Title: The Living Codebook: Documenting the Process of Qualitative Data Analysis

Victoria Reyes, Elizabeth Bogumil and Levin Elias Welch
University of California, Riverside

Acknowledgements: We'd like to thank the editor and the anonymous reviewer(s) for their helpful feedback. We'd also like to thank the undergraduate research assistants involved in this project. This work has been supported by the American Sociological Association's Fund for the Advancement of the Discipline, an American Fellowship (Postdoctoral Research Fellowship) from the American Association of University Women, Hellman Fellows Fund, and a Blum Initiative on Global & Regional Poverty Faculty Research Seed Grant.

Citation:

Reyes, Victoria, Elizabeth Bogumil, and Levin Elias Welch. *Forthcoming*. "The Living Codebook: Documenting the Process of Qualitative Data Analysis," *Sociological Methods & Research*

Abstract

Transparency is once again a central issue of debate across types of qualitative research. Work on how to conduct qualitative data analysis, on the other hand, walks us through the step-by-step process on how to code and understand the data we've collected. Although there are a few exceptions, less focus is on transparency regarding decision-making processes in the course of research. In this paper, we argue that scholars should create a *living codebook*, which is a set of tools that documents the data analysis process. It has four parts: 1) a processual database which keeps track of initial codes and a final database for completed codes, 2) a "definitions and key terms" list for conversations about codes, 3) memo-writing, and 4) a difference list explaining the rationale behind unmatched codes. It allows researchers to interrogate taken-for-granted assumptions about *what* data is focused on, *why*, and *how* to analyze it. To that end, the living codebook moves beyond discussions around inter-coder reliability by documenting the process by which analytic codes are created, refined, and debated.

Introduction

Transparency is once again a central issue of debate across types of qualitative research. Ethnographers focus whether to name people, places, or to share data (Reyes 2018a, Jerolmack and Murphy 2017, Guenther 2009, Contreras 2019) and whether our data actually matches the claims we make (e.g., Jerolmack and Khan 2014). Work on how to conduct qualitative data analysis, on the other hand, walks us through the step-by-step process on how to code and understand the data we've collected (Charmaz 2006, Yin 2011, Hsieh and Shannon 2005, Altheide 1987). For example, Deterding and Waters (2018) advocate for the use of a flexible coding schema, laying out a practical method on how to use qualitative data analysis software to analyze large number of interviews in a systematic and transparent manner. However, Biernacki (2012) argues that coding procedures are often hidden and not transparent. Indeed, a common assumption in coder training is that individuals conduct their research in isolation or, if in teams, submit their codes to a PI who then calculates an intercoder reliability score to see how similar two or more research assistants coded data. Although there are a few exceptions for analyzing interview data (Weston, Gandell, J Beauchamp, McAlpine, Wiseman, and C Beauchamp 2001, Deterding and Waters 2018), less focus is on transparency regarding *decision-making processes* in the course of research.

We shift the transparency debate from ethnography and interviews to how transparency operates in the content analysis, or coding, of documents and argue that scholars should create a *living codebook* from which to analyze their data. The living codebook is a set of tools that makes the analysis of documents more transparent among team members and, if researchers decide to make it public, to the scholarly community. It has four parts: 1) a processual database which keeps track of initial codes and a final database that contains completed codes, 2) a

“definitions and key terms” document that keeps track of conversations about codes, including their definitions and situational applicability, and a clean version for reference, 3) memos written during the coding process, and 4) a difference list where team members explain the reasoning behind assigning particular codes that were not assigned by other team members.

Together, this set of tools allow qualitative researchers, particularly those working in teams, to document the precise reasons regarding how and why we analyzed data and came to our conclusions. It allows researchers to interrogate taken-for-granted assumptions about what data they focus on, why, and how to analyze it. In doing so, the living codebook moves beyond discussions around the importance of inter-coder reliability by tracking the process by which analytic codes are created, refined, and debated. That is, the living codebook is one way to document *how* definitions are arrived, *what* the bounds of concepts are, and the *backstage decisions* of the qualitative data analysis process. It also reveals how many codes are situationally dependent and highlights the need for people in the coding process, as words in and of themselves need to be read in context to understand how they relate to the concepts and codes of interest.

The living codebook has implications for other types of qualitative data analysis, such as the coding of interview data. Coding interview data similarly organizes text – transcripts – into patterns to uncover common themes. However, we also recognize differences in coding interviews versus the coding of documents. For example, interviews are conducted by researchers who have particular hypotheses that shape which questions are asked, whereas document data exists prior to the research process. Nevertheless, we suggest that the living codebook can be fruitfully used to analyze interview data because many types of qualitative

researchers, regardless of whether they use ethnography, interviews or documents as data, grapple with how to define and draw the bounds of codes and concepts.

In the following sections, we first situate the paper in research on transparency in qualitative research and discuss coding practices. Next, we describe the living codebook. Then, we outline its four main parts: the processual and final databases, key terms and definitions list, analytic memos, and the difference list. We discuss their uses through examples from our current research project on reputation and empire. Finally, we end by discussing how the living codebook builds on and extends current discussions of transparency and data analysis in qualitative research.

Transparency in qualitative research

Current discussions around transparency in qualitative research typically focus on ethnography: whether and how to name places, name people, or share data (e.g., Reyes 2018). Jerolmack and Murphy (2017), for example, have argued that as a default, researchers should name the exact places and people they study. They should do so in order to combat what Jerolmack and Murphy (2017) call “pseudo-generalizability” –which end up “reifying sites as ideal types and insinuating that they are representative of other unobserved cases” (3). This is problematic, they argue, because doing so does not accurately root researchers’ data and arguments in the “messiness” of the places and people we study and gestures to a representativeness that does not exist (Jerolmack and Murphy 2017, see also Small 2009). In contrast, other scholars have argued that there should be no default in transparency, because whether, how, and when to be transparent differs on a project-by-project basis and that scholars should focus on transparency in the research process regarding which decisions they made and

why (Reyes 2018a, see also Guenther 2009). Transparency can also be partial and involve semibiographical, partial spatial, and invitational disclosures because naming people and places can present real danger when scholars research crime and criminal activities (Contreras 2019).

Yet, these discussions have generally centered on ethnographic transparency in *data collection*—that is in the raw data we collect and the specific and actual places and people we study. In contrast, there has been less discussion in this debate around transparency in *data analysis*—or how we interpret and understand the qualitative data we collect (for this distinction see Reyes 2019b). There are a few exceptions. First, is Corey Abramson’s work. Abramson and his colleagues have pioneered the field of computational ethnography, and they have called for a new tool to analyze qualitative data, what they call the ethnoarray, which “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” (Abramson and Dohan 2015: 275, see also Abramson et al 2018). This tool allows for transparency and sharing of data and its analysis. Second, is work on how to “do” qualitative analysis, though as we’ll discuss below, decisions around coding practices remain hidden from the scholarly community.

We argue that more research needs to take seriously transparency in qualitative data analysis, and the living codebook is one such way to do so. To be clear, when we talk about transparency, our focus in this paper is on transparency in coding decisions. We are not necessarily advocating sharing the living codebook with the scholarly community, as these decisions around what, why, how and when to be transparent must be made on a project-by-project basis (Reyes 2018a). Instead, we focus on transparency in the research process and within teams. That is, the living codebook is a way to reveal and codify the underbelly of how research

is conducted and be transparent in the decisions we make regarding how and why we created particular codes as well as their definitional boundaries and situational applicability.

Conducting qualitative data analysis

If most of the current scholarly discussion around transparency in qualitative research is focused on data collection, how are people taught to conduct qualitative analyses? Training, or teaching the “nuts and bolts” of how to conduct qualitative research and analyze qualitative data, often occurs in the classroom, and throughout the course of research itself. Graduate students, for example, often are required to take one or two qualitative methods classes that might offer a mix of classical and contemporary readings alongside out-of-classroom activities where you learn methods through “doing.” Some university-affiliated centers also offer short-term qualitative workshops organized around more advanced training in a specific method. For example, Indiana University’s Social Science Research Commons features a Workshop in Methods, UC Berkeley’s Center for Ethnographic Research provides ethnographic training via a Summer Student Workshop and Intensive Ethnography Workshop, among others, while the Comparative Methods for Systemic Cross-Case Analysis (COMPASS) holds Qualitative Comparative Analysis Workshops.

Qualitative data consists of a wide array of materials such as field notes, interviews, documents, and photos, among others. Once data are collected, and often as it is collected, scholars must analyze their materials. While discussion on archival and historical/comparative methods tends to focus on case selection, process tracing, event analysis, path dependence and/or a processual approach (e.g., Abbott 2016, Mahoney 2012, Mahoney 2004, Beck 2017, Beck 2018) and how to grapple with issues of representation, access and availability, given that

archives are incomplete (e.g., Fuentes 2016), a primary mode of analyzing qualitative interviews and documents is content analysis or *coding*.¹ Coding involves systematically categorizing, organizing and analyzing texts to uncover patterns. It can be either quantitative – focused on counts – or qualitative. Qualitative coding is reflexive, focused on meaning and context of materials, and may be further delineated by three different approaches: conventional (inductive), directed (deductive) or summation (latent) content analysis (e.g., Altheide 1987, Hsieh and Shannon 2005).

We follow Yin (2011), who describes coding as an iterative process that involves compiling, disassembling, reassembling (and arraying), and interpreting data. Compiling data entails creating a kind of database to order it. It is iterative in that there is a back-and-forth process of cleaning and verifying data. This ensures researchers are consistent when they order the data and that they become deeply familiar with it. Disassembling data involves creating codes, or assigning labels, to the data. There can be *in vivo* codes, those that use the exact words from the text, or codes that scholars assign. Disassembling data entails memoing, or writing down your ideas as you go. The next step, according to Yin (2011) is reassembling and arraying the data, which is reorganizing codes, or thinking of higher-level codes to sort the data. Finally, interpreting the data involves creating a new narrative from the data and this coding process. At each state of the process, coding involves posing questions, thinking of rival explanations and thinking through comparisons (see also Altheide 1987).

Similarly, Kathy Charmaz (2006) updates classic grounded theory (Glaser and Strauss 1967) by providing guidelines on how to “do” grounded theory at each step in the process: from interviewing and ethnography to coding, memo-writing, sampling and sorting, to theory building and writing papers from memos. For the purposes of this paper, here we focus on and follow her

suggestions on coding and memo-writing. For Charmaz (2006), “coding consists of at least two main phases: 1) an initial phase involving naming each word, line, or segment of data followed by 2) a focused, selective phase that uses the most significant or frequent initial codes to sort, synthesize, integrate, and organize large amounts of data” (46). Coding also entails memo-writing, which “is the pivotal intermediate step between data collection and writing drafts of papers. When you write memos, you stop and analyze your ideas about the codes in any-and every-way that occurs to you during the moment” (Charmaz 2006: 72). To that end, memo-writing is the substantive heart of qualitative data analysis.

However, knowing one way to code, for example by examining data line-by-line, linking particular sentences to specific codes, and the importance of writing analytic memos (e.g., Charmaz 2006), leaves large gaps in knowledge of the research process: how it is actually done in practice and the decisions researchers make regarding how codes are defined and what is included or excluded in the definition. Additionally, as Deterding and Waters (2018) point out, the increasing number of interviews for qualitative studies almost necessarily entails multiple people working on a project. This, combined with the sheer number of documents used in historical/comparative methods and those available in archives and other repositories, suggests the need for more understanding of how to actually breakdown the analysis of a given text. It also suggests the need to be more transparent about those steps, both for team researchers and scholars working on their own, something that is largely lacking in the literature today.

Quantitative methods have long dealt with issues surrounding sharing data, sharing formulas and codes that researchers use to analyze said data, and how to manage multiple people and teams working on the same datasets in a number of ways. One of the most institutionalized ways is through the development of a codebook. A codebook, in these methods, are guidelines

and explanations of how the data was collected, the questions asked, and how to analyze the data. For example, the General Social Survey (GSS), a popular dataset used among social scientists, has one such codebook (see: <https://gss.norc.umd.edu/get-documentation>). The current GSS codebook includes, among others, information on what the GSS is, how it was developed, who funded it, and which questions appear consistently, which do not, and why. It also details changes in its design from 1972 to the present, and why these changes occurred and how it impacts analysis, as well as the category names (labeled as “Mnemonic”) and their description (Mnemonic description), which is necessary to interpret the categories and codes within the dataset. It includes an index to the dataset, each GSS module ever in existence and 22 appendices, totaling 3,799 pages that provide a literal “how-to” guide to understanding the GSS.

Do qualitative researchers need to emulate this model? No. But the *idea* of a codebook - something that keeps track of decisions and identifies definitions of concepts - can be useful in the course of qualitative analysis, particularly as it relates to team research.

We are not the first to think of a qualitative codebook nor of how to conduct qualitative team research in a manner that is transparent, scientific, and useful guides for other scholars. For example, Weston, Gandell, J. Beauchamp, McAlpine, Wiseman, and C Beauchamp (2001) discuss their process in developing a coding system and codebook in their qualitative interview research team. They outline how Weston developed the initial coding scheme, the graduate student team members coded the materials, and that the coding was an iterative process, where Weston would draft and revise the codes based on team feedback in the first six months of meeting, though the codes continued to evolve throughout the process. They developed a codebook because they realized they “needed to document our decisions and as a tool for communication among team members. Initial definitions and rules were developed and relevant

examples from transcripts were documented in the codebook to help with consistency of coding” (387). Throughout their process they addressed issues of reliability and validity among the codes and coders to make sure they could train future coders and did continual code checks to ensure robustness. Their process was as follows: they applied “the newly developed codes to ten percent of the transcripts from each of the six professors. This was followed by informal group discussion where we compared our coding. Where agreement was fairly high [...] we often concurred that codes were acceptable. If agreement was low, then data were reexamined and codes refined” (395). Their codebook had two parts. Weston et al (2001) write that “The first section describes a four-step procedure for coding: how to identify the unit of analysis (an episode), how to identify the type of reflection, how to code monitor statements, and how to code control statements. The second section is a complete list of all the codes, each accompanied by a definition and sample words, phrases, or linguistic markers drawn from the interview transcripts” (395). They conclude that coding is an important part of qualitative analysis, team coding builds shared interpretation of the materials, and team research adds important rigor to qualitative research.

Though they don’t call it a codebook, Deterding and Waters (2018) similarly offer a how-to guide on analyzing interviews in team projects in an effort to increase the scientific rigor of qualitative research and transparency in qualitative data analysis. They tackle grounded theory and argue that while many scholars say they use grounded theory to analyze their interview data, in practice, they do not actually follow its tenets. Drawing on their own experience and an analysis of flagship journals in the sociology discipline, they lay out a practical method, which centers on using qualitative data analysis software, to *flexibly* analyze large numbers of interviews. Deterding and Waters (2018) argue that the first step in qualitative analysis should be

indexing the transcripts by the questions asked during the interview and the concepts the scholars were interested in, as a way to get familiarized with the data.

Next, similar to Charmez (2006), researchers should write analytic memos to understand the themes in their data. They build on Charmez's (2006) work and suggest that researchers should write two types of memos, what they call respondent-level memos (or memos per participant) and cross-case memos (which detail emerging themes across interviews). Instead of grounded theory, and similar modes of analysis, which goes from line-by-line coding to broader themes, they suggest the opposite - first broadly indexing and categorizing the data and then go to more detailed analyses, which allows researchers to equitably distribute the material and be more focused on the particular papers and concepts on which they want to write.

Finally, they detail the importance of using qualitative software in data analysis, and outline specific steps to do so, including how to name files, how to transcribe - and what kinds of questions the person transcribing the interviews should ask themselves, including tone of voice, among other characteristics - and assigning attributes to the data in an effort to be able to compare transcripts and people across particular attributes (or categories). These steps allow for increased reliability and validity of their data and analysis, making sure "unique" or "extraordinary" events are not overly relied on and scholars can pinpoint more specifically emerging themes and provide concrete counts of how often these themes emerge. They also argue that these steps are not limited to interview transcripts, but can also include material like fieldnotes, which can similarly be uploaded to qualitative software and analyzed.

These two works present important advancements in qualitative coding and we build on them in a number of ways in our call for a "living codebook." Weston et al (2001) helpfully discuss how one team member initially defined codes and these codes are refined through

feedback from the team, and their codebook usefully outlines how to identify the unit of analysis, how to code statements and provides a complete list of codes, including definitions and examples. We build on Weston et al (2001) by providing logistical and practical suggestions on how to keep track of analytic and methodological decisions, tracking the conversations team members have about the bounds of codes and their definitions, and adding multiple parts to the codebook.

Similarly, Deterding and Waters (2018) call for increased scientific validity, reliability, and transparency in the data analysis process through the use of their flexible coding guidelines and qualitative software. We also focus on increased transparency and do so by tracking the various steps and decisions made in the course of research. To track these decisions, we suggest the use of two additional tools: the processual database and difference list, both of which are discussed below. Finally, we put all of the moving parts of qualitative analysis – databases, key words and definitions lists, memos, and the difference list – into the methodological tool that we call the “living codebook.”

Our call for a living codebook complements Deterding and Waters’ (2018) flexible data analysis and use of qualitative data software, with each approach committed to strengthening transparency in data analysis and rigor of qualitative data and research. Our focus on coding documents presents some differences. Their approach is based on semi-structured interviews, which are conducted by researchers who have created interview guides that are shaped by their hypotheses, while the nature of document data is that the data exists prior to the research process. Although researchers analyzing documents also have hypotheses, shape data collection by their choice of documents and are shaped by availability and accessibility of data (e.g., Fuentes 2016), they do not use interview guides to help facilitate conversations with participants. This may

mean that the coding of documents is more open, “messy” and relatively more inductive than interview data. Yet, we believe that the living codebook provides useful tools that can be adapted for the analysis of interviews, and other types of qualitative data, and suggest that the living codebook and a flexible data analysis plan using software can be combined and adapted to researchers’ needs.

As we’ll discuss in the following section, the living codebook is meant to be a useful and flexible tool that lays out one concrete, logistical way to conduct rigorous coding that occurs across types of qualitative research by documenting each step in the analysis process. This also allows for more clarity, precision, and relatively quick turnaround regarding training new members who join a research team. It also facilitates ease when researchers decide whether to share the data, and to what extent, with the scholarly community.

We think it is important to note that our call for a “living codebook” should not be read as a reification of quantitative methods as a more “pure” science from which qualitative scholars need to draw on in order to justify qualitative methods. Nor are we saying every qualitative study, whether based on ethnography, interviews, or documents, necessarily needs a living codebook. Instead, qualitative methods, as many have pointed out before, do indeed differ from quantitative methods in a number of important ways and those ways are just as scientific, useful, and rigorous as research using quantitative methods (e.g., Small 2009, Katz 1997). The living codebook offers one way to document how definitions are arrived, the bounds of concepts, and the backstage decisions involved in the qualitative data analysis process. However, just like issues of transparency in ethnography (Reyes 2018a), whether or not and how to use the living codebook are decisions that need to be made on a case-by-case basis depending on the data and goals of the research.

Data

The creation of a living codebook stems from our team project on empire and cultural wealth, or the relationship between reputation, place, and socio-economic activity. This project is based off of thousands of documents collected by the PI from 2012-2013 about Subic Bay, Philippines. Our university's Institutional Review Board determined the project (HS 17-269) does not require Human Subjects Approval. Data was systematically collected from a range of sources. We compare the *content* of what's being said, and the accompanying reputations that are created and institutionalized, with *who or what audience* is speaking and the *form* the discussions take. This includes:

- Reviews from Trip Advisor found using the search terms “Subic Bay Freeport Zone” “Subic Bay” and “Olongapo.” According to Tuttle (2009) TripAdvisor is one of the top 5 most useful travel websites for Budget Travel CNN editors and it has a self-reported 570M reviews and opinions, 455M monthly average unique visitors, and 7M accommodations, restaurants, and attractions.² It also provides reviews for lodging, restaurants, things to do, and has a travel forum. While it is a U.S.-owned company, it's used internationally.
- Posts from Reddit, more specifically four sub-reddits found from a search of “Subic Bay” and “Olongapo”: r/FilipinaBarGirls, r/Navy, r/Philippines, r/USMC [US Marine Corps]. And Reddit is an American social news aggregation, web content rating, and discussion website. It is an anonymous webforum where users can discuss controversial places and topics without fear of being judged in their off-line life.

According to World Economic Forum and data from Alexa.com, Reddit is one of the top ten websites around the world (Gray 2017).

- Government websites from Australia, Japan, South Korea, Philippines, U.S.

because these countries are among the most connected to Subic Bay and these sites provide governmental information that directs people where to go and what to expect when traveling to other countries and are the major migrant sending and receiving countries to and from the Philippines.

- U.S. military-related websites that are top hits on a Google search, including: Navyvets.com, subicbaypi.com, and Retired Activities Office Subic Bay. Given Subic Bay's extensive history as a former U.S. military base, these sites provide a wide array of semi-official and unofficial communication with current U.S. military personnel and veterans.

- Official Philippine tourism websites from the national Philippine government, Olongapo City government, and Subic Bay Freeport Zone government as these provide the government's attempts to cultivate a particular reputation.

- Fodders, Lonely Planet, Frommers, and Michelin Travel guides are some of the most popular travel guides in the world that influence travelers on where to go and what to expect; although U.S.- and Western European-based, these guides are used worldwide.

- Newspapers, including the Philippine Daily Inquirer and the New York Times, which provide local and U.S. coverage of Subic Bay that makes headline news, and allows us to see messages distributed to everyday people in the Philippines and U.S. They frame our understanding of the world around us (e.g., Lester 1980; Gamson and Wolfsfeld 1993), though we also know that coverage and what is considered

“newsworthy” are socially constructed (Myers and Caniglia 2004; Oliver and Myers 1999; Lester 1980).

Each set of data provides a different vantage point in which to study the reputation of Subic Bay. By reputation, we refer to how Subic Bay, and any accompanying people and specific locations within it, are discussed, analyzed, and portrayed across these media sources. Each source has a different author - ranging from government officials and journalists to everyday people using social media to express their opinion. They also each have a different target audience - educated consumers, a general “public” or fellow social media users. Taken together, these sources allow us to compare how Subic Bay is portrayed and discussed across different mediums, whether and to what extent Subic Bay’s reputation and cultural wealth is similar across sources, and whether its reputation is differently racialized and gendered across authors and audiences. For the purposes of this paper, we draw on examples from TripAdvisor to show the process of the living codebook.

The Living Codebook

We use the term “living codebook” to encompass a set of methodological tools that can help research teams understand and analyze qualitative data. In contrast to quantitative code books that provides researchers with the final documentation on how to use and interpret a given dataset or qualitative code books that detail a priori definitions of concepts that research assistants use to code the material given to them by a researcher, the living codebook builds the process of revision, discussion, and transparency into data analysis. The living codebook is a tool with four moving parts: (1) processual and final databases, (2) definitions and key terms, (3)

memos and (4) difference lists. Using the living codebook requires continual meetings among team members, and training for new team members as they are brought onboard projects.

The living codebook is centered around three main ideas. First, that science, particularly social science, is a trial and error process that is iterative in practice, where researchers comb back and forth through data and theories (see Timmermans and Tavory 2012) while simultaneously seeking alternative explanations to avoid confirmation bias. Second, that research needs to involve reflexivity, or thinking about how our own social positions shape our data analysis. While most qualitative work on positionality focuses on ethnography (see Contreras 2013, Duck 2015, Flores 2016, Reyes 2018b), using each of the tools within the living codebook is also an exercise in reflexivity. Writing down explanations regarding why we chose particular terms over others and what stood out in the coding process in our memos not only allows us to immerse ourselves in the data but also reveals how not all researchers read the data the same way. Researchers' positionality influences how they interpret data and their logic in doing so. Third, when using these tools in a team environment, it elicits active participation by all members, as Weston et al (2001) noted in their description of their project's codebook. It demystifies the research process for undergraduate and graduate student members, showing the logistics of how data is analyzed and interpreted. Our team consists of the PI, two graduate students and teams of 2-3 undergraduate students. To be sure, the PI leads the research team. However, by emphasizing that each team member has something intellectual to contribute, the project itself is enriched as student researchers provide insights into codes and their definitions that the PI might not have otherwise thought of.

In the rest of this paper, we discuss the four moving parts of the living codebook more in-depth: processual and final databases, definitions and key terms, memos, and difference lists.

Although individual parts of the living codebook have been discussed by previous scholars (e.g., Charmaz 2006, Deterding and Waters 2018, Weston et al 2001), we document the role each plays in the living codebook. To that end, we introduce a novel way to maintain a “definitions and key words” list, and two new methods to document the data analysis process: the processual database and the difference list. Finally, we outline how each part works, questions to consider as scholars use the various parts of the living codebook, and challenges we’ve encountered.

Processual and final databases

The first step in using a living codebook is to organize data by creating database(s). As previously described, scholars have documented and provided insights and guidelines into how to construct a database, including the necessity of writing memos and/or the importance of assigning attributes (e.g., Weston et al 2001 Deterding and Waters 2018) and we agree with many of their guidelines. We also recognize the need for *multiple* databases for the same set of data and project. For example, there is what we are calling the *final* database. This is the database that Deterding and Waters (2018), among others, describe. It has the final codes and data identifiers that researchers use to sort the data after it has been cleaned. We also construct a *processual* database that documents the process of assigning final codes. Let us walk through an example that demonstrates the differences between these databases and the need to introduce a processual database to the qualitative analysis process.

One set of data we are working with are TripAdvisor reviews. The raw data are in PDF format. The first step in creating a database for these was deciding the unit of analysis. For the hotel reviews the unit of analysis was an individual review, and there are 1547 reviews of 18 hotels in the area. The next step is deciding on the information we wanted to document from

each review, what Deterding and Waters (2018) call the data attributes. For each review, we documented the keyword used to search and find that particular review, as the PI used three different search terms “Olongapo” “Subic Bay” “Subic Bay Freeport Zone” to find the data. We also documented the name of the establishment, the type (lodging), subtype (hotel, B&B, specialty), whether the search term matches the physical location of the establishment (since this is not always the case), the reviewer’s name/online handle, reviewer’s contribution level (this level is based on how many reviews the reviewer has written on TripAdvisor.com), date of review, date of stay, and title of review. We also maintained a column that contains the entire review itself for ease of coding, so the research assistants can refer to this column rather than the original documents. We agree with Biernaki’s critique of stripping data from context, and this is one way to include the context in the database. Organizing the information in this way is standard.

Our third step was creating a set of three columns per research assistant: key word review, key word thematic (or analytic code), and overall rating (see figure 1). With the TripAdvisor data, we are working with undergraduates with little to no qualitative coding experience. To that end, we created the “key word review” column, which is meant to be a useful step between the review itself and the analytic codes assigned. The “key word review” column consists of words - or sometimes phrases - that are excerpts taken directly from the review, for example, “cheap,” “simple joys,” “turtle nests at the beach,” “tucked away,” or “hate.” The next column, “key word thematic,” are the analytic codes themselves. Similar to the “key word review” category, this category consists of multiple codes in one column and each of these codes is located in our definitions and key word document. For example, some of our analytic codes include “hotel quality,” “nature,” “gender,” “illicit activities” “cleanliness,” “nationality” and

“food”. Next, we have a column on overall rating of the review. That is, whether the review itself is “mixed” “positive” “negative” or “neutral.” To code the overall rating, we looked at the title of the review, looked at the overall tone of the review, and counted the number of positive, negative, or neutral sentences within the review.

Figure 1: Format of TripAdvisor processual database that includes research assistants and their three types of codes

Research assistant A				Research assistant B		
Categories	Keyword review	Keyword thematic	Overall rating	Keyword review	Keyword thematic	Overall rating
Review						

In a standard database - what we’re calling a final database - each analytic code is its own characteristic or attribute. For this processual database - in which the undergraduate research assistants are working on the TripAdvisor data - we have all the codes in a single column to simplify the process for them. For other databases, this step may not be needed.

Next, there are another set of five columns, labeled: “Same codes,” “Different codes,” “Deleted codes,” “Added codes,” and “Final codes” (see Figure 2). These columns are the substantive heart of the processual database. The researcher working with the students would go through each individual review, comparing the codes listed in “key word thematic” category for each paired group. They input the codes that appeared in both “thematic key word” columns in the “Same codes” category and those that appear in one student’s thematic key word column but not the other into the “Different codes” column. These different codes are listed in both the

“Different codes” column and the difference list, which is discussed below. The “Deleted codes” and “Added codes” columns are a way to document which codes the team, as a whole, decided to delete or add during their weekly meeting. The “Final codes” category lists all agreed upon codes, both those that were added and those that were initially agreed upon. This column, alongside the identifying information of each review, is listed in the final database. Each analytic code within the “final codes” column is also listed as separate attributes in the final database, mimicking the database that Deterding and Waters (2018) advocate.

Figure 2: Format from TripAdvisor processual database

Categories	Same codes	Different codes	Deleted codes	Added codes	Final codes
Review					

Note: The researcher compares the codes listed in the “keyword thematic” category of each student in a group. Those that appear in one student’s column but not the other, are listed in the “Different codes” category, while those that appear in both are in the “Same codes” category.

The processual database is important because it keeps track and documents the decisions made about each code, rather than listing the finalized codes that come at the end of the process (see table 1). It is also flexible. Although we describe the processual database for the TripAdvisor data coded by undergraduates in the sections above, it can be adapted. For example, when two coauthors coded photographs from the Retired Activities Office (RAO), each analytic code was its own column, or attribute, from the beginning, mimicking the conventional database. We did not have a “key word review” category nor a “review” category, though we did have a category where we described the contents of each picture. What makes it a processual database is that we recorded our codes in the same database - for example we had a column titled “Coauthor

A clothing”, a column titled “Coauthor B clothing” - and kept track of the differences in our codes. We did so through a “final codes” column that listed all agreed upon codes (e.g., “final clothing”), those that we initially agreed upon and those we agreed upon after our meetings. We also kept track of differences in our codes within the difference list, which is discussed below.

Table 1: Processual and final databases of TripAdvisor data

Processual database	Final database
<ul style="list-style-type: none"> - Key characteristics of each review - Columns by research assistant - Column that document similarities and differences in codes between research assistants - Column that documents which codes are added - Column that documents which codes were deleted - Column of final codes 	<ul style="list-style-type: none"> - Key characteristics of each review - Column of “final codes” - Separate column for each analytic code

Definitions and Key Terms List

The second part of the living codebook is the definitions and key terms list. Defining codes is one, if not *the*, central function of any codebook. It is a way to help researchers understand and analyze the data, that is to identify patterns and relationships among the data. As previously mentioned, Weston et al (2001) included in their codebook outline how to identify the unit of analysis, how to code statements and provides a complete list of codes, including definitions and examples. We build upon their work to use the definitions and key terms list to document the decision-making process. To that end, in addition to outlining the definition of a term and giving appropriate examples, we also document the questions raised and conversations about what contexts and situations a given code can be applied and when it is excluded. It also includes which codes were initially used but subsequently deleted in favor of a more useful code.

Let us walk through an example using the TripAdvisor data once again. During the first quarter this project was conducted at our current institution, the coauthors each worked with a team of 2-3 undergraduates who would code the data. Each team would meet weekly and the entire team would meet once a month. Because this marked the beginning of our creation of codes, questions often came up in the course of these meetings and we documented them in this list (see figure 3).

Figure 3: Excerpt from the Definitions and Key Terms List

- **“Guests”** : comments on/about other guests. Can be positive, negative, or just generally about other guests at the hotel (e.g. compliments or complaints). Also includes other non-staff people who come to the hotel and its facilities (e.g., vendors) because we can’t infer whether the hotel allows them or not. Also includes absence of other people or crowded
EB: should we categorize just as “Guests” rather than complaints about guests - this way it can reflect any types of comments about them? For example, for service we just say “service” and it incorporates comments on good and bad service VR: I like this idea re: changing it from “Complaints about guests” to “Guests” and have made that change. So perhaps definition should be instead of “complaints about other guests’ behavior, interactions etc” to “comments on other guests of the hotel” (I made this change);
 - What about when mention something/someplace is “crowded”? Would this be under guests? What about if beach is open to non-guests→ but we can think of this as guests to the beach since they presumably don’t live there and aren’t workers

This excerpt concerns our definition of the term “Guests”. The discussion is color-coded, with the code in red, the definition in black, and comments from coauthor A in purple and coauthor B in yellow. A few things to note. First, through this example, you can see the discussion of how and why we changed the code from “complaints about guests” to “guests” so as to capture positive, negative, and other comments about the guests more generally. It expanded the original code that focused solely on complaints. You can also see the comments

raised about whether to include vendors or others who came to the hotel to sell items but were not necessarily staying overnight at the hotel as “guests.” Per the reviews themselves, sometimes vendors were unwanted by guests staying at the hotel. However, there is no way to ascertain whether vendors are “wanted” by the hotel staff or not. We decided to include anyone who was not a hotel staff member as a “guest” since presumably they do not work at the hotel nor do they live there.

After the first quarter, only one coauthor continued to work with undergraduate students in separate teams of two to three students each, while the other coauthors worked on other parts of the project. Although conversations across teams no longer occurred since the coauthor was leading all of them, we were still able to document the questions raised and the rationale for our decision-making process. For example, in review 245, Icaarr states “if you're looking for hotel just to spend night at, then this is a good one. Just can't get over of the frog in the pool. Lol. The service is also good. It's very near to the night life in SBFZ, blocks away from restaurants and malls.” This stimulated a discussion around how to classify particular animals, whether as nature or a symbol of cleanliness. Our discussion, “Where do animals like frogs in an outdoor pool fit? Is it Nature (comment on animals) or is it Cleanliness (a comment on whether clean the pool)” can be found in our definitions and key words list and in the difference list, discussed below.

Similarly, in a meeting the PI had with another pair of undergraduate research assistants, we had a discussion - after two quarters of coding restaurants as a hotel amenity - over whether a separate code on “restaurant” should be added, why this would be important to add, and how it relates to other terms. We decided it was a necessary code to add, and undergraduate researchers recoded the entire TripAdvisor dataset to include this new code. However, we also decided comments about restaurant staff/service would still be categorized as “service” more generally

rather than adding a “restaurant service” code because the purpose of the code was to document distinct comments on the overall quality and feel of the restaurant. After coding hundreds of reviews, we began to understand how the quality of the hotel and restaurant may be quite different. Separating these codes allow us to get more fine-tuned information regarding circumstances around food, a particular subject that is overtly racialized. Both the frog and restaurant examples show the flexibility of the living codebook and demonstrate how conversations over the coding process continue even after initial training. We should note that we keep both the original definitions and key terms document and a clean version, without the ensuing discussions, the latter of which is a helpful reference document.

Memoing

Memos and memo-writing are the third key component of the living codebook, which has been extensively discussed elsewhere (e.g., Charmaz 2006). Indeed, the lead coauthor used Charmaz’s (2006) examples of a memo (e.g., Box 4.1: Example of a Memo - Suffering, pp 73-74) and particular questions memos should address (e.g., Box 4.3: How to Write Memos, pp 80-81) to train all other team members on what memos were and how to write them. That is, to use memo-writing as a way to identify the patterns that the research assistants see as they go through the data, raise questions, discuss what they are seeing versus what is not being said, and always rooting their discussions in particular examples from the data.

In the course of the project, we found that some of the undergraduates needed a bit more structure in how to think about memo-writing. Although the coauthors didn’t use these prompts for their own coding and memo-writing for a different set of data, we found it useful for undergraduate research assistants. We discussed how memo-writing always takes longer than

expected and to set aside 1-2 hours to do so, and list three different sections (coding, reflections, and reflexivity) they should address in each memo. The first section is the main one: coding. Here, we list many of the same prompts that Charmaz (2006) details though we also adapt it to our specific project. For example, it includes questions such as answering what people are doing to thinking about the processes at issue in the review (see table 2). The second section is about process and questions. That is, it is a space where students discuss what is unclear, what is difficult and whether these things have changed since the beginning of the project. The third part is on reflexivity - to what extent do they relate to the themes and concepts they are seeing. These are important pedagogical prompts that ask students to reflect on the learning process and their own experiences to make sure their participation in this project promotes active learning principles. Sometimes the reflexivity section provides more in-depth discussions on concepts, definitions, and analytic codes than other sections of the memo.

Let us walk through an example from two students paired together to code the TripAdvisor reviews. In their memos for the week, they both discussed “safety.” However, despite coding the reviews similarly, the way they discussed the broader concepts differed. Student A discussed safety as more direct and specific--touching on physical safety issues within a particular hotel. This student notes, “some guests who are regulars recommended upgrades or renovations to the hotel. One was very descriptive in upgrades such as the staircase railing. They would mention how they would wobble and were rusting” (Memo: Student A 5/7). In contrast, student B discussed safety in a broader way, exploring the concept of safety from a macro perspective. This student writes “I am seeing more of ‘safety’ because I’ve been going back to recode and revise the themes. Defining safety has changed because in the past, I would not have recognized safety if it was not obviously stated within the review” (Memo: Student B - 5/6/2018

- Coding Section) and elaborates “if safety is an issue with the hotel and its surroundings, then what does that say about the area? What does that say about the country/region overall? If there is a lack of safety, then is the country considered ‘lesser’ or more ‘exotic?’” (Memo: Student B - 5/6/2018 - Reflexivity Section). While weekly memos often will cover similar reviews and/or common themes they coded from their reviews, they also reveal insights into the different ways students analyze and understand what they are coding.

Another example of how people may interpret data differently and use their memos to explain their thought process concerns Review 5 in a set of practice codes that the PI had all eight teams she was working with that quarter to code, so as to compare coding practices across teams and ensure consistency. Both students in one team commented that how the reviewer talked about the housekeeping staff related to racism and classism. One of these students wrote in their memo:

I found the third review interesting because it stated that the guest lost a pearl earring in their room and it was returned to the front desk where they were ultimately able to retrieve it. They claimed the hotel have “have honest and efficient housekeepers.” I wonder whether this statement is genuine excitement or if the statement is holds racist or a “poverty blaming” perspective the reviewer holds. In my experience, those who do not have faith in the security of an establishment, hold this view because they have discriminatory beliefs. For example, believing all housekeepers are Mexican and therefore they will steal your valuable belongings. I also wonder if this is an American view that only pertains to the United States. Personally, when I travel, I do not necessarily worry my valuables will be stolen; I don’t even touch the safe that is commonly provided in hotels. However, there may be other individuals who have negative experiences or discriminatory beliefs that leads them to be extremely cautious. This is something I would like to explore further (Memo – Reflexivity section, Student C, 10/8/19).

Here, the student links a review where the reviewer lost a pearl earring and was seemingly surprised that the housekeeping staff found and returned it to issues regarding stereotypes of housekeeping staff across the Global South. In the team meeting, these two students made a convincing argument on how surprise at the hotel employing honest

housekeeping staff related to possible issues of stereotypes, racism, and classism. Only two other teams brought up a similar reading of this review. The other five teams did not. This memo also reveals how analysis can be explicitly related to researchers' positionality, as this Latina student links her own experiences, including traveling abroad, and knowledge regarding harmful racist and prejudice stereotypes regarding Mexican housekeepers by white Americans, to the content of the review.³

Table 2: Analytic Template for Undergraduate Student Researchers, which uses Charmaz (2006: 80-81)

Section 1: Coding. Here are a few things to think about as you're coding. Remember that you do not have to address each bullet point; however, we ask that you try to address at least 2-3 of these points in each memo.

- What are the main themes you are seeing in the accounts? How would you define those themes (similarly or differently than we already have)? If it's a new theme, how are you defining it and why?
- Describe how the themes have emerged or changed. What data supports it?
- What do particular themes you saw this week look like from various vantage points (e.g. from a review of a person from the Philippines vs. someone from abroad)
- What are people doing? What are people saying? How do these connect?
- What do the reviewers' actions and statements take for granted?
- How do structure and context serve to support, maintain, impede or change their actions and statements?
- What processes are at issue in your reviews (e.g. racializing, poverty blaming, elitism)? And under which conditions does this process develop (e.g. where is the reviewer from or the reason for their visit?)
 - Are the reviews aware of these processes? If so, how do they seem to think, feel, or act while they're involved?
- Does the process appear to change for any reason? If so, when, why, and how does the process change (e.g. does it change depending on where the participant is from or by the type of hotel they are staying at?)
- Make comparisons amongst themes or processes.
 - Compare different reviewers (based on beliefs, situations, actions, accounts, or experiences)
 - Compare themes in the data with other themes

Section 2: Reflections on learning the coding process

- Are all the parts of the coding process clear?
- Do you need clarification about the process or what any of the documents are for?
- Is this harder/easier than you thought? Why?

- Has the process of coding changed for you from the beginning stages to now?

Section 3: Reflexivity

- How you are relating to the themes and concepts you are finding in the reviews and coding?
- How do the themes and concepts relate to your own experiences and your own social location?
- Compare the themes or sub-themes as you have personally experienced them.
 - Compare sub-categories of themes with the fit within categories? Are there clear differences between sub-categories and what differentiates those differences?
 - Compare the entire analysis with other literature you've read on the topic?

Difference list

The final component of the living codebook is the most novel: what we are calling the difference list. As the name suggests, the difference list tracks differences in codes. In other words, it documents the codes that one person on the team assigned but the other did not. It also includes the reasoning behind assigning or not assigning any given code, the final decision on whether to include the aforementioned code or not, and the reasoning why. Let us walk through an example from the TripAdvisor data (see table 3).

Table 3: Excerpt of a Difference List for the TripAdvisor Data

Review 4 (Attractions, Room Quality, Safety)

- **Student A**
 - Room Quality: the executive room is “more than adequate”. Based on that comment, I assumed that the room was big enough or has the essential necessities, which speak to the quality of the room.
 - Safety: “guarded parking”. I took this as a comment on the safety of the hotel -> the guests won't have to worry about losing their belongings if they left it in their cars or moped/motorcycle.
 - Didn't put attraction: looking back, I should have put this term, because of “near shopping”.
- **Student B**
 - Added attractions because “near shopping”

- Did not add room quality because it did not speak about the room quality.
- Did not add safety but would add because of guarded parking.

Decision:

- Add: attractions, room quality, safety
- Delete: .

Each team, composed of 2-3 students, met weekly with one of the coauthors. In preparation for the weekly meetings, student researchers were required to code a certain number of reviews, write a memo, and - after initial training was completed and students felt more comfortable with the purpose and use of the difference list - fill out the difference list. The difference list for the TripAdvisor data is a list of each review and each code that was different than the other team member(s) (see table 3). That is, the processual database was set up in such a way that the coauthor was able to compare codes for each review and identify similarities as well as any differences. The coauthor would then list out each review and the accompanying different codes from that review, since each review had multiple codes assigned to it. Student researchers would then go through, explaining why they did or did not assign that code for the review, with evidence from the text itself. The team meetings were then centered on discussing the results of the difference list - going through each review, having each student researcher explaining why they did or did not assign a given code, and discussing whether or not the code should ultimately be added or deleted to the “Final codes” column. As the process became streamlined, teams met on a biweekly basis to discuss their memos and difference list, because the student researchers relied on one another to complete their codes in a timely manner. During the first week, student research teams coded a set number of reviews. For the second week, sometime before the team meeting, each member would fill out the difference list that the team leader (coauthor) created based on their codes the week before (see figure 4).

Figure 4: Living Codebook Process

Note: This is the process of using the living codebook after training

This process, although time intensive, is important for many reasons. It allows us to immerse ourselves in the data and investigate why there were deviations, rather than taking the deviations for face value vis-a-vis an intercoder reliability score. Were these differences due to simple mistakes, vague definitions of codes, ambiguous data, differences in social positions or something else? We found it to be the case for all of the above, at different times. Regarding simple mistakes: after revisiting the review, student researchers often didn't understand why they coded the way they did and would recode the review. For example, in a March 15, 2019 meeting, one student remarked as we're going through the difference list, that "old [student name]" coded it that way but now he would code it differently and in his memo for the week - written on March

6, 2019 - he wrote “Finally, like I’ve said numerous times during our team meetings, the difference list has really helped me see my progress in the coding process. I’ve even expressed that I want to go back and redo most, if not all, of my original coding because I’m confident that I can do it much better now.” The difference list helps students reflect and explain why they coded a review the way they did, and mistakes may be due to just getting familiar with the data. Yet, it is not simply a matter of inadequate training, nor something that occurs only at the beginning of the process. It is also a matter of the coding process is both an iterative process (see Altheide 1987 for his work on reflexivity in ethnographic content analysis) and a tedious task where simple mistakes can be made.

The difference list, and the accompanying process, also reveals when definitions of codes are insufficiently clear as to when it is applicable, and how codes are often situationally dependent. This includes the “guest” code that we described in the definitions and key words section, where our questions and subsequent discussions of who counts as a guest and whether a vendor would be counted as such was initially brought up in the difference list. Another example is our code “deception,” which we define as “any time a reviewer mentions misrepresentation about a hotel, experience, food, price, etc. This is an additional code, so if there is a misrepresentation about price, for example, a review would be coded both “deception” and “price”; explicit mentions of why they feel misrepresentative or deceived based on information from the hotel (which is different from expectations based on information not coming from hotel).” Madison723 writes in review 386 about their experience at The Lighthouse Marina Resort: “We were first booked at the other hotel but we transfer to this hotel for the rest of our 2 more night stay in subic last april. When in agora its fully book better check there web site coz its not really fully book and its more cheaper to walked in or call in there hotel.” Two student

coders paired together disagreed on whether this statement should be coded as deception. One student said in the difference list “I put deception because it mentioned that if it says fully booked on the website, it’s not” while the other wrote “I didn’t put deception: I think the reviewer was talking about agoda.com, which is a third party booking site. I don’t think that should count against the hotel, since the hotel itself isn’t responsible for the information on that website.”

The code “deception” often needed to be discussed because it was more situationally dependent than other codes. For example, one group had three student research assistants and they did not agree whether review 171 should include deception as a code. Student E explains “I chose illicit activity over deception because the charges that the hotel illegally put on the credit card is an illicit activity against the guest, rather than something about a stated price that was later increased without consent. It seemed more to me that the hotel staff/owners just stole their credit card information” while Student F explains “The reason for choosing deception because the overall review was explaining how the hotel was not being clear about their actions, especially in terms of their charges and amenities provided. The reviewer then explains as to how though the hotel was missing a lift and having poor location, their greatest shock was that of the surprise fees that appeared after they returned home as well as the missing website for the hotel when searched online.” Finally, Student G states “The reason I chose deception is because the main purpose of the review is to warn others of credit card fraud going on in the hotel. Although this was a scam, as described by the definition for illicit activity, the tone of the reviewer implies that they were deceived. It seems as though the reviewer is blaming the management because they state that when he tried complaining, the hotel’s website was no longer available.” This

experience of having to discuss situationally dependent codes occurs both early in the process and later.

Checking codes on a sample of the data misses the ambiguity of situations that occur in data not sampled (see also Altheide 1987). As such, the difference list also allows us to pinpoint when the data itself may be ambiguous with regard to whether a code applies. Since many codes are often situationally dependent, simply counting words or sampling parts of data to qualitatively code is inadequate. By documenting the decision process and reasoning behind codes, we provide our justifications for our interpretations, and leave these decisions open to scrutiny on the validity of these codes.

Conclusion

In this paper, we present a methodological tool to document the process of qualitative data analysis: the living codebook, which has four parts: processual and final databases, definitions and key word list, memoing, and the difference list. In presenting a way to reveal the often hidden decisions that go into the coding process, we build upon the recent work on transparency in qualitative research (e.g., Reyes 2018a, Jerolmack and Murphy 2017, Contreras 2019) and the previous work scholars have done in setting a foundation for what a qualitative codebook could look like (Weston et al 2001, Deterding and Waters 2018). The living codebook 1) gathers already existing tools qualitative scholars use in their data analysis, such as memo-writing, 2) introduces a new way to use a “definitions and key words” list by keeping track of decisions and conversations around codes, and 3) presents novel tools, such as the processual database and the difference list. It also combines these moving parts into a single entity to highlight their interconnections.

Some of these tools overlap in content. That is, both the processual database and the difference list document differences in codes among team members. This is purposeful and follows the reasoning behind the “3-2-1” data backup plan for quantitative scholars, where there are three different copies of the data maintained in at least two mediums, one of which is located on a remote drive. Qualitative scholars do not necessarily need to follow these tenets, but the idea of making sure our decision-making process is documented in different ways is important and helps prevent data loss. Furthermore, documenting the differences in codes in both the processual database and the difference list is useful and needed in order to use them to their fullest capacity. The different moving parts of the living codebook also build and reinforce one another, where the difference list explains reasoning behind including or excluding codes found in the Final codes column. Students often reiterate and expand on these processes and thoughts in their memos, demonstrating how each part of the living codebook is interconnected. In this way, it also is one way to show how abductive analysis (Timmermans and Tavory 2012) works in practice.

By highlighting examples from our TripAdvisor data - coded by undergraduate student researchers - and discussing how the tools were adapted for the data coded by two of the coauthors, we also show how these tools are meant to be flexible and adaptable to any given set of data and a research team’s needs, in a similar way that Deterding and Waters (2018) discuss how their codebook is flexible to the source of data. To that end, the living codebook has theoretical and methodological significance. Theoretically, it provides a conceptual way to think about the various moving pieces, whether memo-writing or databases, that researchers draw on in the course of qualitative data analysis. Methodologically, it provides concrete, but flexible, steps to document the various stages of qualitative analysis.

We acknowledge the limitations of the living codebook, and its accompanying processes, the most obvious of which is how time and labor intensive it is. The increasing pressure to publish for graduate students, postdoctoral fellows and adjuncts on the job market and for assistant professors on the tenure-track (e.g., Warren 2019) may mean that this level of documentation isn't as feasible as we suggest since calculating intercoder reliability is quick and sufficient to conduct and publish research. The multiple moving parts of the living codebook also requires a clear and detailed organization system so that all team members know the steps required of them at each stage in the process and where all the documentation is located. While our data involves the coding of documents, we also believe the living codebook is a useful tool for other types of qualitative research, such as interviewing. However, we also recognize, like Deterding and Waters (2018), that some types of qualitative research, such as intensive ethnography, and some goals of researchers may not align well with this process and using the living codebook for these situations is not advisable. Still, we believe that despite the time and labor this process required, the living codebook is an important conceptual and methodological tool for documenting the precise process of qualitative data analysis, which is often invisible and undocumented. Since we've outlined how the living codebook operates in our team process, future research can examine whether and to what extent these different tools can be adapted for use by solo researchers and for different types of qualitative data.

Endnotes

1. Some scholars critique coding as a means of conducting qualitative data analysis. For example, Biernacki (2012) argues that, in practice, coding is not a “social scientific” enterprise, but instead, a performative one based on ritual, where scholars strip so-called “data” from context, engage in circular arguments and data collection, and tell the story that they want to impose rather than the actual meanings that emanate from the materials. Lee and Levi Martin (2015) agree with many of Biernacki (2012)’s critiques about coding as a failed blending of “science” and “interpretation” and they suggest that a formal analysis - a la counting and maps of counts, what they call cartography- is a way forward for the *scientific* analysis of cultural materials. We agree that Biernacki (2012) book raises important critiques and questions that sociologists engaged in cultural analyses and coding need to address. However, we are not ready to disregard coding as a useful scientific method. It is not the purview of this paper to address all of Biernacki’s critiques. Instead, we address the potential problems of secrecy and the “black box of coding” through what we call the living codebook, which brings to light and documents the often invisible, cloaked decisions that surround coding procedures and practices that both Biernacki (2012) and Lee and Levi Martin (2015) bring up.
2. <https://tripadvisor.mediaroom.com/us-about-us>, last accessed 12/4/2017
3. Both students in this team were Latinx. However, Latinx students in other teams did not make these same connections. We are not suggesting that all Latinx, or all members of a certain group, would have the same particular “special insight” into the data. Rather, we are saying that this student (and her coding partner) both drew on their own experiences and positionality to analyze the data, which resulted in their making these connections.

References

- Abbott, Andrew. 2016. *Processual Sociology*. Chicago, IL: University of Chicago Press.
- Abramson, Corey and Daniel Dohan. 2015. "Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data." *Sociological Methodology* 45(1):272–319.
- Abramson, Corey, Jacqueline Joslyn, Katharine Rendle, Sarah Garrett, and Daniel Dohan. 2018. "The Promises of Computational Ethnography: Improving Transparency, Replicability, and Validity for Realist Approaches to Ethnographic Analysis." *Ethnography* 19(2):254–84.
- Altheide, David L. 1987. "Reflections: Ethnographic Content Analysis" *Qualitative Sociology* 10(1): 65-77
- Beck, Colin. 2017. "The Comparative Method in Practice: Case Selection and the Social Science of Revolution." *Social Science History* 41(3):533–54.
- Beck, Colin. 2018. "The Structure of Comparison in the Study of Revolution." *Sociological Theory* 36(2):28.
- Biernacki, Richard. 2012. *Reinventing Evidence in Social Inquiry: Decoding Facts and Variables*. London: Palgrave MacMillan.
- Burawoy, Michael. 2003. "Revisits: An Outline of a Theory of Reflexive Ethnography." *American Sociological Review* 68(5):645–79.
- Charmaz, Kathy. 2006. *Constructing Grounded Theory*. Thousand Oaks, CA: Sage Publications.
- Contreras, Randol. 2019. "The Broken Ethnography: Lessons from an Almost Hero." *Qualitative Sociology* 42(2):161–79.
- Contreras, Randol. 2013. *The Stickup Kids: Race, Drugs, Violence, and the American Dream*. Berkeley, CA: University of California Press

- Deterding, Nicole and Mary Waters. 2018. "Flexible Coding of In-Depth Interviews: A Twenty-First-Century Approach." *Sociological Methods & Research* 1–32.
- Duck, Waverly. 2015. *No Way Out: Precarious Living in the Shadow of Poverty and Drug Dealing*. Chicago: University of Chicago Press.
- Flores, Glenda. 2016. "Discovering a Hidden Privilege: Ethnography in Multiracial Organizations as an Outsider Within." *Ethnography* 17(2):190–212.
- Fuentes, Marisa. 2016. *Dispossessed Lives: Enslaved Women, Violence, and the Archive*. Philadelphia, PA: University of Pennsylvania Press.
- Gamson, William and Gadi Wolfeld. 1993. "Movements and Media as Interacting Systems." *The ANNALS of the American Academy of Political and Social Science* 428(1):114–25.
- Glaser, Barney and Anselm Strauss. 1999. *Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Routledge.
- Gray, Alex. 2017. "These are the world's most popular websites," World Economic Forum, <https://www.weforum.org/agenda/2017/04/most-popular-websites-google-youtube-baidu>, last accessed 8/8/2019
- Guenther, Katja. 2009. "The Politics of Names: Rethinking the Methodological and Ethical Significance of Naming People, Organizations, and Places." *Qualitative Research* 9(4):411–21.
- Hsieh, Hsiu-Fang and Sarah E Shannon. 2005. "Three Approaches to Qualitative Content Analysis" *Qualitative Health Research* 15:1277-1288
- Jerolmack, Colin and Shamus Khan. 2014. "Talk Is Cheap: Ethnography and the Attitudinal Fallacy." *Sociological Methods & Research* 43(2):178–209.

- Jerolmack, Colin and Alexandra K. Murphy. 2017. "The Ethical Dilemmas and Social Scientific Trade-Offs of Masking in Ethnography." *Sociological Methods & Research* 1–27.
- Katz, Jack. 1997. "Ethnography's Warrants." *Sociological Methods & Research* 25(4):391–423
- Lee, Monica and John Levi Marin. 2015. "Coding, Counting, and Cultural Cartography" *American Journal of Cultural Sociology* 3(1):1-33
- Lester, Marilyn. 1980. "Generating Newsworthiness: The Interpretive Construction of Public Events." *American Sociological Review* 45(6):984–94.
- Mahoney, James. 2004. "Comparative-Historical Methodology." *Annual Review of Sociology* 30(1):81–101.
- Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41(4):570–97.
- Myers, Daniel and Beth Schaefer Caniglia. 2004. "All the Rioting That's Fit to Print: Selection Effects in National Newspaper Coverage of Civil Disorders, 1968-1969." *American Sociological Review* 69(4):519–43.
- Oliver, Pamela and Daniel Meyer. 1999. "How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events." *American Journal of Sociology* 105(1):38–87.
- Reyes, Victoria. 2018a. "Three Models of Transparency in Ethnographic Research: Naming Places, Naming People, and Sharing Data." *Ethnography* 19(2):204–26.
- Reyes, Victoria. 2018b. "Ethnographic Toolkit: Strategic Positionality and Researchers' Visible and Invisible Tools in Field Research" *Ethnography*, published online first October 25, 2018: <https://doi.org/10.1177/1466138118805121>

- Reyes, Victoria. 2019. *Global Borderlands: Fantasy, Violence, and Empire in Subic Bay, Philippines*. Stanford, CA: Stanford University 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.
- Timmermans, Stefan and Iddo Tavory. 2012. “Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis.” *Sociological Theory* 30(3):167–86.
- Tuttle, Brad. 2009. “Five great travel booking sites” CNN Travel Companion, <http://www.cnn.com/2009/TRAVEL/traveltips/06/02/travel.booking.sites/index.html>, last accessed 8/8/2019
- Warren, John Robert. 2019. “How much do you have to publish to get a job in a top sociology department? Or to get tenure? Trends over a generation” *Sociological Science* 6:172-196
- Weston, Cynthia, Terry Gandell, Jacinthe Beauchamp, Lynn McAlpine, Carol Wiseman, and Cathy Beauchamp. 2001. “Analyzing Interview Data: The Development and Evolution of a Coding System.” *Qualitative Sociology* 24(3):381–400.
- Yin, Robert. 2011. *Qualitative Research from Start to Finish*. New York: The Guilford Press.
- Yin, Robert. 2002. *Case Study Research: Design and Methods*. Third. Thousand Oaks, CA: Sage Publications.