. 13). New forms of data, they argued, "do not necessitate a research-driven hypothesis in order to be generated" and instead allow research to "inductively build theory from the ground up" (Pp. 16–17). Practically, the approach becomes a "computational ethnography" that emulates Glaser and Strauss's (1967) grounded theory and unfolds as a "discovery process [that] begins with an initial hunch or intuition" (McFarland et al. 2016:18). Importantly, McFarland et al. (2016) warned that induction may cause its own problems: machine learning "results in little to no understanding if there is no effort to derive a theory of explanation." Therefore, "the iterative combination of atheoretical induction and theory-led deduction can be quite powerful" (P. 21).

Wagner-Pacifici et al. (2015) summarized the familiar tension between deduction and induction in computational sociology. They noted, with Goldberg (2015), that deductive hypothesis testing leads to limiting one's sociological imagination because it centers the analysis on a number of preset variables. Instead, the challenge with computational data is to decide which, out of a thousand possible variables, are the most relevant. Here, inductive reasoning can help to distinguish which variables stick, but it may have limited theoretical payoff if it remains on the level of classifying data into broad themes (Tavory and Timmermans 2019). Abduction, where surprising research findings are surprising exactly because they violate our

prevailing theoretical preconceptions, offers a way out by asking how unusual results make theoretical sense, requiring theoretical innovations. Wagner-Pacifici et al. (2015) therefore endorse abductive reasoning for computational research, but the implementation of an abductive logic of inference remains underdeveloped (e.g., Goldberg 2015; Ignatow 2020; Kitchin 2014).

Building upon this foundation, we outline two tactics for working abductively with quantitative data in the digital age. The first tactic is about leveraging the large scale at which data have become available to locate surprising empirical findings. We can methodologically cultivate abductive insights out of voluminous and complex data because they enable researchers to recover observations previously too minute or isolated to appear systematically in most conventional data formats. The “messiness” of big data constitutes an opportunity for theory construction. The second tactic takes advantage of technical developments around data processing, such as machine learning and visualization. These data processing techniques enable researchers to uncover surprising observations. We illustrate these tactics using examples from existing social scientific research, which retroactively we can diagnose as following an abductive logic of inquiry even if the researchers themselves do not describe their logic in these terms.

Scaling Up Data

The received view is that large-scale data sets challenge conventional research practices because they capture novel settings and observations (Evans and Foster 2019; McFarland et al. 2016; Wagner-Pacifici et al. 2015). Indeed, scaling up data sources often comes with a loss of data granularity and variables common in social science research (including critical demographic variables) but also constitutes a gain of data volume, variability, velocity, and complexity that allows researchers to ask novel research questions. This trade-off is particularly apparent in found data (Salganik 2018:14). Our first tactic is to play to the strengths of big data to cultivate empirical surprises. In fact, detailed, rich, and complex data are a prerequisite for abductive inference (Van Maanen et al. 2007). The history of social science contains several moments in which a shift in the volume and amount of data led to empirical surprises that lent themselves to abductive inferences and generated new theoretical insights, even if the breakthroughs were not described or recognized in abductive terms. We provide two such examples where researchers turned an apparent data weakness into a strength when they discovered surprising findings that generated new theoretical insights.

Network data. Milgram designed experiments to test the small-world theory in the 1960s (Milgram 1967; Travers and Milgram 1969), but Watts and Strogatz (1998) used more comprehensive and extensive “trace” data sets to build on the small-world theory, in the process jumpstarting network science. Milgram’s study investigated whether two individuals were connected by asking a group of randomly chosen participants to forward a letter to a target individual through personal contacts. He found chains of about five intermediary contacts, supporting the idea of a small world. In analyzing how letters found their target, Milgram considered, for example, the significance of private or work contacts and whether they were

more likely to pass through same-sex contacts. Although groundbreaking, Milgram's original study reported results of just 64 chains in which letters reached the target, a fraction of the initial sample (Travers and Milgram 1969).

Thirty years later, Watts "wanted to prove the theory that everyone in the world is connected by just six degrees of separation" (Watts 2014). Watts and Strogatz tested the small-world theory on a complete social network data set of co-starring ties between all 225,226 movie actors in U.S. films. The size of the data set allowed Watts and Strogatz to depart from Milgram's explanatory strategy of focusing on demographic information. Instead, they exclusively developed the overall network structure. Whereas Milgram (1967) saw implications for the "sociological characteristics" of a "communication net" (P. 67), Watts and Strogatz (1998) concluded that the small-world phenomenon "is probably generic for many large, sparse networks found in nature" (P. 441). Indeed, their computational approach showed small-world characteristics in biological and technological networks (Watts and Strogatz 1998).

Thus, rather than mourning the lack of easily accessible demographic data, Watts and Strogatz cultivated the strength of their upscaled data. They started with an existing theory but then found surprising findings about the structural similarities across professional, biological, and technological networks that required new abductively derived hypotheses to explain the surprise (such as the general occurrence of small-world structures).

Survey data. Next, we turn to research on social class to illustrate the potential of abductive inference for large-scale survey data analysis. In their study of the reproduction of intergenerational class inequalities in the United Kingdom, Laurison and Friedman (2016) responded to literature that defined social mobility as a problem of "*occupational access*" (P. 669; italics in original) and ignored the differences in resources people bring to occupations and the rewards they gain once employed in an occupation. They discovered a "class ceiling" effect in social mobility: people from lower social classes gaining access to higher-status occupations do not reap equivalent benefits (in terms of earnings) as people entering these same occupations from an upper social class background.

This important finding was suggested by Laurison and Friedman's earlier work that relied on a large but imperfect data set. Previously, the limited size of data sets had been a barrier to advancing this field of research: Weeden and Grusky (2005:156), for instance, noted that the size of their data curtailed their ability to show the benefits of a micro-class view. To overcome this barrier, Friedman, Laurison, and Miles (2015) used the Great British Class Survey (GBCS). Savage (2020) recalls that because this data set "had such a large sample size, we were able to descriptively compare pay rates at a very fine-grained level." The GBCS's large size was a result of an online data collection strategy that imposed no restrictions on who could answer the questionnaire. Although these data first revealed variation within occupations by class background, these findings were not yet representative (Friedman et al. 2015). The sense of a lack of representativeness could have rendered the data set unsuitable for a sociological analysis. However, the 2015 data set produced the surprising finding of the importance of class background at occupational entry that Laurison and Friedman were able to develop subsequently with the U.K. Labour

Force Survey, “the first large-scale and representative study of social mobility into and within Britain’s higher professional and managerial occupations” (Laurison and Friedman 2016:669). With findings at odds with the “occupational access” view in the literature, Laurison and Friedman (2016) advocated for “an approach rooted in Bourdieusian theory” that demands a comprehensive understanding of an actor’s resources, including earnings and education, as well as other forms of economic, cultural, and social capital (P. 691).

As in the case of Watts and Strogatz, the GBCS came with distinct disadvantages that could have limited the data set’s utility for social science research. The larger data size, however, allowed a surprising finding to emerge, which prompted an abductive inference in theorizing and streamlined the analysis in a more representative data set. The take-home message is that although larger data sets often have more distinct data limitations in what variables are included or how the data were collected, when researchers capitalize on the opportunities of upscaling the data and their heterogeneity, they can cultivate surprising findings in light of existing research based on smaller data sets. McFarland et al. (2016:13) consider large-scale data a potential “trading zone” where researchers from entirely different paradigms, despite differences in language and culture, collaborate with each other to exchange tools, information, and knowledge.”

New Procedures

In order to take advantage of the large repositories of social traces, networks, and texts proffered by computational data, social scientists need to acquire new technical skills (Evans and Foster 2019; Lazer et al. 2009; Mützel 2015). Some methods, such as unsupervised machine learning approaches in text analysis (Molina and Garip 2019), were designed for purposes other than hypothesis testing. The programming packages that facilitate this work have also opened opportunities for data visualizations that were not systematically available previously and therefore not part of our methodological thinking (Healy 2018).

New data manipulation. Radicalism entails opposition against powerful sociopolitical systems and institutions and takes the form of championing vanishing traditions, justifying violence against innocents, or pursuing nonviolent extra-institutional action. Studying more than two decades of radical discourse in Afghanistan, Karell and Freedman (2019) find, consistent with the literature, a rhetoric of subversion that evokes an historical fight against powerful external enemies. They, however, also discover a second body of discursive radicalism that does not fit the literature. The rhetoric of reversion draws upon the notions of everyday morality to uproot undesirable developments in the personal and local contexts.

This surprising finding relies not only on a comprehensive corpus of materials produced by Afghani mujahideen organizations between 1979 and 2001, but also on the adaptation of new data analytic techniques. Karell and Freedman used topic modeling, a form of unsupervised learning (Blei, Ng, and Jordan 2003; Mohr et al. 2013; Mohr and Bogdanov 2013). The authors, working in an abductive logic of inference, use this approach to “defamiliarize” existing theoretical concepts (Karell

and Freedman 2019:272) and generate new observations that a coding strategy modeled on existing theories and literature would have missed. Five of the 10 topics consist of words around war and international conflicts, consistent with the familiar rhetoric of subversion. Two of the remaining topics consist of words that are associated with local economic issues and private life (Karell and Freedman 2019:733).

To make sense of these results, the authors took a closer look at documents that are associated with these words and consulted additional social scientific theories about the ability of social discourses to foster social relationships. This led them to interpret the topics as “a sociotemporally intimate gaze, stitching together concepts of morality, the everyday, and the personal and local to challenge members of radicals’ own groups and communities, as well as the radicals themselves” (P. 734).

Visualizations. Powell et al. (2005) studied the expansion of the commercial biotech field between 1988 and 1999. The network Powell et al. analyze consists of dedicated biotech firms, venture capitalists, public research organizations, government research institutes, pharmaceutical companies, and financial institutions. Network ties occur when these different actors engage in collaboration agreements for different purposes, such as licensing, research and development, finance, and commercialization.

Preferential attachment, in which early success begets later success (Barabási 2002), became Powell et al.’s main theoretical explanation of network expansion (P. 1135). Three additional hypotheses from the sociological literature on economic ties and other social structures inform network growth: attachment bias may be due to homophily, following-the-trend, or multiconnectivity. The analysis of the expansion of the biotech network shows that the theory of accumulative advantage and the proposed multiconnectivity theory operate together in a way that is “suggestive of a systemic pumping action” (Powell et al. 2005:1183). Instead of rejecting an accumulative advantage mechanism in favor of multiconnectivity, or vice versa, as the initial theories would have suggested, Powell et al. find evidence of both.

The authors deploy novel network visualizations to arrive at this surprising insight. They use a “spring-embedded” network drawing algorithm that presents a network in a two-dimensional space as it places nodes with ties between them in close proximity and others further away (Powell et al. 2005:1154). They produce visualizations for six of the 12 years for which they have network data and describe each of their hypotheses in terms of those visualizations. They stress, however, that these visualizations only serve to illustrate their hypotheses and then turn to McFadden discrete choice models to test them. Yet, they acknowledged that their “figures suggest something the logit models cannot reveal—a very small core of perhaps one or two dozen organizations are routinely placed in the center, and their node size grows somewhat over the period. . . . The new tie visualizations do not reveal large nodes at the center, growing even larger as the field matures” (Powell et al. [2005:1189]; our emphasis). They call this group an “open elite.” This set of actors captures the main drivers of the field’s pumping action. The visualizations were able to reveal this open elite because they capture “the overall population rather than specific nodes” and thereby show “shifts in the dominant activities as well as changes in the composition of the nodes, which illustrate the overall trends in the

field" (P. 1156). Powell et al. use visualizations to find a core of elite actors and then abductively infer a pumping process of theoretically distinct mechanisms. They cite Bourdieu to make the connection between social fields and networks of relations but also explain that these relations are subject to social pressures.

In sum, researchers can stack the deck in their favor for creative theory development if they recognize empirical surprises in light of existing theories. In computational research, an abductive logic of inquiry can be cultivated when researchers take advantage of the scaling up of data sets and develop new analytical and visualization techniques to draw out counterintuitive findings. This requires leaning into apparent disadvantages of computational data and turning them into analytical opportunities. Although these are broad methodological tactics rather than precise methodological steps, abductive reasoning is a better fit for computational research than a deductive logic forcing the analysis in a narrow theoretical straitjacket or open-ended inductive reasoning that risks producing atheoretical findings. In the next section, we turn to an original research example to document abductive inferences as they unfold during the actual research project to show abductive inference in action.

Abduction in a Computational Analysis of Large-Scale Trace Data

The iconic New York City yellow cab industry rests on the work of thousands of taxi drivers working as independent contractors by leasing a cab from fleet operators. Taxi drivers are low wage workers with a relatively high level of autonomy and flexibility over their work. This industry makes a good case for showcasing abductive reasoning in quantitative research because of the presence of large-scale administrative data sets and a burgeoning research body using deductive approaches (for a review, see Occhiuto 2017).

Data

Our data set consists of all yellow cab trips during 2013. New York City taxi trip data have served as an empirical basis for social science research since the 1990s (e.g., Camerer et al. 1997; Farber 2005). Where earlier researchers were required to transcribe individual trip sheets, the New York City Taxi and Limousine Commission recently made digitally recorded trip data available (e.g., Buchholz 2018; Farber 2015). These data contain machine-generated, time-stamped, and geo-coded pick-up and drop-off information for each trip, the distance traveled, the fare amount, and additional fees. They also contain information on tips and the number of passengers, but this information is less reliable because tip information is primarily recorded for electronic transactions. Although large, minute, and rich, this data set lacks demographic characteristics of drivers that are necessary for answering many sociological questions.

Literature

Economists have examined the taxi trip data to understand labor decisions (Agarwal et al. 2016; Camerer et al. 1997; Chou 2002; Farber 2005, 2008, 2015). The core question animating this literature is what motivates taxi drivers to keep working a shift, considering that they have some flexibility in choosing their hours and their wages fluctuate daily and seasonally. Do taxi workers optimize the opportunity to work more hours and make more money (positive wage elasticity), or do they end their shift once they make enough money (negative wage elasticity)?

Camerer et al. (1997:407) observe that taxi drivers have daily income targets and once they hit those, they quit working, rather than maximizing their earnings. This is a counterintuitive finding in a field where the life-cycle approach to labor supply predicts that employees allocate leisure and work time across the life course to maximize their utility. The authors marshal social psychological factors on risk-taking and practical constraints to show the plausibility of a one-day income target rather than making as much money as possible.

In several response articles, Farber (2005, 2008, 2015) shows that Camerer et al.'s findings are due to erroneous methodological assumptions. Camerer et al. analyze labor supply at the level of working days. They regress a logged daily wage rate, a ratio of income to hours, on logged daily hours and find negative elasticities.² Farber (2005) instead calculates a model of the decision to stop work or continue driving at the conclusion of each fare. He finds that drivers stop working based on the number of hours they drove their cab, presumably because past earnings suggest future earning opportunities during the shift. The cab drivers' behavior is thus consistent with the intertemporal labor supply model.

Still, there are some weak spots in the data³ and the economists' perspective. The divide between the economists comes down to a question of unobserved variation in the data. Farber essentially recovered variation within shifts to show that Camerer et al.'s results are flawed. Both Camerer et al. and Farber, however, overlooked variation across longer time periods during which drivers didn't transport any passengers. These gaps in driving are irrelevant for the economists because they presume that drivers are unproductive—the gaps fall outside their scope conditions. Therefore, the economists dismiss these gaps as lunch or leisure breaks. Still, we don't know what the drivers did during that time. This weak spot may seem negligible from Camerer et al. and Farber's theoretical perspective, but it may matter for understanding the ups and downs of taxi work.

In contrast, sociological studies of taxi work have mostly focused on factors around daily labor decisions, such as the vulnerability of taxi drivers to criminal attacks or their need to solicit tips from unstable customer groups (Davis 1959; Hamill and Gambetta 2006; Occhiuto 2017). In an ethnographic study, Occhiuto (2017) encountered a taxi driver who chose this work to accommodate their night school schedule and another who returned to their home country for a few months a year. Occhiuto's study does not indicate how widespread these motivations are, though *New York Times* investigative reporting suggests that the poverty of many taxi drivers systematically constrains their work decisions.⁴ This qualitative research raises questions about how the work rhythms are influenced by the con-

straints, priorities, and opportunities facing taxi drivers that go beyond maximizing their take-home pay. How can we get at some of these life factors?

Analysis

Our methodological innovation is designed to reveal social dynamics that the economic approach might have missed. We look at how the work patterns of the taxi drivers may reflect their lives beyond making money. We designed a visualization that allows us to see all the trips throughout a day and across all the days of a year. The visualization builds on the intuition behind sequence analysis and uses the logic of heatmaps and the technical functions for line charts. Sequence analysis draws attention to patterns of states, such as school, apprenticeship, and work, during the life course. In the raw data, we only have two states for taxi drivers, passenger trips or not. Because we suspect more complex states in their lives, such as longer absences, we could not yet specify a sequence analysis. Line charts are designed to visualize temporal information as lines that are stable over time, change, or both. As our lines are part of sequences, they are straight connections between the start time and end time of a trip. Typical line charts with more than a dozen lines become illegible. Our line charts display up to a few thousand lines, depending on the number of trips a driver has made throughout a year. The collection of individually illegible lines creates images of taxi driver lives.

Initial exploration. Because temporal rhythms structure human lives (Zerubavel 1985), we use the x axis to locate trips at a specific time of the day between the early morning and midnight, and we stack days on top of one another so that each line on the y axis is a day of the year, from January to December. Each row thus shows a day in a cabbie's life; each plot shows the driving life of a cabbie during one year. Figure 1 shows the work-lives of three taxi drivers out of more than 100 pictures that we have generated. These three pictures are no outliers. More than 17,000 drivers in the data are active during 49 weeks or more during the year, similar to those in Figure 1(A) and (C). More than 6,000 drivers are active for only part of the year, between 22 and 38 weeks, similar to Figure 1(B).

Figure 1(A) comes close to the ideal-typical taxi driver work pattern from both Camerer et al. and Farber's perspectives. This driver begins a shift around 6:00 or 7:00 a.m. and usually stops between 4:00 and 5:00 p.m., with a short break just after 10:00 a.m. If we subtract that break (light gray area in the center of the figure), then this driver makes the eight to nine hours between the same starting and end times that Farber reports in his data (Farber 2005, 2015). This driver occasionally works until late into the evening, recalling the original motivation for economists of turning to the taxi industry to study "labor supply responses in settings in which workers are free to set their hours of work" (Farber 2005:47). This image does not tell whether the drivers drove longer on some days because they had not yet earned enough (Camerer et al. 1997) or because they still felt fresh enough to keep going and earn more (Farber 2005). But they are consistent with both theoretical perspectives in the economics literature.

Figure 1(B) shows a slightly different pattern. We again see some variation in the times when first trips start and last trips end, possibly reflecting that this driver,

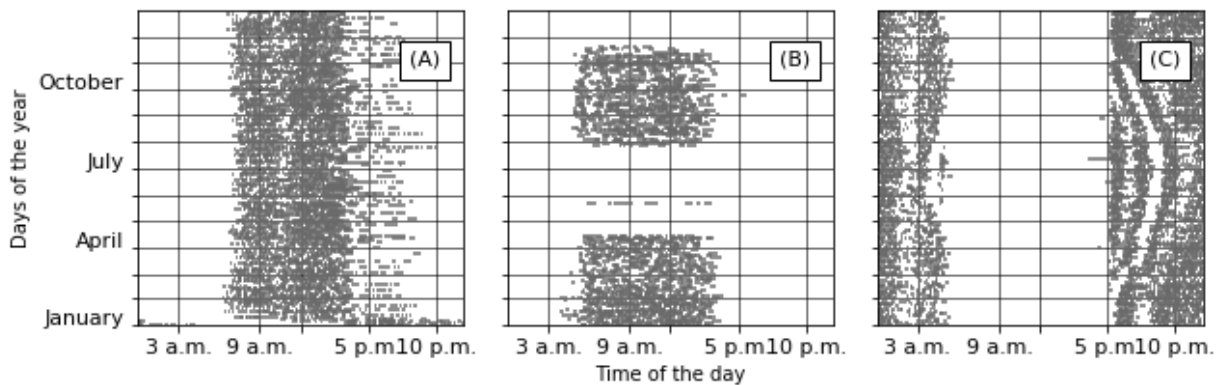


Figure 1: Taxi work during 2013 for three drivers.

too, chooses on some principle how much to work. We also see extensive white spaces, however, indicating no trip activity for weeks if not months at a time.

We don't know what drivers do when they don't drive. Camerer et al. and Farber labeled longer periods without trips lunch or leisure breaks. Although this decision suits a longer break within a shift, breaks of several weeks or months raise more questions. Some possible answers can be found in Occhiuto's (2017) ethnography of the yellow cab industry. Figure 1(B) would fit immigrants who work intensively for several months, save up money, and then return to their home country for a few weeks. It would also fit with workers who alternate cab driving with jobs in other industries, such as construction. Or alternatively, it may suggest that taxi driving is highly stressful or unhealthy and that drivers take long breaks. The visualization suggests that for a large number of taxi drivers, whether to end a shift may be less critical than how many shifts they drive on a yearly basis and what their alternative sources of income are. Taxi driving is embedded in alternative priorities and economic opportunities. At a minimum, this visualization begs for additional research to figure out whether taxi drivers following this seasonal rhythm opt for travel and rest (leisure) or for higher paying jobs.

So far, so deductive. Although Figure 1(B) shows how the economists may have missed seasonal patterns of taxi driving that may have ramifications for their interpretation of economical decision-making, hypotheses derived from existing theories can account for these fluctuations.

Abductive inference. Figure 1(C) calls for abductive inference. One obvious difference is that this driver works at night. After considering that difference, this picture constitutes a completely surprising driving pattern, unaccounted for in the existing economics literature. This pattern emerges from the regular breaks that this driver seems to take, as the systematic lack of passenger trips in the early evening indicates. Farber's (2015) operationalization would control for these breaks, even though they cannot be lunch breaks, as for the driver in Figure 1(A). Critically, the breaks seem to shift systematically over the course of the year. Although evening breaks start after the first recorded trip, at 5:00 p.m., initially, they appear at increasingly later times as the days pass. This pattern continues until halfway

through the year, when the breaks are earlier again. The breaks seem to follow the sunset.

Economists are sensitive to effects that weather may have on driving activity, controlling for rain and temperature. Sunset plays no role for economists or sociologists who study taxi driving. Why would a taxi driver always stop right at or around sunset?

Although we have seen this pattern in several pictures of our initial exploration, we don't know how common it is throughout the population of taxi drivers. To get a sense, we obtained the specific sunset times for New York City on each day in 2013, the year for which we have taxi trip data. Then we designed an algorithm that queries all work days of each driver who has generated more than \$30,000 in trip fares (to focus on drivers with substantial driving activity and hence systematic evidence of breaks). We considered drivers who have passengers before and after sunset, coding the days on which they have no passengers right when the sun sets. This step produced a sample of 23,410 drivers. Then we inspected the distributions of the share of observed sunsets across these drivers. We retained drivers who observe sunset breaks on 70 percent of the days on which they work before and after sunset, if there are 20 or more of those days. Of the initial 23,410 drivers, this strategy identifies 979, or 4.18 percent of the sample, as systematic sunset observers. What is so important to make almost a thousand taxi drivers in New York City interrupt their work at sunset?

Our abductive inference is that these taxi drivers time their breaks to engage in religious practices. Prayer times in Islam follow sunset for scheduling (Wiedemann and King 2012). We don't know the exact distribution of different religions among New York City cabbies, let alone how many observe their faith. We know, however, that the largest groups of New York City taxi drivers come from predominantly Muslim countries. In 2000, 84 percent of taxi drivers were immigrants. The most common countries of origin were Pakistan (14 percent) and Bangladesh (14 percent) (Schaller Consulting 2004). Even if we were to commit the ecological fallacy of inferring individual religious beliefs and practices among taxi drivers from these statistics, our estimate remains far smaller. And any coding strategy that uses sunsets necessarily produces false negatives but also false positives. Drivers cannot plan when they find a passenger and where their passengers want to go. Figure 2, which shows three drivers from our estimated sunset observations, illustrates that the driver in Figure 2(A) seems to resist the demands of his profession and almost always stops right when the sun sets (the dark curved line). But some will have to miss their sunset break because they have a passenger on a long trip, which may explain the inconsistencies of the driver in Figure 2(B). Others may occasionally have a break during sunset because there was no passenger. For the driver in Figure 2(C), the measurement detected systematic breaks around sunset that the visual inspection of trips for an entire year does not reveal, except between July and August.

To gain confidence in our interpretation, we asked how many drivers take systematic breaks at other times of the day, expecting lower numbers if there is no cultural reason. At 3:00 p.m., only three and half percent take regular breaks. This suggests that sunsets organize breaks more systematically. Taking another time,

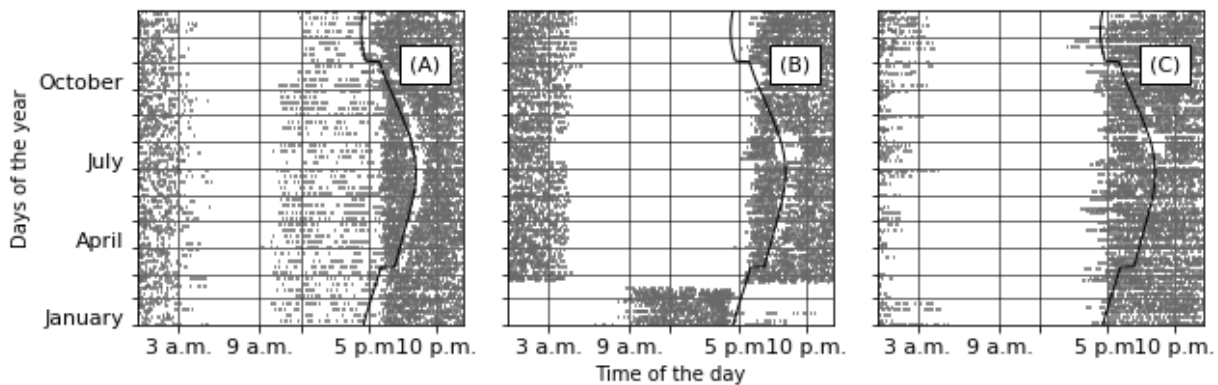


Figure 2: Three taxi drivers with systematic breaks around sunset.

10:30 a.m., however, more than six percent of relevant drivers regularly have no passengers (see Figure 3), suggesting that systematic breaks are at different times of the day and that the sunset times may have no distinct meaning. Specific points in time differ from breaks that shift systematically throughout the year, however. For example, if we consider breaks at 6:45 p.m., a specific point in time between the earliest sunset times in winter and the latest in June, we only retain about one percent of the drivers. Although the number of drivers who take breaks at sunset rather than a specific time in between was higher, we get a lower number, just below two percent, when we let the sun set in the afternoon. Imagining morning sunsets, to retain the shape of the breaks, we retain a higher number of drivers in absolute terms but lower than the one for the specific 10:30 a.m. break. The pattern around sunset is unique. And the high number in the morning is plausible as the time after the 8:00 a.m. rush hour is less busy for taxi drivers. It is a good time to take breaks (recall the driver in Figure 1(A)), and there will be longer pauses between trips. These breaks thus further strengthen our abductive process as they reveal another cultural practice, the morning coffee break, that also shapes labor supply decisions. This familiar practice enhances our confidence in an explanation based on the potentially less prevalent Islamic sunset prayer.

These additional steps strengthen our abductive inference and specify our estimate of drivers who observe sunset breaks. Further corroborating evidence comes from the observation that many drivers take much longer breaks during the late summer months (recall the driver in Figure 2(C)). In the context of religion, Ramadan is a time during which we would expect longer breaks. And indeed, Ramadan was held between July 8 and August 7 in 2013. We thus propose the hypothesis that the beginning of these longer breaks correlates with sunset, and in the summer sunrise, because of religious practices.

Our estimates remain preliminary as our main concern has been to demonstrate the abductive logic of inquiry. We have robust evidence that the sunset break patterns are systematic for a large group of drivers. The hypothesis that follows from this discovery in the context of the larger debate around taxi work might thus be that labor supply elasticities vary not only by economic opportunities but also by

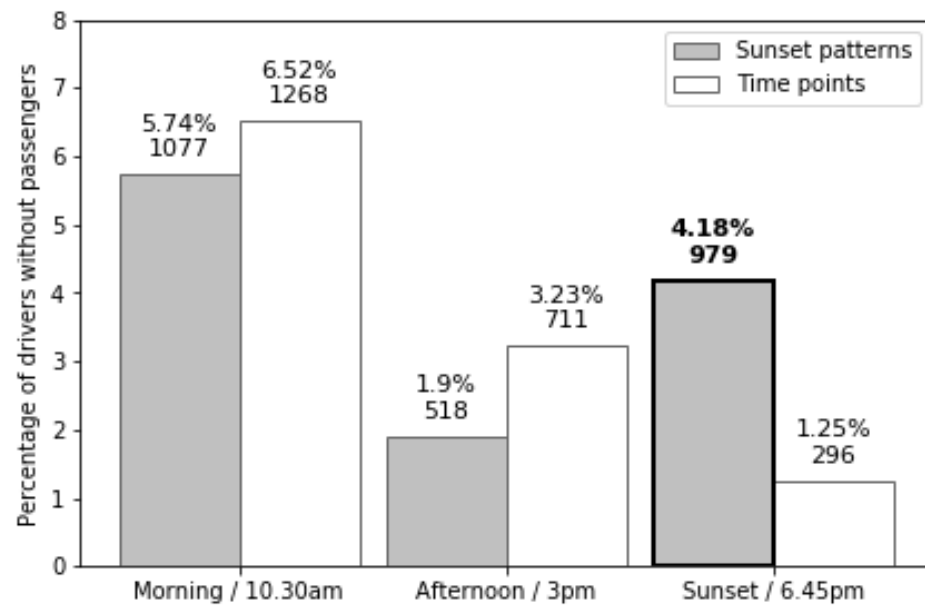


Figure 3: Systematic breaks at sunset and other times. Light bars indicate specific time points; gray bars indicate shifting time points that follow the sunset.

other sociological factors such as religion (see also Smith and Bender 2004). As part of the sociological challenge to economics (e.g., Stark 2009), this abductive inference may stimulate additional qualitative interview research to explore how drivers with different levels of religiosity and different religions decide to start or stop a shift or to conduct ethnographic research at mosques and check the cars in the parking lot. A survey among taxi drivers could then connect the economic and noneconomic activities, and perhaps their temporal unfolding, to the trace data to test competing hypotheses.

The taxi example shows how an unexpected finding may abductively open the analysis to a novel theorization. The economics literature did not predict any religious patterning of taxi shift work, and we could not get to this finding with a deductive hypothesis drawn from this literature. It just doesn't fit. From a purely inductive perspective, the visualization divorced from a theoretical framing could have been easily dismissed as noise rather than indicative of social patterning of the daily and seasonal rhythms of work decisions. An abductive inference rests on three intertwined strategies. First, in order to be attuned to a surprising finding, we need to familiarize ourselves with the existing literatures because a finding is only surprising in the context of previous work. In our case, this meant familiarizing ourselves with the economics literature but, once the surprising finding emerged, also drawing from the literature in sociology of religion. Second, for computational data, flexible analytic strategies are critical to get a handle on the data. Here, novel visualizations helped reveal unexpected data patterns. Third, unexpected findings then prompt the need for new hypotheses to explore deductively and inductively. The abductive inference is a hypothesis on probation requiring further exploration

and testing. In the taxi example, it may lead to theorizing all the different ways taxi drivers straddle the work–life balance beyond religion to include family obligations throughout the life course and to evaluate this balance against other employment options.

Conclusion

Quantitative computational research could continue its deductive agenda with some minor adjustments into the digital era. Researchers can derive hypotheses from theories and test them, selecting data sets or collecting observations that contain the relevant indicators and variables. We could approach the taxi data from a deductive perspective, wondering about race or gender effects, and if this information is missing, move on. (In fact, we were advised to pass on the taxi data for that reason.) Many found data sets lack traditionally relevant variables, especially demographic indicators, thereby stifling the research potential from a deductive perspective. An alternative path is to follow the inductive legacy of grounded theory and approach analyses unencumbered by pre-existing theories to let theoretical themes emerge spontaneously from research findings (Charmaz 2014).

A more fruitful way is to take advantage of the versatility of digital data and adopt an abductive logic of inquiry (e.g., Goldberg 2015; Wagner-Pacifici et al. 2015). An abductive inference is required when findings surprising in light of existing theory call for a new hypothesis and theorization. The key here is that we approach the data with relevant pre-existing theoretical notions. Abductive analysis is then the most creative but also most tentative form of scientific inference. It helps select theoretically relevant variables, winnows down possible hypotheses, and has the potential to break new theoretical ground, but it requires further testing to gain confidence in the hypothesis. Many researchers have serendipitously encountered and recognized surprises. By developing the abductive approach here systematically, we aid its wider adoption and render applications more rigorous. Abduction responds to recent calls for “embracing of variation” in statistical analyses (Gelman 2016) and enriches the explanatory ambitions in social sciences (Shmueli 2010).

Abductive inferences depend on the discovery of research findings unaccounted for in existing theories. In computational sociology, these can be cultivated by taking advantage of opportunities in large-scale data. The other tactic builds on new procedures that reveal patterns in the data, either quantitatively or visually. Our visualizations have illustrated an important challenge for these tactics. Modern programming packages make data work easier, such as designing plots that combine multiple dimensions (Healy 2018). Often immensely productive, these routines quickly impose the way of thinking of their designers on a specific sociological research problem (Martin 2019). Surprises require a flexible and innovative approach to quantitative data analysis, which, as it turns out, is a great fit for computational sociology.

Notes

- 1 In order to be surprised, researchers need to know what to expect based on existing theories. This same knowledge also allows the production of novel inferences. Acquiring in-depth familiarity with a range of theories goes beyond the scope of this article.
- 2 Farber notes two technical problems with this approach. First, he notes that income varies over the hours of a day and that a daily ration is therefore inappropriate. Second, both Farber and Camerer et al. note that the usage of hours on both sides of the equation may lead to a negative bias in case of misspecification or measurement error. Camerer et al. use an instrumental variable approach to address this issue, which Farber criticizes.
- 3 Both Camerer et al. and Farber initially relied on handwritten trip sheets provided by taxi leasing companies (70 trip sheets of 13 drivers for Camerer et al. and 584 trip sheets of 21 drivers for Farber). These are small, convenience-based data sets selected specifically to test a theory, considering that more than 51,000 cab drivers are licensed to drive the 13,000 yellow medallion taxis in New York City. However, Farber (2015) sampled electronic data of all taxi trips between 2009 and 2013 to come to the same conclusion that cab drivers respond positively to both anticipated and unanticipated increases in earnings opportunities, finding even stronger support for neoclassical theory in these more granular data by showing heterogeneity across drivers, with poorly optimizing drivers dropping out of taxi work.
- 4 Winnie Hu, "Taxi Medallions, Once a Safe Investment, Now Drag Owners into Debt," *New York Times*, September 10, 2017 (<https://www.nytimes.com/2017/09/10/nyregion/new-york-taxi-medallions-uber.html>); Brian M. Rosenthal, "As Thousands of Taxi Drivers Were Trapped in Loans, Top Officials Counted the Money," *New York Times*, May 19, 2019 (<https://www.nytimes.com/2019/05/19/nyregion/taxi-medallions.html>); Brian M. Rosenthal. "'They Were Conned': How Reckless Loans Devastated a Generation of Taxi Drivers," *New York Times*, May 19, 2019 (<https://www.nytimes.com/2019/05/19/nyregion/nyc-taxis-medallions-suicides.html>).

References

- Agarwal, Sumit, Shih-Fen Cheng, Jussi Keppo, and Ryoko Sato. 2016. "Learning by Driving: Evidence from Taxi Driver Wages in Singapore." Technical report, National University of Singapore.
- Bail, Christopher A. 2017. "Taming Big Data: Using App Technology to Study Organizational Behavior on Social Media." *Sociological Methods & Research* 46(2):189–217. <https://doi.org/10.1177/2F0049124115587825>.
- Barabási, Albert-László. 2002. *Linked: The New Science of Networks*. Cambridge, MA: Perseus.
- Black, Thomas R. 1999. *Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics*. London: SAGE Publishing.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.
- Buchholz, Nicholas. 2018. "Spatial Equilibrium, Search Frictions and Dynamic Efficiency in the Taxi Industry." Working paper, Princeton University.

- Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *Quarterly Journal of Economics* 112(2):407–41. <https://doi.org/10.1162/003355397555244>.
- Charmaz, Kathy. 2014. *Constructing Grounded Theory*. London: SAGE Publishing, 2nd edition.
- Chou, Yuan K. 2002. "Testing Alternative Models of Labor Supply: Evidence from Taxi-Drivers in Singapore." Working paper, University of Melbourne.
- Davis, Fred. 1959. "The Cabdriver and His Fare: Facets of a Fleeting Relationship." *American Journal of Sociology* 65(2):158–65. <https://doi.org/10.1086/222656>.
- Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." *Annual Review of Sociology* 46:61–81. <https://doi.org/10.1146/annurev-soc-121919-054621>.
- Evans, James, and Jacob G. Foster. 2019. "Computation and the Sociological Imagination." *Contexts* 18(4):10–15. <https://doi.org/10.1177%2F1536504219883850>.
- Evans, James A., and Jacob G. Foster. 2011. "Metaknowledge." *Science* 331(6018):721–725. <https://doi.org/10.1126/science.1201765>.
- Fann, K.T. 1970. *Peirce's Theory of Abduction*. The Hague: Martinus Nijhoff.
- Farber, Henry S. 2005. "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers." *Journal of Political Economy* 113(1):46–82. <https://doi.org/10.1086/426040>.
- Farber, Henry S. 2008. "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers." *American Economic Review* 98(3):1069–82.
- Farber, Henry S. 2015. "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers." *Quarterly Journal of Economics* 130(4):1975–2026. <https://doi.org/10.1093/qje/qjv026>.
- Friedman, Sam, Daniel Laurison, and Andrew Miles. 2015. "Breaking the 'Class' Ceiling? Social Mobility into Britain's Elite Occupations." *Sociological Review* 63(2):259–89. <https://doi.org/10.1111%2F1467-954X.12283>.
- Gelman, Andrew. 2016. "The Problems with P-Values Are Not Just with P-Values." *American Statistician* 70.
- Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing Company.
- Goldberg, Amir. 2015. "In Defense of Forensic Social Science." *Big Data & Society* 2(2). <https://doi.org/10.1177%2F2053951715601145>.
- Gray, Paul S., John B. Williamson, David A. Karp, and John R. Dalphin. 2007. *The Research Imagination: An Introduction to Qualitative and Quantitative Methods*. Cambridge, MA: Cambridge University Press.
- Hamill, Heather, and Diego Gambetta. 2006. "Who Do Taxi Drivers Trust?" *Contexts* 5:29–33. <https://doi.org/10.1525%2Fctx.2006.5.3.29>.
- Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton, NJ: Princeton University Press.

- Heiberger, Raphael H., and Jan R. Riebling. 2016. "Installing Computational Social Science: Facing the Challenges of New Information and Communication Technologies in Social Science." *Methodological Innovations* 9. <https://doi.org/10.1177%2F2059799115622763>.
- Hofman, Jake M., Amit Sharma, and Duncan J. Watts. 2017. "Prediction and Explanation in Social Systems." *Science* 355(6324):486–88. <https://doi.org/10.1126/science.aal3856>.
- Ignatow, Gabe. 2020. *Sociological Theory in the Digital Age*. New York: Routledge.
- Karell, Daniel, and Michael Freedman. 2019. "Rhetorics of Radicalism." *American Sociological Review* 84(4):726–53. <https://doi.org/10.1177%2F0003122419859519>.
- Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1(1). <https://doi.org/10.1177%2F2053951714528481>.
- Laurison, Daniel, and Sam Friedman. 2016. "The Class Pay Gap in Higher Professional and Managerial Occupations." *American Sociological Review* 81(4):668–95. <https://doi.org/10.1177%2F0003122416653602>.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323(5915):721–23. <https://doi.org/10.1126/science.1167742>.
- Lazer, David, and Jason Radford. 2017. "Data ex Machina: Introduction to Big Data." *Annual Review of Sociology* 43:19–39. <https://doi.org/10.1146/annurev-soc-060116-053457>.
- Martin, John Levi. 2019. "See It with Figures." *Contemporary Sociology* 48:607–12. <https://doi.org/10.1177%2F0094306119880195>.
- McFarland, Daniel A., Kevin Lewis, and Amir Goldberg. 2016. "Sociology in the Era of Big Data: The Ascent of Forensic Social Science." *American Sociologist* 47:12–35. <https://doi.org/10.1007/s12108-015-9291-8>.
- Milgram, Stanley. 1967. "The Small World Problem." *Psychology Today* 1(1):60–67.
- Mohr, John W., and Petko Bogdanov. 2013. "Introduction—Topic Models: What They Are and Why They Matter." *Poetics* 41(6):545–69. <https://doi.org/10.1016/j.poetic.2013.10.001>.
- Mohr, John W., Robin Wagner-Pacifici, Ronald L. Breiger, and Petko Bogdanov. 2013. "Graphing the Grammar of Motives in National Security Strategies: Cultural Interpretation, Automated Text Analysis and the Drama of Global Politics." *Poetics* 41(6):670–700. <https://doi.org/10.1016/j.poetic.2013.08.003>.
- Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45:27–45. <https://doi.org/10.1146/annurev-soc-073117-041106>.
- Mützel, Sophie. 2015. "Facing Big Data: Making Sociology Relevant." *Big Data & Society* 2(2). <https://doi.org/10.1177%2F2053951715599179>.
- Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1):3–42. <https://doi.org/10.1177%2F0049124117729703>.

- Occhiuto, Nicholas. 2017. "Investing in Independent Contract Work: The Significance of Schedule Control for Taxi Drivers." *Work and Occupations* 44(3):268–95. <https://doi.org/10.1177%2F0730888417697231>.
- Pajo, Bora. 2017. *Introduction to Research Methods: A Hands-on Approach*. Los Angeles: SAGE Publishing.
- Peirce, Charles Sanders. 1934. *Collected Papers of Charles Sanders Peirce. Vol. 5, Pragmatism and Pragmaticism*. Cambridge, MA: Harvard University Press.
- Powell, Walter W., Douglas White, Kenneth Koput, and Jason Owen-Smith. 2005. "Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences." *American Journal of Sociology* 110(4):1132–205. <https://doi.org/10.1086/421508>.
- Salganik, Matthew J. 2018. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.
- Savage, Mike. 2020. "What Makes for a Successful Sociology? A Response to 'Against a Descriptive Turn'." *British Journal of Sociology* 71(1):19–27. <https://doi.org/10.1111/1468-4446.12713>.
- Schaller Consulting. 2004. "The Changing Face of Taxi and Limousine Drivers." Report, Schaller Consulting.
- Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science* 25(3):289–310. <https://doi.org/10.1214/10-STS330>.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22(11):1359–66. <https://doi.org/10.1177%2F0956797611417632>.
- Smith, Elta, and Courtney Bender. 2004. "The Creation of Urban Niche Religion: South Asian Taxi Drivers in New York City." In *Asian American Religions: The Making and Remaking of Borders and Boundaries*, edited by Tony Carnes and Fenggang Yang, pp. 76–97. New York: New York University Press.
- Stark, David. 2009. *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton, NJ: Princeton University Press.
- Tavory, Iddo, and Stefan Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research*. Chicago: University of Chicago Press.
- Tavory, Iddo, and Stefan Timmermans. 2019. "Abductive Analysis and Grounded Theory." In *The SAGE Handbook of Current Developments in Grounded Theory*, edited by Antony Bryant and Kathy Charmaz, pp. 532–46. Los Angeles: SAGE Publishing.
- Timmermans, Stefan, and Iddo Tavory. 2012. "Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis." *Sociological Theory* 30(3):167–86. <https://doi.org/10.1177%2F0735275112457914>.
- Travers, Jeffery, and Stanley Milgram. 1969. "An Experimental Study of the Small World Problem." *Sociometry* 32(4):425–43.

- Van Maanen, John, Jesper B. Sørensen, and Terence R. Mitchell. 2007. "The Interplay between Theory and Method." *Academy of Management Review* 32(4):1145–54. <https://doi.org/10.5465/amr.2007.26586080>.
- Wagner-Pacifi, Robin, John W. Mohr, and Ronald L. Breiger. 2015. "Ontologies, Methodologies, and New Uses of Big Data in the Social and Cultural Sciences." *Big Data & Society* 2(2). <https://doi.org/10.1177/2F2053951715613810>.
- Watts, Duncan J. 2007. "A Twenty-First Century Science." *Nature* 445:489. <https://doi.org/10.1038/445489a>.
- Watts, Duncan J. 2014. "Should You Go to Grad School?" BuzzFeed, June 17. <http://www.buzzfeed.com/duncanwatts/should-you-go-to-grad-school>.
- Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393:440–42. <https://doi.org/10.1038/30918>.
- Weeden, Kim A., and David B. Grusky. 2005. "The Case for a New Class Map." *American Journal of Sociology* 111(1):141–212. <https://doi.org/10.1086/428815>.
- Wiedemann, E., and D.A. King. 2012. "al-Shafak." In *Encyclopaedia of Islam, Second Edition*, edited by P. Bearman, Th. Bianquis, C.E. Bosworth, E. van Donzel, and W.P. Heinrichs. Leiden: Brill.
- Zerubavel, Eviatar. 1985. *Hidden Rhythms: Schedules and Calendars in Social Life*. Berkeley: University of California Press.

Acknowledgments: We gratefully acknowledge detailed comments on earlier drafts from Fabien Accominotti, Alina Arseniev-Koehler, Byungkyu Lee, Kinga Makovi, and Josh Whitford, as well as anonymous reviewers and the Sociological Science editors. Karen Barkey, James Evans, and John Levi Martin, as well as audiences at the Séminaire de Sociologie Quantitative du CREST and at LIPADE, provided helpful feedback on initial ideas for this article.

Philipp Brandt: Department of Sociology and Center for the Sociology of Organizations, Sciences Po. E-mail: philipp.brandt@sciencespo.fr.

Stefan Timmermans: Department of Sociology, University of California, Los Angeles. E-mail: stefan@soc.ucla.edu.