$\mathfrak{E}$ 6 $\mathfrak{E}$

BEYOND TEXT: USING ARRAYS TO REPRESENT AND ANALYZE ETHNOGRAPHIC DATA

Corey M. Abramson*
Daniel Dohan†

Abstract

Recent methodological debates in sociology have focused on how data and analyses might be made more open and accessible, how the process of theorizing and knowledge production might be made more explicit, and how developing means of visualization can help address these issues. In ethnography, where scholars from various traditions do not necessarily share basic epistemological assumptions about the research enterprise with either their quantitative colleagues or one another, these issues are particularly complex. Nevertheless, ethnographers working within the field of sociology face a set of common pragmatic challenges related to managing, analyzing, and presenting the rich context-dependent data generated during fieldwork. Inspired by both ongoing discussions about how sociological research might be made more transparent, as well as innovations in other data-centered fields, the authors developed an interactive visual approach that provides tools for addressing these shared pragmatic challenges. They label the approach "ethnoarray" analysis. This article introduces this approach and explains how it can help scholars address widely shared logistical and technical

*University of Arizona, Tucson, AZ, USA
†University of California, San Francisco, San Francisco, CA, USA

Corresponding Author:
Corey M. Abramson, University of Arizona, The University of Arizona Social Sciences Building, 1145 E. South Campus Drive, Room 400, Tucson, AZ 85721, USA
Email: coreyabramson@email.arizona.edu

complexities, while remaining sensitive to both ethnography's epistemic diversity and its practitioners shared commitment to depth, context, and interpretation. The authors use data from an ethnographic study of serious illness to construct a model of an ethnoarray and explain how such an array might be linked to data repositories to facilitate new forms of analysis, interpretation, and sharing within scholarly and lay communities. They conclude by discussing some potential implications of the ethnoarray and related approaches for the scope, practice, and forms of ethnography.

1. INTRODUCTION

Recent methodological debates in sociology and related social science disciplines have focused on how data and analyses might be made more open and accessible (Duneier 2011; Freese 2007), how the process of theorizing and knowledge production might be made more explicit (Leahey 2008; Swedberg 2014), and the importance of developing means of data visualization in addressing these issues (Moody and Healy 2014). The presence of basic common understandings found in many quantitative approaches, such as concerns with replicability, reliability, generalization, inference, and validity, facilitates these discussions by providing a shared cultural basis for developing new tools (Durkheim [1893] 1984; Latour and Woolgar 1986). In ethnography,[1] where scholars from various traditions do not necessarily share basic epistemological assumptions about the research enterprise with either their quantitative colleagues or one another, these methodological debates are particularly complex.

Ethnographers wishing to work toward producing new tools for transparency, sharing, and visualization must first confront a lack of consensus about whether this form of scholarship can or should attempt to follow the traditional models of inquiry in the social sciences. Following the postmodern turn, a host of scholars have argued that ethnography should be seen as a field of humanistic inquiry rather than social science (Clifford and Marcus 1986). While recognizing the influence and importance of this critique (for a discussion, see also Reed 2010), in this article, we discuss shared challenges faced by those who argue that ethnography remains a viable social science method. Even

among these empirically inclined ethnographers, however, the question of how to situate ethnography relative to other methods is hotly debated. Some argue that concerns about validity, reliable representation, and generalizability are universal to social science and that ethnographers must contend with them in fundamentally the same way as scholars using other methods (Goldthorpe 2000; King et al. 2001; Sánchez-Jankowski 2002). Others argue that ethnography's value lies in discovering the unexpected and hidden, as well as in developing theory, and in these roles it provides a necessary and critical alternative to "positivist" social science (Burawoy 1998; Duneier 2011; Tavory and Timmermans 2009). Another group holds that ethnography is commensurate with, but different from, mainstream social science and thus necessitates a specialized logic and language of inquiry (Brady and Collier 2004; Lofland 1995; Small 2009). These divisions are deeply ingrained and often manifest as contentious public exchanges (cf. Becker 2009; Duneier 2002, 2006; Klinenberg 2006; Wacquant 2002). Consequently, ethnographers wishing to develop new tools must do so without many of the points of common ground (i.e., shared ontological, axiological, and epistemic assumptions) that researchers using other methods might take for granted.

In this article, we bracket debates about which approach to ethnography is most legitimate or appropriate. Rather, we focus on a core set of shared practical challenges faced by empirical investigators who aspire to data sharing, openness, and visualization. We focus here on two such practical challenges. First, ethnographers who do fieldwork must develop strategies to manage and make sense of the large volumes of context-rich data they collect over the course of research. Typically they do so without fully relying on quantitative data reduction techniques, as quantitative reduction is believed to strip data of the depth that makes it valuable in the first place. Second, ethnographers need to present evidence and communicate insights so readers will be engaged and convinced by their findings—a particular challenge, as typical ethnographic warrants mean that readers are often unfamiliar or misinformed about the social settings and actors that ethnographers describe (Katz 1997).

Following recent calls for a move from "tribalism" to pragmatic pluralistic engagement among divergent "qualitative" perspectives (Lamont and Swidler 2014), we grapple here with and introduce new techniques for addressing these challenges by proposing an approach that can help researchers identify, construct,[2] and present rich ethnographic data in a

way that allows analysts to make sense of patterns in their data, characterize them in a way that allows readers to appreciate their context and contingency, and open up possibilities for collaboration and data sharing. We do so with the full acknowledgment that although we have found the resulting methods to be a useful complement to traditional approaches, it is inconceivable that a single tool can serve as a common platform for all ethnographic analysis. However, we proceed under the belief that reconnecting to broader methodological debates about transparency, process, innovation, and visualization can enhance ethnography's contribution to social science.

In the pages that follow, we outline a new approach featuring an interactive graphical display for representing and sharing data that we call an "ethnoarray." The "ethnoarray" is loosely adapted from the microarray, a graphical "heatmap"-based approach that is used in the biological sciences to present large volumes of complex data.[3] Functioning as more than aesthetics, graphical displays potentially provide a flexible way for sharing information, seeing patterns, and blending narrative and explanation, characteristics recognized by quantitative analysts (Moody and Healy 2014; Tufte 1983, 1997). Like other visual display approaches currently being developed in qualitative social science and the humanities (Henderson and Segal 2013; Mohr and Bogdanov 2013; Tangherlini and Leonard 2013), the ethnoarray allows the display of data in ways that can facilitate the discovery of relationships and thus help researchers understand the contextual richness of their own data, whether those insights ultimately manifest as a richer interpretation of theoretical constructs or comparative analyses and causal explanation. In other words, the ethnoarray is a part of a growing class of visual-analytic tools that facilitate data exploration and can yield insights for both "confirmatory" and "exploratory" data analysis projects (Moody and Healy 2014; Tukey 1977).

Our goal in developing this approach goes beyond within-project discovery, however. The ethnoarray may also allow ethnographers to enjoy greater empirical and analytical transparency by providing new ways of sharing data with colleagues, readers, and the public (Freese 2007). That is, it may open new possibilities for the analysis, representation (Moody and Healy 2014), and sharing (Freese 2007) of ethnographic data, an arena in which there is still much room for innovation. In sum, we introduce the ethnoarray as a pragmatic tool for preserving and bolstering ethnography's traditional strength in sharing deep contextualized

narratives while speaking to new possibilities for exploring ethnography's relationship to causal analysis and the production of generalizable findings enabled by technological advances.[4]

To illustrate the ethnoarray approach, this article proceeds as follows: In section 2, we examine several pragmatic challenges faced by ethnographers and some extant responses. In section 3, we describe the microarray that is used in the biological sciences and explain how and why ethnographers can fruitfully adapt it for certain social science projects. We also provide an illustration using data from a five-year comparative ethnographic project examining the social and technical management of terminal illness. In section 4, we address the potential implications of the ethnoarray for current and future ethnographic practice, as well as the limits of this approach. In section 5, we conclude with a summary of how the ethnoarray and related approaches can lead to new possibilities for ethnographic representation and scholarly engagement that can go beyond traditional text. Two appendixes provide additional illustrations of how the array approach might be used for different units of analysis and cases.

2. BACKGROUND: APPROACHING ETHNOGRAPHY'S DATA CHALLENGES

Analyzing data produced during fieldwork creates substantial logistical challenges. Even brief episodes of ethnographic research can produce hundreds of pages of field notes or interview transcripts, as well as audio and video recordings, drawings, maps, or objects (Emerson et al. 1995; Sanjek 1990). Many ethnographers spend years in the field and produce commensurately large volumes of data. Upon commencing analysis, researchers cannot easily sample or thin their data lest they lose the richness that motivated the ethnographic engagement in the first place. Moreover, those researchers sensitive to the issue of generalizability may undertake multisite or comparative ethnographic studies as they explore counterfactual possibilities, flesh out social mechanisms, or seek to develop explanatory models (cf. Abramson 2015; Dohan 2003; Sallaz 2009; Sánchez-Jankowski 2008). Even though notes written by these researchers may address a more narrowly defined question, format, and unit of analysis that impart structure, the data's breadth and volume can still be daunting (Sánchez-Jankowski 2002). In either situation, as the notes, transcripts, recordings, documents, and objects

accumulate, analysts may struggle to focus on the experiences, themes, or patterns they care most about.

Ethnographers who successfully grapple with large volumes of data during analysis then confront a second pragmatic hurdle: how to share data and analyses in order to substantiate findings and conclusions. Social scientists typically illustrate how their findings were produced by sharing data, describing analytical procedures, or both. However, in ethnography, there is no widely shared standard regarding what constitutes appropriate analysis or representation. Whichever method is used may be criticized and rejected by those using alternative approaches. This is intensified by a growing "tribalism" found among camps of qualitative researchers (Lamont and Swidler 2014). This issue, combined with the unique role of the ethnographer as an instrument of data collection, has incited heated exchanges (Duneier 2002, 2006; Klinenberg 2006; Wacquant 2002). Even if ethnographers could agree on the value (perhaps even the morality) of pluralism, the volume, sensitivity, and context dependence of ethnographic data make sharing a nontrivial challenge. Despite repeated calls for making research processes transparent to peers, informants, communities, and the general public (Burawoy 2004; Duneier 2011), there is little consensus about how this can be accomplished.

Within the social science community, in which an expectation of open and shared data is widespread among quantitative researchers, the inability to share data easily has substantial implications for ethnographic claims-making (Becker 1958; Cicourel 1964; Sánchez-Jankowski 2002; Small 2009). The horns of this dilemma appear clear. Not sharing data raises concerns about validity, transparency, and even the veracity of fieldwork in a way that has the potential to delegitimize hard-won ethnographic findings. At the same time, sharing all of the ethnographic data generated by a research project is typically neither ethical nor feasible, nor is it necessarily useful; sharing small amounts of selected data diminishes interpretive richness and can impede understanding; and no singular protocol can describe the various interpretive processes through which analysts immersed in the field assess whether a behavior such as the contraction of an eyelid is a wink, a twitch, or a conspiratorial gesture (Geertz 2000).

Contemporary quantitative social science research is impossible to imagine without computers, but computing has had a relatively smaller impact in ethnography. The emergence of computer-assisted qualitative

data analysis software (CAQDAS) over the past two decades has pro-vided some promising new ways to analyze and present data. In terms of data logistics, a growing number of CAQDAS platforms help analysts enter, structure, code, organize, and retrieve large qualitative data sets including text and other evidence (Dohan and Sánchez-Jankowski 1998; Miles and Huberman 1994). New methods based on quantifying and conducting formal textual analyses have emerged as well (Franzosi, De Fazio, and Vicari 2012; Mohr and Bogdanov 2013; Mohr et al. 2013). However, even for those interpretivist and humanist analysts who are opposed to quantitative reduction, or even the notion of "coding" text (cf. Biernacki 2014), these platforms can potentially offer a way to organize and quickly cycle through voluminous data (Dohan and Sánchez-Jankowski 1998).

Although CAQDAS has provided new options for approaching eth-nography's data challenges, widely available commercial packages have limitations.[5] Although increasingly flexible, most commercial software emphasizes coding and retrieving textual "chunks" and exploring pat-terns of codes. In many ways, this originates from and reflects (and per-haps reinforces) an attempt to implement the code-heavy approach often associated with grounded theory (Reeves et al. 2008). In terms of shar-ing, CAQDAS can help investigators share data within a research team. Recently, software has even allowed networked collaboration in shared data clouds. Nevertheless, there has been less attention around how to share data with readers or other researchers. Although some software features the basic underpinnings of interoperability that can facilitate data sharing (such as an extensible markup language [XML] output), techniques for doing so are relatively nascent. Furthermore, CAQDAS output capabilities have not been widely invoked as a way to share data itself, that is, as a way to put data into the hands of readers and allow them to explore and reproduce analyses. Likewise, although online repositories for qualitative data are beginning to emerge as a location for hosting data (cf. Perez-Hernandez 2014), shared approaches for summarizing the data while protecting both context and confidentiality remain elusive.

What could help advance ethnographic inquiry beyond coding? First, although not all ethnographic approaches are concerned with mapping associations in data in either an exploratory or an explanatory manner, this is central to a number of qualitative approaches, ranging from grounded theory to more positivistic techniques. Analysts from

divergent camps frequently need support discovering patterns in their data, yet they also acknowledge that such support cannot come at the cost of disconnecting data from context. Second, among those concerned with transparency and replication, analysts need support for sharing data so that readers can assess claims making. Many contemporary CAQDAS packages provide tools for tagging themes and for team analysis, which constitutes a form of data sharing. Still, disseminating data more broadly is not a core goal of most CAQDAS packages available today, and shared methods that would make this more plausible have yet to be developed.[6] Finally, new approaches must go beyond the specific proprietary software architectures of CAQDAS platforms by offering adaptable public approaches that researchers can implement in their attempts to advance these conversations. In sum, although CAQDAS provides important tools for analysis, substantial opportunities exist for new tools and techniques to advance more open and transparent forms of ethnography.

2.1. *The Microarray: A Potential Tool from an Unlikely Source*

What might potential tools and techniques related to the microarray look like? Although ethnographers use a unique set of methods in their studies of social life, sharing and analyzing large volumes of context-dependent data is not a challenge unique to the social sciences. In molecular biology, a technique known as microarray analysis has proved powerful because it uses an interpretable heatmap visualization to help analyze and depict complex multilevel biological systems and processes to varied audiences. The term *microarray* refers to both a process for analyzing biological samples—typically patterns of gene expression in tissues—and a graphical product displaying results (Belacel, Wang, and Cuperlovic-Culf 2006; Eisen et al. 1998; Schena et al. 1995). The introduction of microarrays and their exploratory use has led to important advances. For example, microarrays helped scientists identify genetic patterns (overexpression vs. underexpression) in breast cancer tumors by analyzing and displaying expression profiles for a large number of tumor samples simultaneously (see Figure 1), differences that can help explain the course of illness, distribution within populations, and responsiveness to different types of therapy (Prat and Perou 2011).

Using microarrays, biologists can display, aggregate, analyze, and share complex, multilevel data using exploratory statistical procedures,
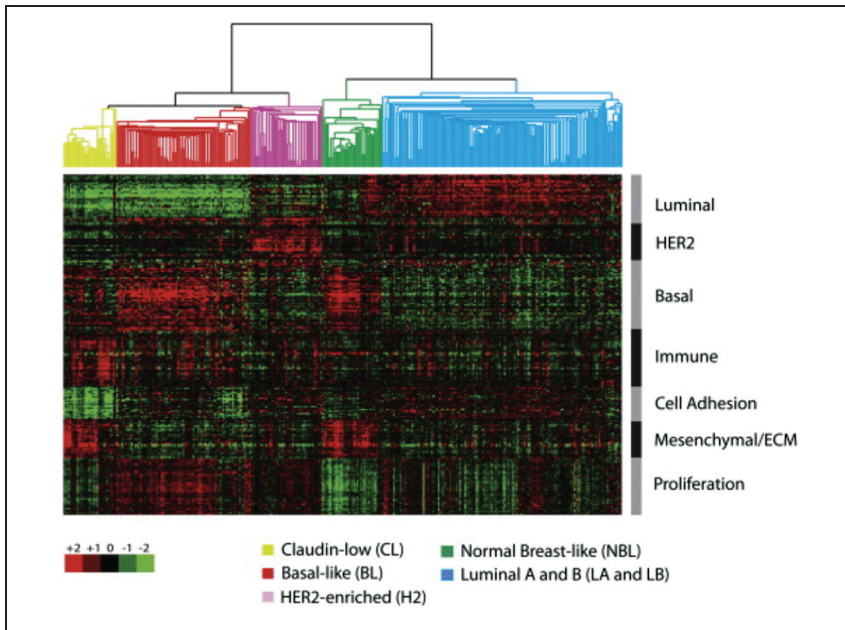
Figure 1. Microarray based on gene expression profiling data from 337 breast samples (in columns; 320 tumors, 17 normal tissues) and approximately 1,900 genes (rows).
Source: Prat and Perou (2011).

such as principal component analysis (PCA), that allow systemic inductive identification of group boundaries and pattern recognition (Stears, Martinskey, and Schena 2003). Figure 1 provides an existing example using gene expression data from breast tissue specimens.[7] At the same time, the microarray retains microlevel information about individual specimens so that analysts do not lose context. Analysts can "zoom in" to examine characteristics of an individual case in the array as readily as they can "zoom out" to see how that case fits within the array's overall pattern. Incorporating individual-level data within the microarray means that molecular biologists can use arrays not only to share findings but also to share the data and process by which those findings have been generated from voluminous underlying data.

Ethnographic field notes are quite different from the gene expression profiles found in microarrays. The former are typically more interpretative; the latter are expressed via quantitative reduction. Ethnography

necessarily involves self-reflection; there is no directly comparable activity for biologists. But analyzing either kind of data requires scholars to shift their gaze between distinct analytical levels and to represent their interpretations to a wider audience. Biologists use the microarray to examine genes, markers, individuals, and populations. Ethnographers examine microlevel interactions, emergent themes, theoretical constructs, and social contexts, and in this way they engage their sociological imagination to explore connections among and between behavior and narrative, group and organization, institution and society (Mills 1959).

3. ETHNOARRAY: AN EXAMPLE

Just as a microarray facilitates the multilevel exploration of biological data, we suggest that an ethnoarray may similarly facilitate, document, and reveal the richness of ethnographic data in ways researchers and readers find useful. Bearing in mind the caveats mentioned above, we use data from a study of the technological and social management of serious illness to develop an ethnoarray mock-up.

Our data are drawn from the Cancer Patient Deliberation Study (PtDelib), which uses ethnography to explore, understand, and explain how patients move along different treatment pathways with a specific interest in which patients end up embarking on clinical trials compared with seeking out less aggressive palliative care as they approach the end of life. All of the patients in our study have metastatic cancer and typically are within one to three years of death. The study uses ethnography to examine not only interactions between providers and patients but also the physically and analytically distant social processes that structure those interactions, and how these are understood by actors, with an ultimate goal of tracing how happenings in the exam room reflect the institutional contexts of patients and clinicians.

Patients are recruited to the study as their disease progresses and as treatment options begin to dwindle. Recruitment occurs in person during a routine clinic visit, and patients are then followed longitudinally. The study uses a multifaceted approach. Data consist of semistructured interviews with patients conducted at multiple points in time, direct observation of clinical encounters (including the recruitment visit), a semistructured interview with a family member or caregiver, review of medical records, and surveys administered at each interview with

patients and caregivers. The PtDelib cohort includes 82 patients as well as 31 caregivers and 63 providers. We have recorded approximately 4,000 pages of observational field notes and 8,000 pages of interview transcripts.

The research team includes four fieldworkers and three researchers who review and analyze transcripts and field notes. We use commercial qualitative data analysis software (ATLAS.ti) to organize the data, and we developed a coding scheme using both deductive and inductive techniques to facilitate retrieval of field notes, transcripts, and other data. This database must support multiple analytical goals and be accessible to multiple audiences. The study's research team and audience span diverse disciplines, including sociologists, bioethicists, linguists, health services researchers, and medical professionals. Consequently, PtDelib findings need to be interpretable and responsive to various viewpoints and questions along a continuum of analytical approaches. We began development of the ethnoarray approach as a way to address this core project need, but we found that its utility extends beyond this goal.

3.1. *Developing an Ethnoarray*

Using preliminary data from the PtDelib study, we developed a small-scale model of an ethnoarray. All of the data we use in this mock-up are drawn from field notes and interviews we had previously entered and coded via an iterative interpretive analysis using the ATLAS.ti software platform. An example of the coded database is presented in Figure 2, which shows a single paragraph from the transcript of a PtDelib patient interview. As this figure illustrates, passages of text typically include many codes. The software allows analysts to flexibly search the database to retrieve passages of interest using Boolean search procedures and even basic inductive tools such as co-occurrence tables. However, given that the ethnoarray involves a new approach, current software packages are not designed at this time to directly facilitate the production of ethnoarrays. In translating the data into an array, our first—and most analytically consequential—decision for this mock-up was to focus the ethnoarray on analyzing patients' trajectories into clinical trials. That is, we chose to organize this array to facilitate understanding differences and similarities between individuals.[8] After selecting the unit of analysis for the array, we chose five substantive domains that prior literature and our early iterative interpretive analyses suggested
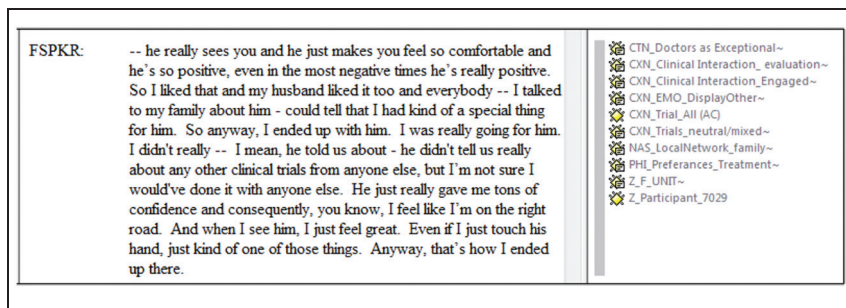
FSPKR: -- he really sees you and he just makes you feel so comfortable and he's so positive, even in the most negative times he's really positive. So I liked that and my husband liked it too and everybody -- I talked to my family about him - could tell that I had kind of a special thing for him. So anyway, I ended up with him. I was really going for him. I didn't really -- I mean, he told us about - he didn't tell us really about any other clinical trials from anyone else, but I'm not sure I would've done it with anyone else. He just really gave me tons of confidence and consequently, you know, I feel like I'm on the right road. And when I see him, I just feel great. Even if I just touch his hand, just kind of one of those things. Anyway, that's how I ended up there.

CTN_Doctors as Exceptional~
CXN_Clinical Interaction_ evaluation~
CXN_Clinical Interaction_Engaged~
CXN_EMO_DisplayOther~
CXN_Trial_All (AC)
CXN_Trials_neutral/mixed~
NAS_LocalNetwork_family~
PHI_Preferances_Treatment~
Z_F_UNIT~
Z_Participant_7029

Figure 2. A single paragraph from the transcript of a Cancer Patient Deliberation Study patient interview, coded as an ATLAS.ti data set.

were relevant, discussed with team members how to properly represent those domains, ultimately selected three to four measures for each domain, and arranged the domains and measures as the array's 16 rows.[9]

For columns, we selected a sample of patients for whom we had sufficient data (i.e., for whom we had at least two interviews and field observations before they either died or left the study). We then debated how to represent time variation in their experiences. In Figure 3, we show an array in which all information has been aggregated into a single column (to create a 10-column array); Figures 4 to 6 show arrays in which patients' experiences and statuses at different times are shown in distinct columns (baseline [T1] and first follow-up [T2]), which creates a larger and perhaps less intuitively interpretable array that includes greater richness about patients' experiences and trajectories. The key to Figures 3 to 6 is provided below. Each cell reflects all interview and participant observation data associated with that individual, domain, and measure. The domains, measures, and rows are ordered with a temporal logic following our particular research questions, but future arrays need not follow this model. The rows of a grounded theory approach, for example, might include general emergent themes generated entirely inductively. Columns could represent organizations, events, interaction sequences, or other units depending on the analyst's goals. Appendix A shows a brief example of how ethnoarrays can use other units of analysis (such as neighborhood contexts) in a comparative participant observation study, but for the sake of clarity, we focus on individual trajectories from PtDelib data in the main text.
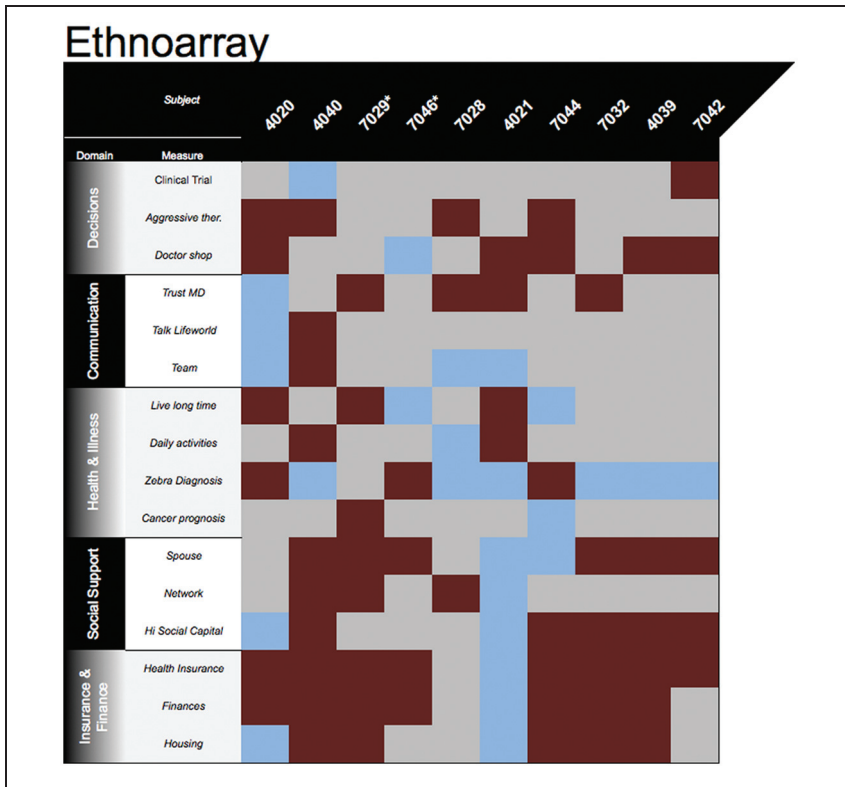
Figure 3. An ethnoarray based on data from 10 participants in the Cancer Patient Deliberation Study. The "Key to Figures 3–6" on p. 288 provides additional information regarding the domains, measures, and color assignment.

The sample array we present is organized by discrete units (characteristics and experiences of individuals at a given point in time evidenced by field notes and interviews). This corresponds with the analytical goals of the PtDelib project but raises important points about array construction. First, which type of data can be included in arrays? In any study, analysts must answer this question on the basis of the particular research question. For illustrative purposes here, we included only traditional ethnographic data derived from field observations and interviews, but the larger study also includes data from medical records, focus groups, and surveys, which could also be incorporated. A related question is which portions of data become part of the array. Again, analysts will make this decision on the basis of the nature of their study and
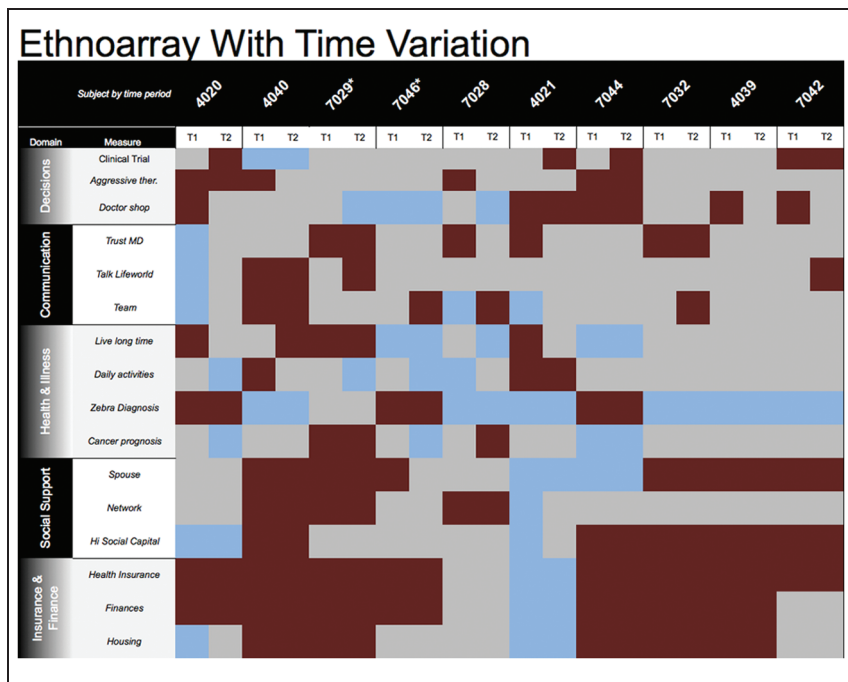
Figure 4. An ethnoarray based on data from 10 participants in the Cancer Patient Deliberation Study, with baseline (T1) and follow-up (T2) data. The "Key to Figures 3–6" on p. 288 provides additional information regarding the domains, measures, and color assignment.

question. Their decision will likely reflect the specific epistemic and intellectual tradition within which they situated their work.

For the sample array in this article, we used a broadly interpretive approach. We examined coded data in the ATLAS.ti database and narrative summaries of each patient's experience, and we held discussions among the team of researchers and fieldworkers who had firsthand knowledge of the patients, providers, and clinics represented in the array. Within this broad contextual framework, we interpreted specific interview passages and field notes according to whether and how they were related to array domains. We used all such passages in constructing the arrays in this article. The temporal structure of the array reflects the longitudinal design of the study, in which interviews and observations were conducted in a coordinated sequential fashion. Others who use ethnoarrays need not follow this model. Analysts might use only a
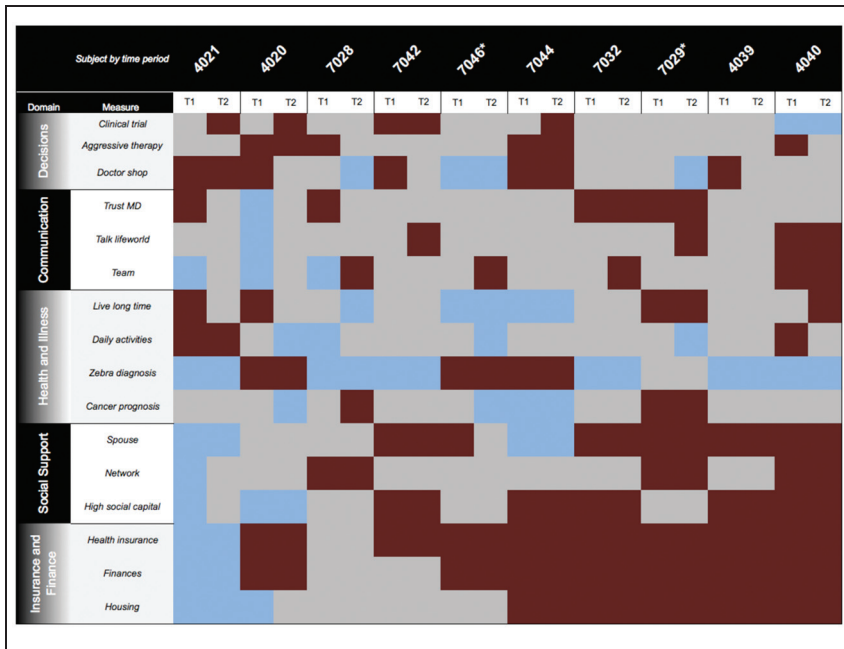
Figure 5. A sorted ethnoarray based on data from 10 participants in the Cancer Patient Deliberation Study. The "Key to Figures 3–6" on p. 288 provides additional information regarding the domains, measures, and color assignment.

single form of data (e.g., interviews or field notes). They might choose to deal with the question of inclusion differently as well. They might use formal linguistic tags to aid in categorization rather than relying on interpretive coding. "Uncoded" data (e.g., data that are not formally categorized according to substantive domains but are still associated with the columnar unit of analysis, such as individuals or neighborhoods) could be linked under a broad category labeled "other" (again, see Appendix A). Finally, researchers might decide that including all data for a person, site, and so on, is not feasible, ethical, or relevant. In this case, they would still be able to share summaries and still provide data beyond those in the traditional ethnographic report, but the array would not be inclusive. In short, like the quite varied notion of "coding," the array is a flexible tool whose use depends largely on researcher decisions and justifications (i.e., what to examine, how to measure it, and how to represent it) that parallel those found in other forms of social science (Cicourel 1964).
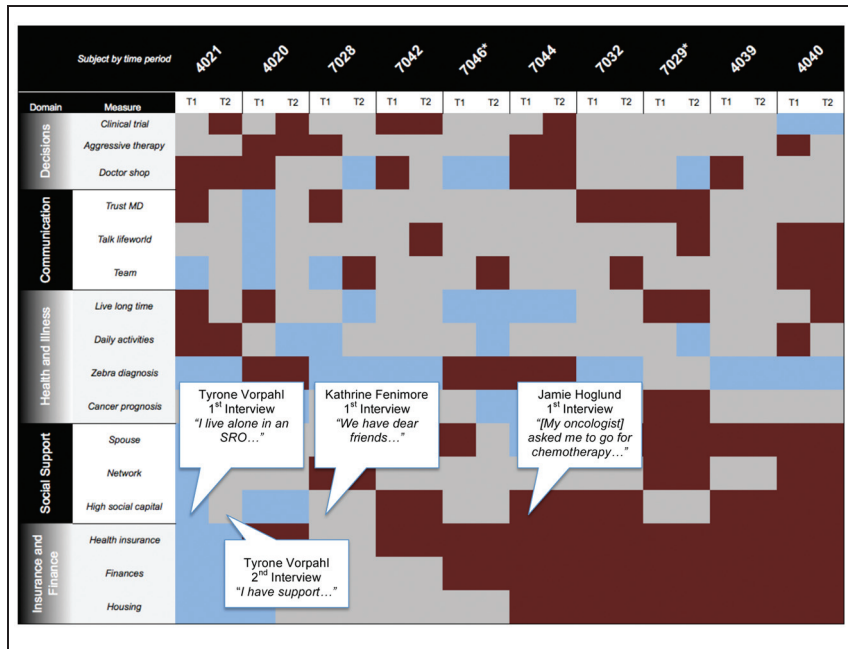
Figure 6. A sorted ethnoarray with selected quotations from linked data. The "Key to Figures 3–6" on p. 288 provides additional information regarding the domains, measures, and color assignment.

With a unit of analysis selected, inclusion criteria identified, and rows and columns defined and ordered, a final design decision concerns how to assign colors to the resultant cells in an analytically useful way. The goal of color coding in this array is not simply aesthetics; it is also to enable a visual summarization to facilitate examining data patterns. To this end, the model ethnoarray features a three-color matrix that indicates the degree to which a given characteristic is present (blue = less; gray = typical or unremarkable among study participants; red = more). These colors were based on the team's review, discussion, and interpretation of each patient's case and associated transcripts and field notes. That is to say, similar to the procedures used for deciding on data inclusion, color assignment in this particular example was based on the interpretations of fieldworkers who were deeply familiar with the site and individuals rather than formalistic procedures or automated approaches for text mining. Existing coding in our CAQDAS data set, which already tagged text according to themes of interest in the project, were used to

Domain	Measure	Color Assignment**
Decisions	Clinical trial	red = more aggressive therapy; blue = less aggressive
	Aggressive therapy	red = more doctor shopping or change in doctor; blue = relatively stable in doctor choice
	Doctor shop	red = more serious consideration of trial; blue = little consideration of a clinical trial
Communication	Trust MD	red = express/display trust in provider; blue = express/display mistrust in provider
	Talk lifeworld	red = communication includes topics other than illness; blue = talk is only of cancer
	Team	red = good provider-patient communication; blue = poor communication
Health and Illness	Live long time	red = live a long time; blue = end of life
	Daily activities	red = daily activities unaffected by illness; blue = daily activities are highly impacted
	Zebra diagnosis	red = unusual or rare condition; blue = unremarkable diagnosis
	Cancer prognosis	red = patient believes cancer is slow-growing; blue = patient believes cancer is fast-growing
Social Support	Spouse	red = highly involved spouse; blue = no spouse or spousal support available
	Network	red = large non-spouse network; blue = no non-spouse network
	High social capital	red = network w/high social capital; blue = lower capital in social network
Insurance and Finance	Health insurance	red = comprehensive medical insurance coverage; blue = poorer coverage
	Finances	red = finances are quite stable; blue = experiencing financial instability
	Housing	red = highly stable housing; blue = less housing stability including changes in housing

* Indicates participant is deceased

** Based on the interpretive procedures described in the text. Grey is typical for the sample of participants, red indicates higher or more intense than normal, blue indicates lower or less intense than typical.

Key to Figures 3–6. Overview of domains, measures, and color assignment.

help identify information about each cohort member and provide a level of confidence that we were not overlooking relevant data as we developed our interpretations of patients' experiences and understandings within the measures of each domain.

Embedding a traditional interpretation of rich ethnographic data within a structured tabular framework of domains and measures is, of course, only one approach to achieving a balance between more interpretive and formalistic approaches to analysis.[10] Other approaches may involve an explicit scoring procedure for determining cell color, for example, density or co-occurrence of codes from a CAQDAS database, linguistic algorithms, or word frequency counts. Color assignment schemes could be developed to indicate the absence of data for a

particular theme of interest, for example, to capture different degrees of theoretical saturation or other types of unevenness in data collection that must be addressed in analysis. The tabular format may not be suitable for some projects, including those that do not have a clearly identified unit of analysis or take a more humanistic approach. Still, even within the broad spectrum of approaches falling under the umbrella of sociological participant-observation, many studies maintain a clear unit of analysis and examine variation within and across groups or contexts (cf. Abramson 2015; Cicourel 1968; Dohan 2003; Lareau 2011; Lutfey and Freese 2005; Sallaz 2009; Sánchez-Jankowski 2008).

Inset 1 outlines several different approaches to color assignment and illustrates how the ethnoarray can be used with a variety of analytical styles, from formal rules geared toward the quantification of observed behavior to flexible integration of interpretative insights. The ability to integrate and even combine these divergent approaches exemplifies the ethnoarray's ability to accommodate different analytical approaches, goals, and styles from diverse intellectual and epistemic traditions.

INSET 1: STRATEGIES FOR COLOR ASSIGNMENT

A fundamental task in constructing any ethnoarray is deciding upon a procedure for assigning cell color. A variety of techniques are possible, with different implications for different methodological approaches. Three common strategies drawn from existing forms of qualitative analyses include the following:

1. Analyst imputation and interpretation
2. Counting (frequency or co-occurrence)
3. Scale construction

It is important to note that just as these strategies represent different approaches for representing ethnographic data, color assignment does not necessarily privilege theory testing or deductive inference (although it may facilitate such an approach). As we illustrate below, color assignment can summarize valences of text without reducing it to quantitative information and can thus remain connected to the underlying narratives and subject-centered accounts. It can also be used as an exploratory as well as an explanatory tool. In either case, color assignment is a key part of making sense of observations,

---

seeing how they are related, and communicating findings. Yet it does not devalue or replace the underlying qualitative data.[11]

In the interpretive strategy, an analyst (or a group of analysts) uses his or her own understanding of the social scene being analyzed, as well as judgment and experience based on immersion in the field, to constitute measures and domains and to assign colors to cells. This is the approach we followed in the mock-up ethnoarrays shown in Figures 3 to 6. Ethnographic data have often been analyzed using interpretive approaches that focus on meanings and the explanations that people give for their actions (e.g., Geertz 2000), which are reflected in this mode of color assignment. These analyses often do not rely on explicit rules but, rather, focus on creating a coherent understanding based on research subjects' actions and accounts. Interpretative strategies have also been formalized. In several strains of grounded theory, patterns within ethnographic data are identified inductively and leveraged to construct theories or explanations of social phenomenon (Glaser and Strauss 1967; Strauss and Corbin 1990). A grounded theory approach to color assignment might start by grounding emergent measures and domains. Although interpretative approaches typically rely on researchers' assessments rather than formal rules to assign colors, pattern analysis of the ethnoarray can still be done in a systematic or even rule-bound way (as in the PtDelib example below). A potential drawback of interpretive strategies is that some audiences may find their procedures opaque and, therefore, the findings unconvincing. Interpretative strategies, which necessarily involve substantial amounts of manual analysis, may also prove more burdensome for researchers than approaches that automate the color assignment processes.

The second strategy for color assignment relies on counting: calculations of the adjusted or unadjusted frequency of codes. Because the raw number of occurrences for the given outcome, factor, or phrase being measured is often less important than its likelihood to come up under particular circumstances in ethnographic research, analysts may find density or co-occurrence functions preferable to raw counts. An application of this approach is found in forms of conversation analysis that focus on the semiotics, words, and meanings of ethnographic data via procedures governed by formal rules about phrasing patterns, turns of speech, or thematic proximity. A conversation analysis approach to color assignment would use formal rules, for example, how frequently a phrase appears in a fixed segment of

the database (a co-occurrence density function) or the proximity of key words. This approach could inform statistical techniques for automated color assignment so that arrays could be constructed without an analyst having read through all the data. This could also potentially facilitate approaches such as QNA (Franzosi et al. 2012) or topic modeling (Mohr and Bogdanov 2013; Mohr et al. 2013). The risk with frequency and co-occurrence functions is the increased possibility of generating erroneous associations based on statistical rather than substantive associations (i.e., false positives and the validity issues associated with automated imputation).

The third way of assigning colors is to construct scales using a series of measures. This approach comes closest to replicating strategies most commonly used in quantitative social science. Color assignment could reflect nominal, ordinal, or interval scales related to observable and measurable acts. In the PtDelib study, a scalar approach might assign colors on the basis of whether a cancer patient sees a doctor when she feels a lump in her breast (nominal), a person's highest academic degree (ordinal), or how many times clinical trials are mentioned to a potential enrollee (interval). In preliminary tests, with PtDelib data, we have found that when these measures are binary and comparatively objective (e.g., whether someone went to a doctor), the reliability of color assignment increases. Pattern analysis of the ethnoarrays produced from scalar color assignment can leverage the informational richness created through the use of nominal, ordinal, or interval measures. This increases the specificity with which results can be reported and may improve the credibility of reported findings for some audiences. On the other hand, this approach involves a substantial risk: the process of manipulating data into scales may make it harder to appreciate or interrogate the contextual richness that spurred ethnographic engagement in the first place. That is, it forces rigidity on the data that can impede connecting with the object of study (Cicourel 1964). Ethnoarrays may thus reasonably be seen as a complement to, rather than a replacement of, existing ethnographic approaches.

In practice, analysts likely will experiment with and combine approaches to color assignment. For instance, the PtDelib study primarily relies on an interpretative approach for color assignment, but we specified some domains and measures deductively, and when possible (e.g., when constructing measures of social standing), we developed simple scales to govern color assignment. If widely adopted, it

would be important to investigate the advantages and drawbacks of developing arrays that shared common parameters and allowed cross-study comparisons.

3.2. *Reading the Ethnoarray: The Experiences of Wayne Burley*

Once constructed, the ethnoarray can be used in numerous ways to understand and represent large volumes of data. For analysts interested in examining and visually representing the experiences of specific patients to understand an outcome (e.g., whether they enter a clinical trial), data can be read along a single column or across the adjacent columns of a single patient's baseline and follow-up data. This can be useful to both contextualize interpretive insights and to provide information about the validity of inferences drawn using statistical methods that can group like trajectories such as sequence analysis (SA).

For example, the first columns of Figure 4 illustrate the experiences of Wayne Burley[12] (the array shows his study identifier number, 4020) derived from analysis of two in-depth interviews with him, an interview with his live-in girlfriend Heather Okeefe, and direct observation of two appointments around the time of the interviews with his oncologist, Antonio Akin, who was also interviewed (as well as observed in numerous other interactions with patients and colleagues). Wayne and Heather had been living on opposite coasts prior to his cancer diagnosis, and over the course of several months, he lived for short periods in three cities as he sought diagnosis and treatment for his cancer (a rare form of the disease; note the red cell under *Health and Illness>Zebra diagnosis*). He finally settled in northern California to make it easier for Heather to care for him, and the couple moved from a small one-bedroom apartment (where we conducted his baseline interview) to a larger two-bedroom apartment by the time of his follow-up interview (at which point we also interviewed Heather). Given this history, we characterized his *Insurance and Finance>Housing* as relatively unstable with respect to the others in our cohort at the baseline interview (blue cell) and more typical by follow-up (gray cell). Wayne had a long career with a government employer, retired early, and had begun a second career teaching public school when his cancer was discovered. Although no longer working because of his illness, Wayne's employment history provided a stable pension and generous health benefits; we thus classified his *Insurance and Finance>Finances* and *Insurance and*

Finance>Health insurance as higher than typical (red cells). Our interpretation of Wayne's situation in the domains of *Social Support, Health and Illness, Communication*, and *Decisions*, drawn in comparison with dozens of other study participants, can be read in a similar fashion by examining the color of each relevant cell.

Wayne and Heather's first visit with Dr. Akin was among the most contentious we have seen in the PtDelib study. Wayne relocated to northern California in part because Dr. Akin is acknowledged as an expert in his unusual cancer, but Wayne also received treatment from other oncologists. Before meeting Wayne and Heather for the first time, Dr. Akin reviewed Wayne's medical record in the clinic hub room (a physicians' work room out of earshot of patients) and commented to one of his colleagues that the other oncologists had been overly aggressive "cowboys" in their treatment approach. Although frank commentary is the norm in the hub room, we observed Dr. Akin repeat his "cowboy" comment to Wayne and Heather in the exam room. Their interactions became tense as a result, something Wayne and Heather commented on after the encounter. They both discussed this uncomfortable visit during their interviews but acknowledged that their relationship with Dr. Akin improved with time. They characterized him as a "straight talker" whose frank assessments of Wayne's progress and prospects were valuable, and they brushed aside his more insensitive remarks. In the ethnoarray, this trajectory in their relationship is reflected in the *Communication* domain, which we coded blue at baseline (indicating atypically poor communication) and gray (typical communication) at follow-up.

The ethnoarray also reflects other changes between baseline (T1) and follow-up (T2) observations. Initially Wayne's daily activities continued uninterrupted, and he believed that his life span would be unaffected by his illness (*Health and Illness>Daily activities* and *Health and Illness>Live long time*). At follow-up, he was experiencing substantial fatigue and was unable to do many of the things he had enjoyed just a few months previously (we characterized this as a blue cell for *Health and Illness>Daily activities*); like some of the patients in our study, at this point he acknowledged that his cancer would not be cured (*Health and Illness>Live long time* is now gray). Finally, examining the *Decisions* domain, we note that Wayne remained aggressive in his approach to his illness but that he seemed to be less interested in finding other doctors to manage his treatment; at his follow-up interview, he said he planned to stick with Dr. Akin. Although Wayne initially had

said he was not interested in participating in a clinical trial during his baseline interview, a few months later he had begun to actively research trials to join (*Decisions>Clinical trial* is gray at T1 and red at T2). The resulting representation in the array summarizes key aspects of Wayne's trajectory and provides a useful visualization that helps contextualize his experience relative to other subjects. This also facilitates further pattern analysis and possibilities for data sharing that we now examine.[13]

3.3 *Relational Mapping*

The colored cells of the array can be used not only for reading narratives but also for mapping and understanding relationships among actors, institutions, and concepts, a fundamental goal for many ethnographic approaches. Take, for example, Wayne Burley's relationship with his physician. The array visualization summarizes that Wayne's trust in his physician changed over time, and on the basis of preliminary study data, it appears other patients have experienced similar shifts. Moreover, the array allows analysts to readily see that these relationships of trust occur not only between patients and their physicians but also in patients' experiences as members of the health care team that includes physicians, nurses, and other health care providers. Analytic memos provide one way of documenting and interpreting an individual patient's experiences of these relationships. The array provides a way to supplement those analyses by considering broader contextual elements that might also influence these experiences.[14]

   Our preliminary analyses suggest that patients' experiences of trust and team membership reflect their estimations of physicians' competence and the congruence between clinicians' treatment preferences and their own, but these factors do not operate in a simple or mechanistic way. The array allows analysts to examine patients' experiences of trust and team membership within a much broader context, for example, whether their cancer is affecting their daily activities, how the progression of illness over time shapes these relationships, and how patients' own beliefs about whether their illnesses are life limiting color the trust and connection patients feel with their clinicians. For Wayne Burley, the progression of his illness, its impact on his daily activities, and the exhaustion of available treatments appeared to reshape his engagement with his oncologist and care team. Our preliminary data suggest that

Burley is not alone in this experience of illness progression—an element of physiology that can force patients of diverse values and expectations to rethink coping, engagement with family caregivers, and their relationship to the illness of cancer itself.

3.4. *Arranging and Sorting to Examine Patterns*

In addition to summarizing and representing ethnographic data and facilitating the interpretation of relationships, arrays can help bridge quantitative and qualitative analytical techniques by allowing researchers to combine statistical techniques for pattern recognition with interpretation of the underlying field notes and transcripts. It is important to recognize that although observations are tagged, they are not "reduced" to numbers or codes. That is to say, code patterns are meant to be orienting rather than reductive.[15] Depending on how they are sorted, ethnoarrays can also help facilitate either *explanatory* or *confirmatory* analysis. Examining cells within a column still facilitates interpretation of patient experiences or narratives, and sorting the ethnoarray can bolster interpretive insights through *exploratory* logics, for example, helping reveal or stimulate interpretations in the data that the analyst might have otherwise missed while reviewing or searching field notes and transcripts. That is, arrays can also be useful in examining whether a typology or pattern implied by a researcher or theory maps onto his or her data. Like the microarray on which it is based, the sorted ethnoarray would ideally allow analysts to identify patterns of similarity and difference in data and explore how these patterns resolve and translate into socially meaningful behaviors and theoretically meaningful categories and constructs.

Dating back to the popularization of exploratory data analysis (Tukey 1977), numerous quantitative techniques have been used to identify patterns in data that lack the strong sampling assumptions, claims to directionality, or assumptions about generalization that are typically associated with techniques like ordinary least squares regression. For instance, PCA, SA, latent structure analysis (LSA), multiple correspondence analysis (MCA), qualitative comparative analysis, and various applications of linguistic algorithms for mining large quantities of text data each provide useful ways of investigating patterns of colored cells that might be fruitfully integrated with the array approach. An in-depth discussion of the procedures involved in these techniques

can be found elsewhere. However, it is worth noting that PCA, SA, LSA, and MCA are the most directly comparable with the simpler model of clustering we use in Figure 5. PCA is commonly used in biological microarrays to define groups on the basis of nominal, ordinal, or interval gene expression data. PCA requires only a shared and directional scale. LSA and MCA allow categorical data to be clustered without assuming directional scale. These techniques are more common in the social sciences. SA groups like sequences and trajectories of longitudinal data. Because any of these approaches can be applied to ethnoarray data, researchers must decide which approach to clustering (if any) is most useful for their projects, as well as whether the use of interval approaches provides more worthwhile insights than the categorical approaches. Depending on how domains are measured and organized within the ethnoarray, statistical patterns revealed via these techniques could address a range of research questions such as ascertainment of temporal sequence, explication of causal mechanisms, and discovery of new grounded-theoretical constructs.

In our mock-up, we used a simple scale and sorting procedure based on interpretive color assignment to show how inductive techniques for pattern recognition might be useful even in interpretive analyses. In Figure 4, patients are arranged arbitrarily. In Figure 5, the ethnoarray is based on patients' structural characteristics (the bottom two domains, *Social Support* and *Insurance and Finance*) to examine whether and how those characteristics might shape pathways and tendencies related to seeking aggressive care. Each cell in these domains was assigned an ordinal value on the basis of color (red = 1; grey = .5; blue = 0). For each patient, an index value was calculated as the sum of the 12 cells in the two domains at times 1 and 2; the index had a potential range of 0 to 12; the actual range in the 10-patient array in Figure 4 was 1 to 11.5. The ethnoarray was then sorted according to these scores. Patients with higher index scores had their columns of data moved to the right side of the array; those with lower scores had their columns placed toward the left. Thus, reading Figure 5 from left to right roughly corresponds to examining experiences of patients with fewer to greater social structural resources.[16]

In the case of the PtDelib project, prior research had suggested that more advantaged patients were more likely to be seen as "good study patients" whom clinician-investigators targeted for clinical trials recruitment (Joseph and Dohan 2009), and the ethnoarray provides an

opportunity to examine this expectation in our preliminary PtDelib data. To examine the plausibility of this notion, we turned to the clustered array. The patients on the left side of Figure 5 tend to have less security in terms of *Insurance and Finance* and weaker *Social Support*. The distribution of the red cells in the *Decision* domain (at the top of the ethnoarray) may suggest that these patients are more aggressive in their pursuit of treatment and participation in clinical research. Given that we have arranged the domains in causal-temporal order, analysts and readers can then scan the array to try to identify patterns of color in the "intervening" domains—*Health and Illness* and *Communication*—that might suggest plausible social mechanisms. Analysts can then examine the underlying data (in this case interview transcripts and field notes) to see if these associations are likely real or spurious and to explain *how* the linking mechanisms operate in specific social contexts, a classic goal of ethnography.[17]

3.5 *Representing Data*

Ethnoarrays have the advantage of being able to summarize large amounts of data in a compact yet flexible form, a key feature of many forms of sociological analysis that has been underdeveloped in many ethnographic approaches. Figures 3 to 6 each summarize data from hundreds of pages of interview transcripts and participant-observation field notes from multiple sites. Just the data from Wayne Burley include dozens of pages of text—too much evidence to include in a journal article or even a monograph. As in the microarray, each color-coded cell reflects a rich storehouse of meaningful information. The color assignment both summarizes the data as an interpretable visual representation and enables new analytics, such as using clusters to identify new patterns in data or verify whether the typologies ethnographers create map onto the underlying data they represent. These visual summaries are meant to supplement, but do not replace, the narratives that form the standard for ethnographic representation.

Arrays also provide new possibilities for sharing information. Ideally, the data underlying cells could be bundled and shared along with the array, and interested readers could explore the underlying data for any array cell. We refer to this type of array as a "data-linked" ethnoarray, in contrast to the arrays shown in Figures 3 to 5, which we characterize as "flat" (i.e., a noninteractive summary representation). We do not yet

have the technology to produce and publish a data-linked array, though tabular and XML output functions of current CAQDAS platforms could facilitate this production, as we discuss and illustrate below. Inset 2 presents segments of underlying data from interviews with three participants that informed our coding of the *High Social Capital* measure in the ethnoarray. Figure 6 illustrates how these excerpts are situated within the ethnoarray. In a fully data-linked array, each cell of the ethnoarray would be associated with one or more fragments of ethnographic data, perhaps a quotation from an interview, a document, or an extract from a field note. Here, we use quotations from patient interviews to illustrate the kind of data that would underlie each cell in a data-linked array.

INSET 2: EXCERPTS OF UNDERLYING DATA LINKED TO THE SAMPLE ARRAY

Kathrine Fenimore (ID 7028): GRAY—coded as typical social capital: We have dear friends at church and just dear life friends that pray with me, and then that has a ripple effect. I mean, they ask others to pray and I've got just so many people, and then people at work, you know, too, that pray and that are wonderful friends as well. And just good support that way.

Tyrone Vorpahl (ID 4021): Change from BLUE (low for this study) at T1 to GRAY (typical for this study) at T2; Baseline/T1 (summer, 2011; BLUE): I live alone in an SRO and it's just a miserable environment. I want to relocate. I want to move someplace where I can live with family and friends. I have options right now . . . one of those options from the beginning were to go to [a city in this state] . . . and looking at—you know, my home is, where I'm from but [that State] is like, from my research, the most difficult state of all so I kind of ruled that out. And then the other option was [another state], 'cause I have close life-long friends there that I can go live with in a house outside of, you know, out of the city with like a bathroom and—I mean, just like I can live in a family home environment as opposed to in an SRO. So I'm trying to weigh—That's kind of a big part of my decision-making.

Follow-up/T2 (fall, 2011, GRAY): [Fieldworker note: since baseline interview, 4021 has moved to the third state mentioned above.] I have support in terms of like, you know, they're just there every day.

> You know, just somebody to say good morning and goodnight and have dinner with every night. And they're very concerned and doing everything, basically, you know, and dealing with my insanity. You know, I've been going through, like I said, kind of a roller coaster and they've been very tolerant of my sort of roller coaster emotions and my depression and anxiety I've been dealing with and sort of tolerant of that and kind of very accommodating in terms of opening up their home to me and letting me live here with them.
>
> ***Jamie Hoglund (ID 7044): RED — network with higher social capital:*** [My oncologist] asked me to go for chemotherapy and I asked my neighbor, who is a kidney transplant surgeon . . . he consulted with I think four or five other doctors. Half of them told me to go for the chemotherapy and the other half, including the liver transplant surgeon—not the one who actually operated on me but one of them who is on the team—they told me don't even bother to go for chemotherapy 'cause I should just live out the last few months in comfort instead of suffering.

Data-linked arrays are dynamic and interactive and thus a poor fit for paper. However, modern computer interfaces are well suited to publishing and sharing such arrays, and we are working on developing the tools and techniques that will allow the construction and electronic publication of data-linked ethnoarrays. Using such an application, readers could explore particular cells or groups of cells within the ethnoarray by reading through the underlying data. Such an interface would also allow readers to sort or reorder the ethnoarray's rows (domains) and columns (cases) using various procedures to highlight or discover patterns. Computer applications and tablet "apps" would ideally allow the reader to navigate a data-linked ethnoarray as one currently navigates online maps: clicking or tapping on cells to reveal the underlying data, zooming in and out of the ethnoarray to focus on particular patterns of data, dragging domains and cases to explore alternative patterns in the data. However, before implementing a data-linked array, important questions of how qualitative data might be adequately deidentified for sharing must be addressed, an issue we discuss in the next section. These discussions are consistent with calls for more transparent "open-source" social sciences (e.g., Freese 2007) and the shift toward digital models of publication that can facilitate new connections between scholarship and underlying data.

4. IMPLICATIONS

The ethnoarray's visual approach to presenting and analyzing data may provide new opportunities for work at the boundary of ethnography and other forms of social scientific scholarship. In constructing an ethnoarray, researchers can decide between and perhaps even balance various analytical approaches when they define conceptual domains and measures, assign colors to array cells, and bundle (or not) data with the array. Analysts can use arrays to scan large amounts of ethnographic data and to explore the data in new ways—explorations that may reveal new narratives, elucidate patterns in the data reflective of social mechanisms, add broader context to individual experiences and events, or suggest contingencies or limitations related to study data. If data are appropriately anonymized, they can be bundled with the array and shared so readers can examine the ethnographic evidence more directly and probe cell-to-data links. Providing access to data and analysis in this way helps readers see patterns, understand the analyst's interpretations, evaluate reliability, and gain a sense of an argument's scope and grounding. The flexibility of the ethnoarray—in presenting and analyzing data as well as providing readers with additional options for exploring data—provides the beginnings of an approach that can help make the vast troves of ethnographic data more available to diverse audiences without resorting to reductionism.

We now examine some implications of arrays for ethnography, including the ethnoarray's potential to spur and cultivate a novel research infrastructure, opportunities for new avenues in claims-making and evaluation, the potential scope of ethnography in large studies and its impact on the traditional solo ethnographer, as well as some key limits of this approach.

4.1. *Research Infrastructure to Support Ethnographic Arrays*

A robust ethnoarray research infrastructure would include (1) computer, Web, or tablet applications to facilitate the creation, distribution, and examination of arrays; (2) policies and procedures for anonymizing ethnographic data; and (3) servers to store and share data. Software to support ethnoarrays would differ from—though ideally integrate with and extend the capabilities of—presently available analysis programs. Current software helps experts manipulate data using technical

procedures. CAQDAS platforms help analysts code, sort, and explore data as well as tag or memo excerpts that will ultimately be presented to readers. Statistical analysis software allows researchers to fit models and produce tables or graphs. These packages all focus on manipulating data and producing output, and in most instances the output, not the underlying data, is all that is distributed to audiences. Array software would include similar tools to manipulate and analyze data (e.g., clustering and search functions) as well as provide a new form of output in "flat arrays." However, this software would also provide a mechanism for ethnographers to distribute findings. In short, ethnoarray software would help ethnographers not only *produce* output but also *be* part of their output and allow them to engage in communal research activities currently common to quantitative research such as archiving data and replicating analysis.[18]

As a data management tool, ethnoarray software would help analysts link data to array cells and to arrange and rearrange the cells to explore alternate definitions of domains or configurations of cases. Links between data and cells occur when assigning cell color (we describe multiple strategies for color assignment in Inset 1). Ideally, applications would remain agnostic about the process of color assignment to allow analysts flexibility in how they link cells to data. This would also allow analysts the freedom to arrange the data set on their own terms, albeit in a way that aids in making their work more transparent. Some analysts might rely on interpretation alone, while others might develop an automated formal process for coding, sorting, and linking data to cells. No matter how the data-cell link is created, however, array applications should help analysts rearrange data to examine patterns or explore new relationships.

The ethnoarray also allows the representation of ethnographic data in two key ways. First, analysts can publish arrays online or in printed articles or monographs. Used this way ethnoarrays, like any other visual representation of data such as graphs or charts, allow researchers a way to summarize a large volume of information. For some researchers, such a summary might represent a key finding of an ethnographic study. Other ethnographers might use the arrays color-coded tabular representation to supplement interpretative analyses of field notes, interview transcripts, and other data that are presented using more traditional narrative approaches. The second way is to share an entire array, including cells and linked data, with readers. Readers then have the ability to

examine the array, to iteratively explore and arrange the cells, and to examine the links between cells and data.

Such dissemination strategies differ substantially from the dominant ethnographic practice of publishing monographs and research articles with solely narrative evidence, but it is not unprecedented. The Human Relations Area Files (HRAF), a nonprofit international consortium, aimed to provide a resource of ethnographic data focused on comparative societal analysis, and full text from early ethnographies exists online. However, to provide comparable data across societies, HRAF used rigid coding and analytical constructs, and the archive has been interpreted as a historical document illustrating the challenges—and perhaps folly—of a cumulative approach to knowledge production in cultural analyses (Clifford and Marcus 1986; Marcus 1998). The ethnoarray model shares HRAF's interest in making data available to a wide community of scholars, but scholars who use the ethnoarray need not format or categorize their data according to rigid preexisting conceptual schemata. They need not even agree about epistemic assumptions underlying ethnographic scholarship. They need only to specify what they do. In this sense only, data-linked ethnoarrays are more akin to publicly available quantitative data sets, such as the Integrated Public Use Microdata Series (IPUMS) or the General Social Survey, than the HRAF.

Researchers could produce array data sets as a part of their scholarly activities, but they need not adhere to a unified epistemic logic in doing so. They could then provide data sets to the sociological community with full documentation of how they were produced but without placing rigid boundaries on how the data are intellectually deployed. Similarly, data-linked ethnoarrays would not follow the rigid proscriptions of HRAF standardization but would instead exist as a series of independent data sets. Access could be provided via Web portals such as those seen at the Interuniversity Consortium for Political and Social Research (ICPSR). Researchers would ultimately have to decide if and how to reuse data sets and whether they were comparable with other data sets.[19]

This raises the question of how to handle data governance. Sharing IPUMS- or ICPSR-housed data relies on policies for depositing, storing, and distributing data that ensure the safety and rights of research participants. Sharing arrayed ethnographic data would require producing new policies or extending current policies. Although data warehouses that host qualitative data are beginning to emerge, such policies are in their nascent stages, and the lack of a shared format for summary

representation and sharing of ethnographic data remains a major limitation. Providing a comprehensive solution is beyond the scope of this article, but the development of ethnoarray approaches could potentially advance work in this arena. Policies for protecting microlevel quantitative social science data or protected health information may offer some further guidance for ethnographic policies.

Ethnographers' own habits and practices regarding treatment of informants and other data, which have generally been passed along as craft rather than codified in policy, would need to be made more explicit. Sharing ethnographic data via arrays would also require a physical computing infrastructure, which could be provided via Internet-connected servers. Finally, even if the secure research infrastructure developed to accommodate arrays were never used to share data-linked ethnoarrays, it might prove to be a valuable resource for ethnographers to store, analyze, and reanalyze their own data, especially as the ethnoarray provides an analytical approach for linking data from multiple studies and points in time. In other words, although the array approach does not provide a universal solution to the challenges of sharing ethnographic data, it does provide a tool that can advance discussions about if and how this aim might eventually be reached.[20]

4.2. *Using Arrays to Support Ethnographic Claims-Making*

Evaluating ethnographic claims can be circuitous. Often, ethnographers collect and analyze data by themselves, and they can share only a fraction of their data with readers. Readers rely on self-reports of how field notes, interviews, and other data were collected; how these data were analyzed; and how insights were obtained and conclusions drawn. Ethnographers have long recognized that their authority derives in part from these reports of fieldwork and readers' trust in those reports (Rabinow 1997; Whyte 1993). Marked by interpretation and iteration, ethnographic data often gain legitimacy when the insights they produce appear plausible and comprehensible—when, in essence, the data take on the appearance of speaking for themselves. Thus, the quality of their presentation—including richly evoked empirical context and well-developed theoretical framing—helps establish the legitimacy of the data that produce those results.[21]

For many, a description of research procedures is a necessary first step, but an inadequate proxy, for a more direct examination of the links

between data and claims. The limitations of this proxy become apparent when ethnographers debate whether the data support the claims made and even whether the data were collected. Such debates can flounder on irreconcilable divergences about the contextual or historical specificity of evidence and argument (Boelen 1992; Duneier 2002; Orlandella 1992; Sánchez-Jankowski 2002; Wacquant 2002). Sometimes a lack of standardization is associated with a lack of rigor. The combination of the requirement of trust without access to data to reconstruct analyses and the often charged nature of ethnography's research topics can lead to scholarly exchanges that generate as much heat as light. Given that the ethnographer is the data collection instrument, it is not entirely surprising that controversies over the validity of ethnographic claims can devolve into attacks on analysts' legitimacy or even morality (cf. Duneier 2002, 2006; Katz 2010; Wacquant 2002). Explicitly revealing how analysts link data and claims and encouraging readers to assess how the former sustains the latter could provide a more productive dialogue. We believe the ethnoarray represents a potentially useful tool to support ethnographic claims-making by facilitating such examinations. In Appendix B, we provide an example of how an array might be applied to examine the claims made in a well-known comparative ethnography.

Ethnoarrays can facilitate a more detailed examination of claims-making and, ideally, generate explicit discussion of how ethnographic data are invoked for causal or narrative purposes. This does not put the research community on an inexorable road toward an ethnographic equivalent of a $p < .05$ threshold for theoretical or substantive significance or even the conceptual standardization of HRAF, nor do ethnoarrays necessarily privilege causally or hypothesis-oriented research. Rather, we hope new tools can provide a way to examine how and why interpretations overlap or differ. Such discussions may provoke new ways of exploring fertile ethnographic questions.

4.3. *The Scope of Ethnography*

The ethnoarray may provide new capacities for analysis, but these capacities may come at the price of new burdens on those who choose to use the approach. A historical characteristic of ethnographic practice has been minimal barriers to entry; the lone ethnographer requires little more than time, a notebook, and access to enter the field and potentially contribute to the literature. In contrast, developing and contributing to

ethnoarrays introduces new burdens for data collection and analysis. Consider the potential new burdens of an ethnoarray approach for ethnographers who conduct participant observation. When lone ethnographers collect notes in the field, they rely on their own judgment to decide what to observe and document.[22] Although many approaches encourage specifying a unit of analysis and identifying conceptual domains, this is not universally the case. Field notes may include everything from contextualized individual behaviors, to reflexive musings, to researcher descriptions of physical space (Emerson et al. 1995). In the midst of fieldwork, researchers decide what types of data to record and how to record it, but they rarely have the time, energy, or foresight to completely document how these decisions were made. Key background information in the form of schemata and headnotes may still remain unarticulated (Sánchez-Jankowski 2002). Even arrays thoughtfully constructed to include research questions, domains, and units of analysis may be incomplete when it comes to crucial details of how and why particular data were collected or recorded. Teams of ethnographic researchers may strive for more consistency in their procedures for conducting and documenting field sites, but the team's shared understandings of the site and the project may not be formally documented. In short, ethnographers currently write field notes for themselves or for small audiences of fellow fieldworkers. They consider broader audiences when designing a study, when deciding what data to collect, or when writing up results. But the ethnographers themselves are the usual audience for most study data that remains largely private.

In contrast, data bound for arrays must continually consider a broader audience. The broader audience may be unknown, but generally it does not know the field site. Data included in an ethnoarray must be clear to a naïve audience lacking the presumed *Verstehen* of the ethnographer. They must have a defined unit of analysis. Formal field notes may thus require greater attention to detail and context, be longer, and take more time to write. They may contain greater redundancies than field notes that are destined for more traditional ethnographic uses, and ethnographers may feel self-conscious about array-bound notes. In short, ethnographers producing arrays may collect data differently than ethnographers producing traditional monographs or articles.

Using arrays also requires analytical transformations after the field data have been collected. Developing and distributing a flat array means using computer software, while using a data-linked array requires a

series of steps to anonymize and secure data. These steps have the potential to make ethnography more expensive and less nimble, and it seems certain that anonymizing data for use in data-linked arrays will lead to the development of new research tasks, infrastructure, and personnel that could change aspects of ethnographic production for those using the array method.[23]

5. CONCLUSIONS

In an influential article, Jeremy Freese (2007) described "the need to move beyond intermittent discussions of replication to standards of collective action" (p. 220) as a key step toward ushering in a more transparent sociology. The fact that Freese and colleagues are even able to engage in a coherent conversation around these issues points to a luxury that ethnographers do not necessarily possess—basic shared assumptions about the nature, language, and goals of the research enterprise. Although most quantitative researchers typically share concerns with replicability, reliability, generalization, inference, and validity, ethnographers differ remarkably in how they relate to these concepts and traditional social science frameworks more broadly. Consequently, those interested in issues of openness, process, and visualization must first confront not only the daunting methodological challenges this entails but also the lack of consensus and persistence of qualitative "tribalism" in the scholarly field (Lamont and Swidler 2014). Tensions among ethnographers with different epistemic approaches are intellectually legitimate, but ideally these tensions should not preclude attempts to address shared practical issues. Despite their differences, ethnographers from various social science traditions must each grapple with the complex logistical challenges of analyzing and presenting context-rich observations of meaningful human action. Most would like to speak to larger audiences, and some would even like a civil means of talking to one another. For ethnographers, developing tools for these ends is an important precursor to enhancing transparency.

In this article, we introduced the ethnoarray, an interactive visual approach for analyzing, representing, and sharing ethnographic data that we argue is consistent with enhanced transparency. We argued that the ethnoarray approach provides tools for addressing common challenges that face many sociological ethnographers as they seek to manage and analyze the rich, context-dependent data gathered through fieldwork.

We then discussed a number of technical considerations in developing an ethnoarray—how an analyst might define domains and measures, assign colors to cells, sort or reorder an array, interpret patterns within or across columns, link data to arrays, and so on—and how these techniques can potentially open new possibilities for sociological ethnography, particularly when used in conjunction with traditional narrative methods of presenting data. We provided a model to illustrate how an ethnoarray might be constructed.

It is important to reiterate a key limitation here: this tool requires further refinement. Our mock-up is small and noninteractive because of the need to outline its premises. It is also constrained by the limitations of the print medium. A functioning ethnoarray would include both finer levels of detail so patterns would be more striking and instructive as well as interactive links that would allow analysts to drill down to the data to which the patterns refer. It is also clear in the discussion of the mock-up that while the ethnoarray may provide a useful tool for managing or analyzing data, it cannot "solve" the more fundamental epistemic divides separating different types of ethnographic practice. Nor do we try to use it for these ends. Rather, we hope it will serve as a bridge that allows conversation across at least some subdisciplinary chasms.

Even as we bear these limitations in mind, we note that the fundamental goal of the ethnoarray reflects a core tenet of many ethnographic approaches: providing a way to bring readers close to the social phenomenon in question so they can appreciate its context, complexity, and contingency. "When assessing evidence," Tufte (1997) noted, "it is helpful to see the full data matrix, all observations for all variables, those private numbers from which the public displays are constructed. No telling what will turn up" (p. 45). Showing the "full data matrix" from an ethnographic study is likely impossible, but the principle that more data are preferable to fewer nevertheless applies. The ethnoarray provides a way of showing readers more information from the field. It allows them to discover and explore patterns in that information, adding context and breadth to specific observations. In this way, it is a tool that may help analysts and readers turn up new insights and one that may help them make sense of the richness and complexity of the social world using visual tools that are essential in other methods (Moody and Healy 2014), but which ethnographers have been slower to adopt. A fully developed ethnoarray may even help researchers and readers share ethnographic data sets to allow deeper

engagement and understanding. To paraphrase Geertz (2000), the eth-noarray approach can potentially provide scholars and readers with an enhanced ability to converse, even if the end result is only the ability to vex one another with greater precision. In this capacity, it may provide another tool in the pantheon of pluralistic techniques for social inquiry that enables communication and provides ethnographers with a platform to address shared challenges and may perhaps in the process even begin to challenge the growing tribalism found in qualitative methodology (Lamont and Swidler 2014).

APPENDIX A

How an Ethnoarray Can Employ Units of Analysis Other than the Individual

The sample arrays in the main text focus on examining differences and similarities in individual trajectories using both interviews and field observations. Columns in these examples represent individuals (see again Figure 3) or individuals by time (Figures 4–6). However, as we noted earlier, this need not be the case. The array was designed to work with any form of data that assumes a basic structure (e.g., the structure has a unit of analysis and analytical domains or measures). This appendix provides a brief illustration using data from a recent comparative ethnographic work.

In *The End Game: How Inequality Shapes Our Final Years*, Abramson (2015) examined how persistent socioeconomic, racial, and gender divides in the United States create an unequal "end game" that structures the later years of the aging U.S. population. In doing so, the work provides a lens for examining both the social stratification of later life and the lifelong consequences of inequality in the United States. The book is based primarily on two and a half years of comparative ethnographic research conducted in four urban neighborhoods and examines, among other issues, how disparate social contexts and resources shape the way people from different neighborhoods and backgrounds can respond to the shared challenges of "old age." One finding is that the ability of older people to navigate the physical spaces of the real world (sidewalks, buildings, etc.) is an important form of inequality that stratifies how different groups can manage everyday life. This is conditioned in part by disparities in seniors' health (which reflect the stratified timing and severity of physical challenges). However, it is also powerfully shaped by an unequal distribution of services and contextual material resources available to networks of people in
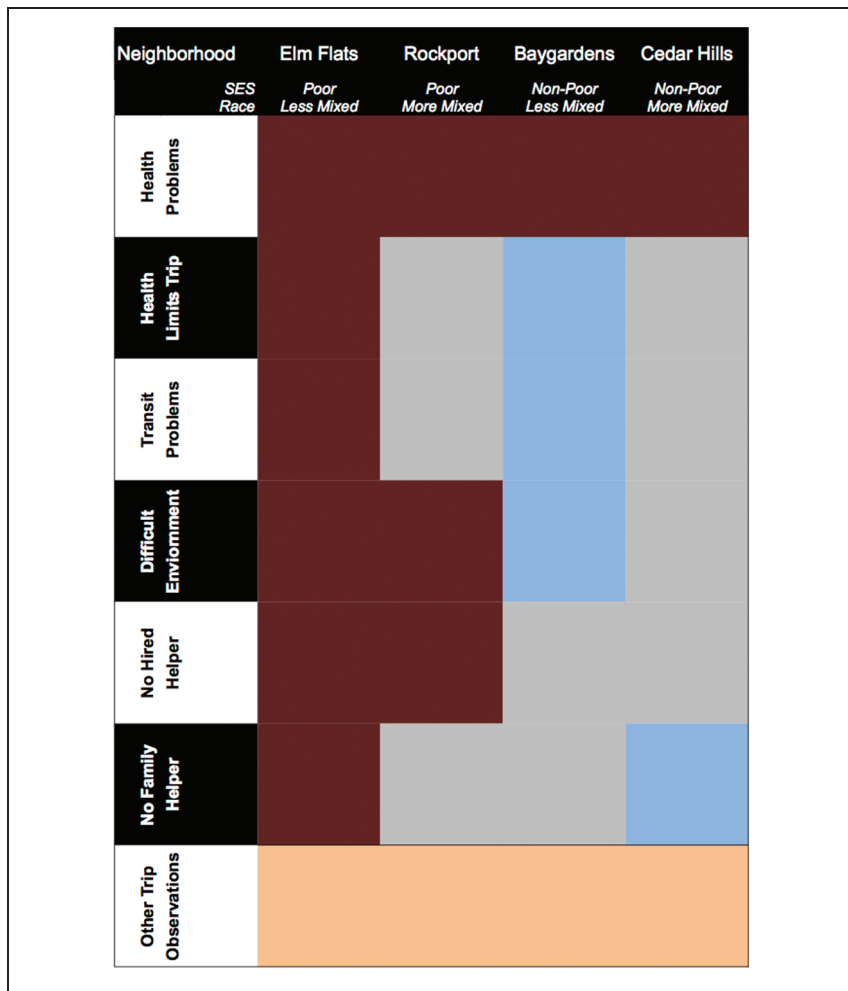
Figure A1. An ethnoarray showing factors that affect trips to medical clinics. Data from *The End Game: How Inequality Shapes Our Final Years* (Abramson 2015).

different neighborhoods—a finding generated using traditional field observations of neighborhood settings such as housing complexes.

An array approach could use this participant-observation data to support the claim of neighborhood-level contextual variation by examining how neighborhood context and other factors affect seniors' ability to navigate the physical environment and make visits to a medical clinic. Although this outcome affects

individuals, the data could be aggregated to correspond to observed events nested in the four neighborhoods. In such a case, columns would represent neighborhoods rather than individuals. Paired with traditional observations of individuals' struggles and accounts of getting to a doctor's office, the array could help illustrate the similarities and differences within and across different communities. An array adapted to this task, following the color coding conventions introduced earlier in the PtDelib examples, might take the form seen in Figure A1. Furthermore, to illustrate where data related to the phenomenon in question but not neatly categorized by domains might be located, this version includes a row for "other trip observations."

APPENDIX B

How an Array Might Complement a Well-Known Comparative Ethnography

This appendix will illustrate how an ethnoarray might be used to present ethnographic findings in a way that complements current practices, and it will explain how the array might be used to explicate the claim making process in Annette Lareau's (2011) *Unequal Childhoods*, a well-known comparative ethnographic study. Lareau's study of class, race, and social reproduction among school-age children includes detailed accounts of how middle-class, working-class, and poor families interact with schools and other institutions. In a detailed appendix, Lareau noted her decision to analyze her data in different ways (sometimes with software assistance, sometimes via the more traditional reading and rereading of notes) and the iterative process by which she sought disconfirming evidence as her arguments began to take shape. Yet it is striking that this description occupies only a small fraction of her appendix, which is largely devoted to describing the logistics and dilemmas of data collection. She also noted that to make her findings more accessible, she presented her data narratively—focusing on one family at a time—rather than adopting a more analytical discursive approach.

This study presents an instance in which ethnoarrays might help achieve the author's ethnographic goals and explicate how conclusions are drawn. Ethnoarrays could maintain the work's accessibility to a wide audience, but at the same time, the use of array visualizations might provide a more explicit analytical schema that would be of interest to specialized sociologists and ethnographers examining claims. Given the multilayered comparisons in the study, a variety of different ways of developing and presenting ethnoarrays

from the data suggest themselves. For example, domains (or rows) of an eth-noarray could be constructed from middle-class study participant Garrett Tallinger's (chapter 3) life: weekly events and activities, conversations with parents and siblings, exhaustion and elation brought on by events in his life, the absence of relatives, attention (or lack thereof) to money, his parents' Ivy League background, their job-related travel schedules, and the extent to which his schedule determines the routines of his siblings. If Garrett were the unit of analysis (column), such an array would paint a picture of how cultivation plays out in the routine activities of a middle-class family. Adding other children (and their activity domains) to the ethnoarray would provide insights into how patterns of cultivation versus natural growth play in the lives of the children Lareau studied, thus providing a more highly explicated analytical approach to her data that could supplement the book's chapter-based case studies.

Ethnoarrays could also enrich the analysis of *Unequal Childhoods* even without engaging Lareau's theoretical arguments about social reproduction. For example, a purely narrative approach might produce an ethnoarray that extends Lareau's Table 2, which currently conveys a sense of Garrett's busy cultivated life just by listing all the sporting and performing events on his calendar. If an individual day in the life of one child is the unit of analysis (column) and his activities during that day constitute the domains (rows), the resultant array would produce a rich, descriptive sense of which children are engaged in which activities when and how intensely. This ethnoarray need not carry the freight of an explanatory logic; it would simply provide a window into how the children Lareau studied spent their days and allow inductive exploration—through grounded or interpretative strategies—of the nature of childhood in contemporary America.

As this exegesis of *Unequal Childhoods* suggests, an ethnoarray can help summarize large amounts of data. As part of a traditional ethnographic article or monograph, an ethnoarray could include a broader sense of the study data and provide some context for analysts' specific claims that may help readers assess internal plausibility. However, data-linked arrays can go even farther and provide readers with the opportunity to examine directly the data that an analyst purports to justify a claim.

participants at the AJS/University of Chicago Conference on Causal Thinking and Ethnographic Research for comments on previous versions of this manuscript as well as to Susan Miller and Matthew Wenger for assistance with initial data analysis. We gratefully acknowledge participants in the Cancer Patient Deliberation Study for their willingness to share their experiences with us.

Funding

Notes

1. Emerson, Fretz, and Shaw (1995) provided a broad definition of the ethnographic craft, which they described as a research method that "involves the study of groups and people as they go about their everyday lives" (p. 1). In the case of sociological research, this normally involves participant observation with individuals, groups, or organizations, conducted over time. More broadly, ethnographers use a collection of methods to understand individuals and the contexts in which they act, often (although not necessarily) spending time with people as they go about their lives, engaging with those individuals, and keeping some sort of notes or diary, then writing findings on the basis of their observations (Becker 1958; Gans 1999).
2. We use the term *construct* because, as scholars have long pointed out, no piece of data simply "speaks for itself" (e.g., Berger 1963; Blumer 1969; Cicourel 1964).
3. Others have also turned to the natural sciences in developing social scientific analytic approaches (e.g., Abbott and Hrycak 1990). Recently, those engaged in automated analyses of text have turned to heatmap-centered approaches as a means of discovery and representation (cf. Mohr and Bogdanov 2013; Mohr et al. 2013).
4. It is important to recognize that the introduction of new techniques and modes of visualization can create new issues and debates as well as advances. One need only look to the discussions around geographic information systems in geography and related fields (e.g., see Pickles 1997). Although it is impossible to anticipate all such issues with the ethnoarray, we endeavor to discuss them throughout this article.
5. For a frequently updated review of the various CAQDAS progams on the market, see http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/.
6. There are, however, some promising basic tools for examining and representing co-occurrence and heatmaps of codes in off-the-shelf programs that might facilitate the array approach with some adaptation.
7. The figure includes 337 columns, each containing data from a single specimen (320 tumorous and 17 normal tissues) and approximately 1,900 rows, each representing a specific gene. The color of each cell in the table reflects the expression profile of one gene in one specimen. Genes that are overexpressed (i.e., that are

more active in metabolic processes) are shown in red; green indicates underexpression. In both cases, color intensity corresponds to the strength of overexpression or underexpression, and typical levels of gene expression are shown as black cells.

8. An alternative analysis using only field notes might examine differences and similarities in interactions across clinics. In that case, the ethnoarray's columns would each represent a clinic. The domains (or rows) of the array could also be redefined.

9. It is important to note a limitation of our mock-up. The 16-row demonstrative ethnoarray creates a relatively crude checkerboard that might mistakenly give the impression that this tool is simply a visual device for quantifying and decontextualizing rich qualitative data. To be useful for capturing the richness of ethnography, however, the ethnoarray would include many more domains and many more measures and submeasures within domains. Rather than constituting a crude checkerboard, a fine-grained visualization would reveal patterns of relationships among domains and facilitate interpretation of the rich data underlying the array. Furthermore, as we discuss in sections below, the cells of the array would be linked back to the underlying notes and transcripts.

10. Unlike critics who argue for the superiority of a given approach (e.g., Biernacki 2014), we believe these approaches can exist pluralistically.

11. The 2012 volume of *Sociological Methodology*, especially Franzosi et al.'s (2012) discussion of quantitative narrative analysis (QNA), explores some of these issues. In addition to introducing QNA, a tool for quantifying the structural properties of narrative, the volume includes useful discussions of the potentials and pitfalls of using computers to analyze qualitative data more generally (Junker 2012; White, Judd, and Poliandri 2012a, 2012b). Although the array differs from QNA in its focus on textual meaning rather than invariant structure, the broader discussions regarding the new avenues of qualitative inquiry and ways of representing and sharing data opened by computers is a useful analog to what we present here; see also Gorski (2004).

12. A pseudonym, as are all proper names in this paper, obtained from a random name generator.

13. Comprehensive array software would facilitate this process but is not yet available, a topic we address in section 4. To generate the representations above, we used data from the master data set for PtDelib, which is currently maintained in a secure data environment using CAQDA software. The field observations, interviews, and survey data were entered into the program, coded using an iterative process, and linked to "case summaries" that gave an overview of individual experiences, chronologies, and outcomes. These case summaries were sortable by various demographic and social characteristics and "hyperlinked" back to the observations on which they were based. The interpretive color coding is based on a reading of these summaries and their underlying data.

14. For a useful related discussion of how the co-occurrence tools of ATLAS.ti can be used to help identify context, see Contreras (2011).

15. A modified array approach might facilitate such a quantitative reduction, but this would involve a fundamentally different form of inquiry.

16. Analysts can decide how many and which domains or measures to use when sorting the array. However, in this article we deliberately provide this relatively

simplistic example to explicate the basic mechanics involved, without presupposing reader knowledge of advanced statistical techniques; see, for instance, the 2012 edition (volume 42) of *Sociological Methodology*.

17. The mock-up in Figure 5 is too basic to draw conclusions on the basis of an interpretation of patterns in the array, but we did arrange the domains to facilitate an explanatory-oriented interpretation of how particular factors, such as social capital, influence whether people enter into early phase trials to show how one might approach this analysis.

18. As our use of CAQDA software in the PtDelib study evidences, the array approach is not antithetical to the use of CAQDA software. ATLAS.ti and MAXQDA both have analytical tools for examining code associations in heatmap-like formats. Ideally array software could operate with, or independently of, commercial CAQA software packages to facilitate ethnoarray production and analysis. The Coding Analysis Toolkit (http://cat.ucsur.pitt.edu) provides a useful exemplar in its Web-based CAQDA suite, which can work either independently or import data from ATLAS.ti (Lu and Shulman 2008).

19. The ICPSR includes references to some ethnographic studies, including Hodson's (2004) data set that allows a comparative analysis of workplace ethnographies. These data sets often include survey data with an ethnographic component, or a representation of ethnographic monographs (as in Hodson's case), but the underlying ethnographic data are not generally available publicly, nor is there a standard mechanism for accessing these data. Recently, there have been moves to improve the repositories of qualitative data (cf. http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter3qual.html).

20. This is a matter of great importance and likely contention. Although a full treatment is beyond the scope of this article, we hope the techniques and suggestions we outline in this text might further this discussion.

21. This intellectual process is not confined to ethnography, of course, but it is one of the fundamental social processes of scientific discovery and the production of authority in any knowledge-based field (Latour and Woolgar 1986; Timmermans and Berg 2003). In addition, applied ethnographers may be urged to establish bona fides by invoking theory or highlighting their use of CAQDA software, a different kind of procedural claims-making (Reeves et al. 2008).

22. Of course, a similar selection process is at play to a large degree in any research method, even when measures are more formalized (Cicourel 1964; see also Witzel and Mey 2004).

23. Once again, although we see this approach as a complement to rather than a replacement of traditional approaches, it is worth noting that arrays may also have unintended consequences for the traditional practice of ethnography. Many sociologists begin their ethnographic research careers as graduate students, entering and exploring a field site under the tutelage of an experienced faculty mentor and, in more than a few cases, producing studies with marked impact on their fields or the discipline (Anderson 1978; Bosk 1979; Burawoy 1979; Small 2004). A hallmark of this relationship is that the mentor need not provide the student with much in the way of material resources. This lack of resource dependence may allow forms of creativity that students training in other methodological

traditions—such as the graduate student researcher whose dissertation analyzes a faculty mentor's grant-funded survey—may experience less frequently. Although the traditional training model appears consistent with producing a flat ethnoarray, students who wish to participate in a community of ethnographic scholars who share their work via data-linked arrays may require further professional support to properly anonymize and contribute data to a central data resource. Obviously, this dynamic would apply to ethnographers at any stage of their career, but the implications on practices and structures of ethnographic training, which has long been the most common entry point to the ethnographic guild, may be worthy of careful consideration.

References

Abbott, Andrew, and Alexandra Hrycak. 1990. "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96(1):144–85.

Abramson, Corey M. 2015. *The End Game: How Inequality Shapes Our Final Years*. Cambridge, MA: Harvard University Press.

Anderson, Elijah. 1978. *A Place on the Corner*. Chicago: University of Chicago Press.

Becker, Howard. 1958. "Problems of Inference and Proof in Participant Observation." *American Sociological Review* 23(6):652–60.

Becker, Howard S. 2009. "How to Find Out How to Do Qualitative Research." *International Journal of Communication* 9:545–53.

Belacel, Nabil, Qian (Christa) Wang, and Miroslava Cuperlovic-Culf. 2006. "Clustering Methods for Microarray Gene Expression Data." *OMICS: A Journal of Integrative Biology* 10(4):507–32.

Berger, Peter L. 1963. *Invitation to Sociology: A Humanistic Perspective*. New York: Anchor.

Biernacki, Richard. 2014. "Humanist Interpretation versus Coding Text Samples." *Qualitative Sociology* 37(2):173–88.

Blumer, Herbert. 1969. *Symbolic Interactionism: Perspective and Method*. Englewood Cliffs, NJ: Prentice Hall.

Boelen, W. A. Marianne. 1992. "Street Corner Society: Cornerville Revisited." *Journal of Contemporary Ethnography* 21(1):11–51.

Bosk, Charles. 1979. *Forgive and Remember: Managing Medical Failure*. Chicago: University of Chicago Press.

Brady, Henry E., and David Collier. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowan & Littlefield.

Burawoy, Michael. 1979. *Manufacturing Consent: Changes in the Labor Process under Monopoly Capitalism*. Chicago: University of Chicago Press.

Burawoy, Michael. 1998. "The Extended Case Method." *Sociological Theory* 16(1): 4–33.

Burawoy, Michael. 2004. "Public Sociologies: Contradictions, Dilemmas, and Possibilities." *Social Forces* 82(4):1603–18.

Cicourel, Aaron V. 1964. *Method and Measurement in Sociology*. Glencoe, IL: Free Press.

Cicourel, Aaron V. 1968. *The Social Organization of Juvenile Justice*. New York: John Wiley.

Clifford, James, and George E. Marcus. 1986. *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley: University of California Press.

Contreras, Ricardo B. 2011. "Examining the Context in Qualitative Analysis: The Role of the Co-Occurrence Tool in ATLAS.ti." *ATLAS.ti Newsletter*, August, pp. 5–6.

Dohan, Daniel. (2003). *The Price of Poverty: Money, Work, and Culture in the Mexican-American Barrios*. Berkeley: University of California Press.

Dohan, Daniel, and Martín Sánchez-Jankowski. 1998. "Using Computers to Analyze Ethnographic Field Data: Theoretical and Practical Considerations." *Annual Review of Sociology* 24:477–98.

Duneier, Mitchell. 2002. "What Kind of Combat Sport Is Sociology?" *American Journal of Sociology* 107(6):1551–76.

Duneier, Mitchell. 2006. "Ethnography, the Ecological Fallacy, and the 1995 Chicago Heat Wave." *American Sociological Review* 71(4):679–88.

Duneier, Mitchell. 2011. "How Not to Lie with Ethnography." Pp. 1–11 in *Sociological Methodology*, vol. 41, edited by Tim Futing Liao. Hoboken, NJ: Wiley-Blackwell.

Durkheim, Emile. [1893] 1984. *The Division of Labor in Society*. New York: Free Press.

Eisen, Michael B., Paul T. Spellman, Patrick O. Brown, and David Botstein. 1998. "Cluster Analysis and Display of Genome-Wide Expression Patterns." *Proceedings of the National Academy of Sciences* 95(25):14863–68.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 1995. *Writing Ethnographic Field Notes*. Chicago: University of Chicago Press.

Franzosi, Roberto, Gianluca De Fazio, and Stefania Vicari. 2012. "Ways of Measuring Agency: An Application of Quantitative Narrative Analysis to Lynchings in Georgia (1875–1930)." Pp. 1–42 in *Sociological Methodology*, vol. 42, edited by Tim Futing Liao. Thousand Oaks, CA: Sage.

Freese, Jeremy. 2007. "Overcoming Objections to Open-source Social Science." *Sociological Methods and Research* 36(2):220–26.

Gans, Herbert J. 1999. "Participant Observation in the Era of 'Ethnography.'" *Journal of Contemporary Ethnography* 28(5):540–48.

Geertz, Clifford. 2000. "Thick Description: Towards an Interpretive Theory of Culture." Pp. 3–30 in *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine Transaction.

Goldthorpe, John H. 2000. *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford, UK: Oxford University Press.

Gorski, Philip S. 2004. "The Poverty of Deductivism: A Constructive Realist Model of Sociological Explanation." Pp. 1–33 in *Sociological Methodology*, vol. 34, edited by Ross M. Stolzenberg. Boston: Blackwell.

Henderson, Stuart, and E. Segal. 2013 "Visualizing Qualitative Data in Evaluation Research." *New Directions for Evaluation* 139:53–71.

Hodson, Randy. 2004. "A Meta-ethnography of Employee Attitudes and Behaviors." *Journal of Contemporary Ethnography* 33(1):4–38.

Joseph, Galen, and Daniel Dohan. 2009. "Diversity of Participants in Clinical Trials in an Academic Medical Center: The Role of the 'Good Study Patient'?" *Cancer* 115(3):608–15.

Junker, Andrew. 2012. "Optimism and Caution Regarding New Tools for Analyzing Qualitative Data." Pp. 85–87 in *Sociological Methodology*, vol. 42, edited by Tim Futing Liao. Thousand Oaks, CA: Sage.

Katz, Jack. 1997. "Ethnography's Warrants." *Sociological Methods and Research* 25(4): 391.

Katz, Jack. 2010. "Review of Cracks in the Pavement." *American Journal of Sociology* 115(6):1950–52.

King, Gary, Robert O. Keohane, and Sidney Verba. 2001. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Klinenberg, Eric. 2006. "Blaming the Victims: Hearsay, Labeling, and the Hazards of Quick-hit Disaster Ethnography." *American Sociological Review* 71(4):689–98.

Lamont, Michèle, and Ann Swidler. 2014. "Methodological Pluralism and the Possibilities and Limits of Interviewing." *Qualitative Sociology* 37(2):153–71.

Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life*. Berkeley: University of California Press.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Leahey, Erin. 2008. "Overseeing Research Practice: The Case of Data Editing." *Science, Technology, and Human Values* 33(5):605–30.

Lofland, John. 1995. "Analytic Ethnography: Features, Failings and Futures." *Journal of Contemporary Ethnography* 24(1):30.

Lu, C.-J., and Shulman, S. W. 2008. "Rigor and Flexibility in Computer-based Qualitative Research: Introducing the Coding Analysis Toolkit." *International Journal of Multiple Research Approaches* 2(1):105–17.

Lutfey, Karen, and Jeremy Freese. 2005. "Toward Some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in the Routine Clinic Visit for Diabetes." *American Journal of Sociology* 110(5):1326–72.

Marcus, George E. 1998. *Ethnography through Thick and Thin*. Princeton, NJ: Princeton University Press.

Miles, Matthew B., and A. Michael Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd ed. Thousand Oaks, CA: Sage.

Mills, C. Wright. 1959. *The Sociological Imagination*. New York: Oxford University Press.

Mohr, John W., and Petko Bogdanov. 2013. "Introduction—Topic Models: What They Are and Why They Matter." *Poetics* 41(6):545–69.

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.

Moody, James W., and Kieran Healy. 2014. "Data Visualization in Sociology." *Annual Review of Sociology* 40:105–28.

Orlandella, Angelo Ralph. 1992. "Boelen May Know Holland, Boelen May Know Barzini, but Boelen Doesn't Know Diddle about the North End!'" *Journal of Contemporary Ethnography* 21(1):69–79.

Perez-Hernandez, Danya. 2014. "New Repository Offers a Home for Data That Aren't Numbers." *Chronicle of Higher Education*. Retrieved January 23, 2015 (http://chronicle.com/blogs/wiredcampus/new-repository-offers-a-home-for-data-that-arent-numbers/50865).

Pickles, J. 1997. "Arguments, Debates, and Dialogues: The GIS–Social Theory Debate and the Concern for Alternatives." Pp. 49–60 in *Geographical Information Systems: Principles, Techniques, Management, and Applications*, edited by D.R.P Longley, M. Goodchild, and D. Maguire. New York: John Wiley.

Prat, Aleix, and Charles M. Perou. 2011. "Deconstructing the Molecular Portraits of Breast Cancer." *Molecular Oncology* 5(1):5–23.

Rabinow, Paul. 1997. *Reflections on Fieldwork in Morocco*. Berkeley: University of California Press.

Reed, Isaac. 2010. "Epistemology Contextualized: Social-scientific Knowledge in a Postpositivist Era." *Sociological Theory* 28(1):20–39

Reeves, S., M. Albert, A. Kuper, and B. D. Hodges. 2008. "Why Use Theories in Qualitative Research?" *BMJ* 337:631–34.

Sallaz, Jeff. 2009. *The Labor of Luck: Casino Capitalism in the United States and South Africa*. Berkeley: University of California Press.

Sánchez-Jankowski, Martín. 2002. "Representation, Responsibility and Reliability in Participant Observation." Pp. 144–60 in *Qualitative Research in Action*, edited by Tim May. London: Sage Ltd.

Sánchez-Jankowski, Martín. 2008. *Cracks in the Pavement: Social Change and Resilience in Poor Neighborhoods*. Berkeley: University of California Press.

Sanjek, Roger, ed. 1990. *Field Notes: The Makings of Anthropology*. Ithaca, NY: Cornell University Press.

Schena, Mark, Dari Shalon, Ronald W. Davis, and Patrick O. Brown. 1995. "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray." *Science* 270(5235):467–70.

Small, Mario Luis. 2004. Villa Victoria: The Transformation of Social Capital in a Boston Barrio. Chicago: University of Chicago Press.

Small, Mario Luis. 2009. "'How Many Cases Do I Need?': On Science and the Logic of Case Selection in Field-based Research." *Ethnography* 10(1):5–38.

Stears, Robin L., Todd Martinskey, and Mark Schena. 2003. "Trends in Microarray Analysis." *Nature Medicine* 9(1):140–45.

Strauss, Anselm, and Juliet Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. 2nd ed. Thousand Oaks, CA: Sage.

Swedberg Richard, ed. 2014. *Theorizing in Social Science: The Context of Discovery*. Stanford, CA: Stanford University Press.

Tangherlini, Timothy R., and Peter Leonard. 2013. "Trawling in the Sea of the Great Unread: Sub-corpus Topic Modeling and Humanities Research." *Poetics* 41(6):725–49.

Tavory, Iddo, and Stefan Timmermans. 2009. "Two Cases of Ethnography: Case, Narrative and Theory in Grounded Theory and the Extended Case Method." *Ethnography* 10(3):243–63.

Timmermans, Stefan, and Marc Berg. 2003. *The Gold Standard: The Challenge of Evidence-based Medicine and Standardization in Health Care*. Philadelphia: Temple University Press.

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. 2nd ed. Cheshire, CT: Graphics Press.

Tufte, Edward. R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press.

Tukey, John. 1977. *Exploratory Data Analysis*. New York: Addison-Wesley.

Wacquant, Loïc. 2002. "Scrutinizing the Street: Poverty, Morality, and the Pitfalls of Urban Ethnography." *American Journal of Sociology* 107(6):1468–1532.

White, M. J., M. D. Judd, and S. Poliandri. 2012a. "Brightening the Bulb: Response to Comments." Pp. 94–99 in *Sociological Methodology*, vol. 42, edited by Tim Futing Liao. Thousand Oaks, CA: Sage.

White, M. J., M. D. Judd, and S. Poliandri. 2012b. "Illumination with a Dim Bulb? What Do Social Scientists Learn by Employing Qualitative Data Analysis Software in the Service of Multimethod Designs?" Pp. 43–76 in *Sociological Methodology*, vol. 42, edited by Tim Futing Liao. Thousand Oaks, CA: Sage.

Whyte, Willliam F. 1993. *Street Corner Society: The Social Structure of an Italian Slum*. 4th ed. Chicago: University of Chicago Press.

Witzel, Andreas,, and Günter Mey. 2004. "'I Am NOT Opposed to Quantification or Formalization or Modeling, but Do Not Want to Pursue Quantitative Methods That Are Not Commensurate with the Research Phenomena Addressed': Aaron Cicourel in Conversation with Andreas Witzel and Günter Mey." *Forum: Qualitative Social Research* 5(3):Art 41. Retrieved January 23, 2015 (http://www.qualitative-research .net/index.php/fqs/article/view/549/1186).

Author Biographies

Corey M. Abramson is an assistant professor of sociology at the University of Arizona and a research associate at the Center for Ethnographic Research at the University of California, Berkeley. Abramson's research uses both qualitative and quantitative methods to examine how persistent social inequalities structure everyday life and how they are reproduced over time. His recent book on this topic, *The End Game: How Inequality Shapes Our Final Years* (Harvard University Press, 2015), provides a comparative ethnographic analysis of how various facets of inequality profoundly shape life for older Americans and examines what this tells us about the mechanisms of social stratification more broadly. Abramson's current methodological work focuses on developing ways to improve rigor and transparency in the collection, analysis, and representation of qualitative data.

Daniel Dohan is a professor of health policy and social medicine at the Philip R. Lee Institute for Health Policy Studies at the University of California, San Francisco (UCSF). His research focuses on the culture of medicine: how it ameliorates and perpetuates societal inequalities, its relationship to science and discovery, and how training creates health professionals. In addition to his research, he is leading the development of a new master's degree in health policy and law, to be jointly offered by UCSF and the University of California Hastings College of Law.