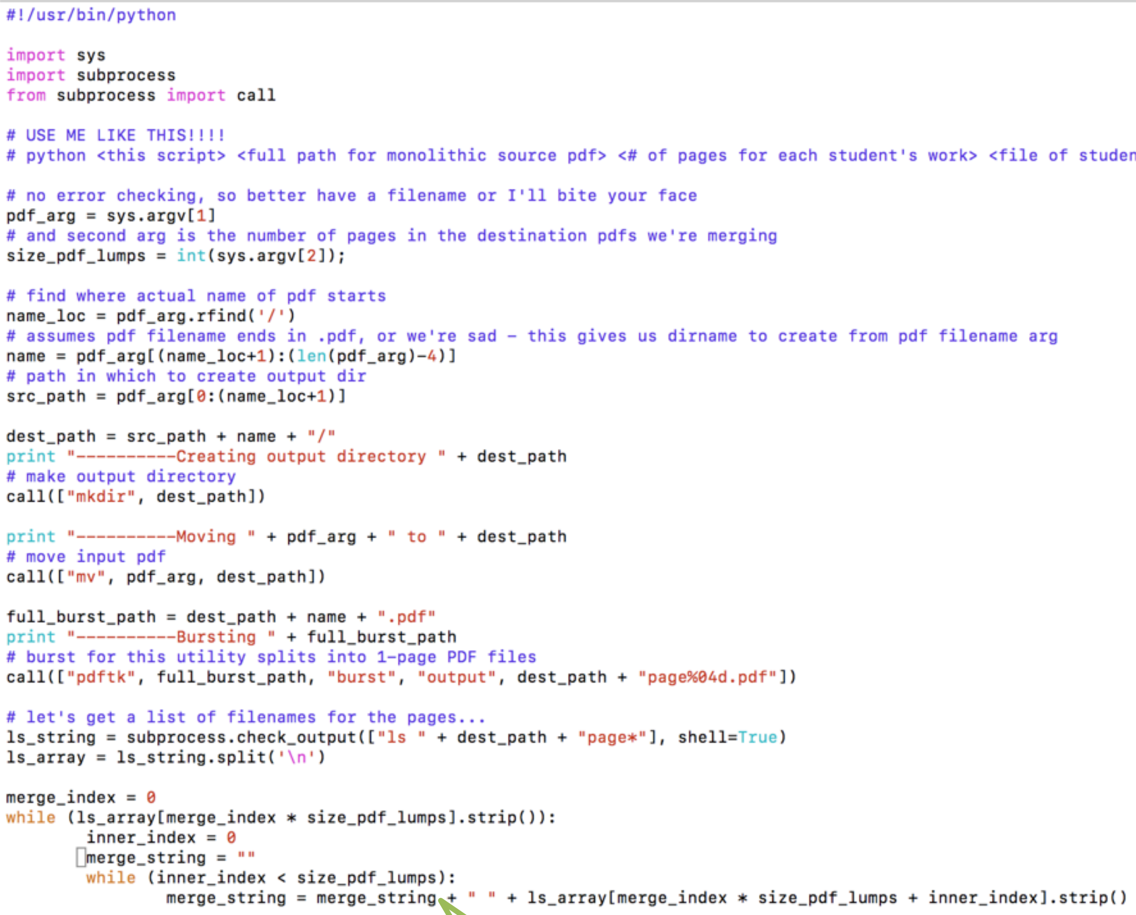
(For Computer Scientists)

CODING

(For Computer Scientists)

CODING

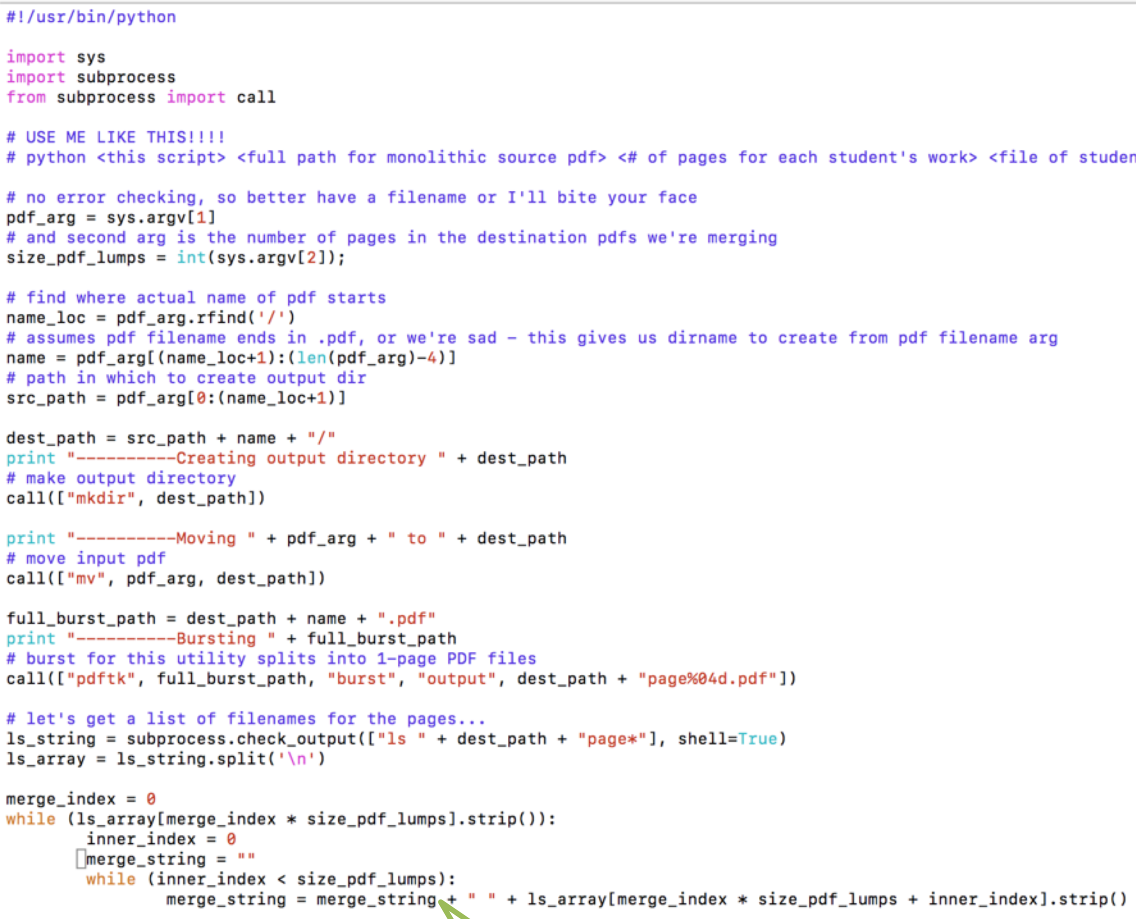


```
#!/usr/bin/python

import sys
import subprocess
from subprocess import call

# USE ME LIKE THIS!!!!
# python <this script> <full path for monolithic source pdf> <# of pages for each student's work> <file of studen

# no error checking, so better have a filename or I'll bite your face
pdf_arg = sys.argv[1]
# and second arg is the number of pages in the destination pdfs we're merging
size_pdf_lumps = int(sys.argv[2]);

# find where actual name of pdf starts
name_loc = pdf_arg.rfind('/')
# assumes pdf filename ends in .pdf, or we're sad - this gives us dirname to create from pdf filename arg
name = pdf_arg[(name_loc+1):(len(pdf_arg)-4)]
# path in which to create output dir
src_path = pdf_arg[0:(name_loc+1)]

dest_path = src_path + name + "/"
print "----------Creating output directory " + dest_path
# make output directory
call(["mkdir", dest_path])

print "----------Moving " + pdf_arg + " to " + dest_path
# move input pdf
call(["mv", pdf_arg, dest_path])

full_burst_path = dest_path + name + ".pdf"
print "----------Bursting " + full_burst_path
# burst for this utility splits into 1-page PDF files
call(["pdftk", full_burst_path, "burst", "output", dest_path + "page%04d.pdf"])

# let's get a list of filenames for the pages...
ls_string = subprocess.check_output(["ls " + dest_path + "page*"], shell=True)
ls_array = ls_string.split('\n')

merge_index = 0
while (ls_array[merge_index * size_pdf_lumps].strip()):
        inner_index = 0
        merge_string = ""
        while (inner_index < size_pdf_lumps):
                merge_string = merge_string + " " + ls_array[merge_index * size_pdf_lumps + inner_index].strip()
```

(not this kind)

(For Computer Scientists)

CODING
AND
QUALITATIVE ANALYSIS



```
#!/usr/bin/python

import sys
import subprocess
from subprocess import call

# USE ME LIKE THIS!!!!
# python <this script> <full path for monolithic source pdf> <# of pages for each student's work> <file of studen

# no error checking, so better have a filename or I'll bite your face
pdf_arg = sys.argv[1]
# and second arg is the number of pages in the destination pdfs we're merging
size_pdf_lumps = int(sys.argv[2]);

# find where actual name of pdf starts
name_loc = pdf_arg.rfind('/')
# assumes pdf filename ends in .pdf, or we're sad - this gives us dirname to create from pdf filename arg
name = pdf_arg[(name_loc+1):(len(pdf_arg)-4)]
# path in which to create output dir
src_path = pdf_arg[0:(name_loc+1)]

dest_path = src_path + name + "/"
print "----------Creating output directory " + dest_path
# make output directory
call(["mkdir", dest_path])

print "----------Moving " + pdf_arg + " to " + dest_path
# move input pdf
call(["mv", pdf_arg, dest_path])

full_burst_path = dest_path + name + ".pdf"
print "----------Bursting " + full_burst_path
# burst for this utility splits into 1-page PDF files
call(["pdftk", full_burst_path, "burst", "output", dest_path + "page%04d.pdf"])

# let's get a list of filenames for the pages...
ls_string = subprocess.check_output(["ls " + dest_path + "page*"], shell=True)
ls_array = ls_string.split('\n')

merge_index = 0
while (ls_array[merge_index * size_pdf_lumps].strip()):
        inner_index = 0
        merge_string = ""
        while (inner_index < size_pdf_lumps):
                merge_string = merge_string + " " + ls_array[merge_index * size_pdf_lumps + inner_index].strip()
```

(not this kind)

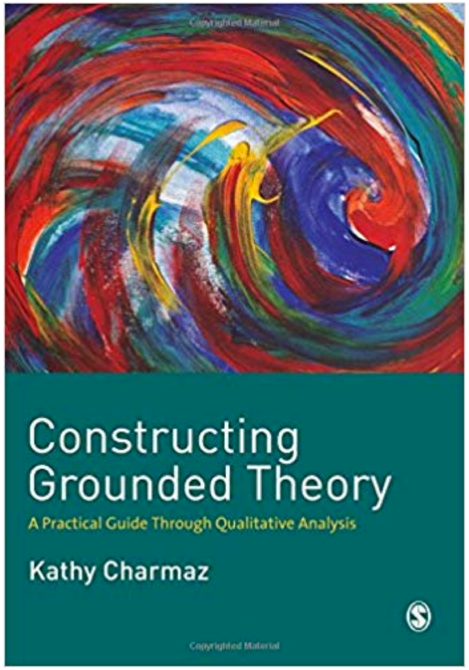
We’re only discussing the tip of the iceberg in this slide deck, not becoming experts.

https://www.flickr.com/photos/kmortara4/27320897699/

As with most things, you learn by doing.

https://www.flickr.com/photos/ucitysoccer/1498720580/

The Good



Constructing Grounded Theory

A Practical Guide Through Qualitative Analysis

Kathy Charmaz

Joseph A. Maxwell

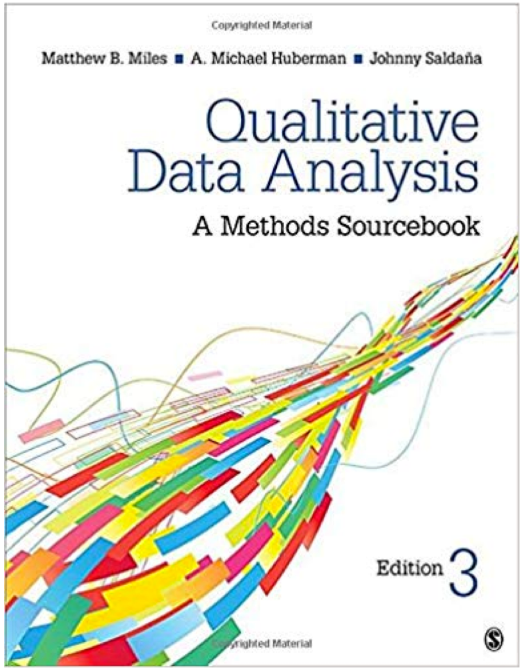
Qualitative Research Design

An Interactive Approach

3 EDITION

41 APPLIED SOCIAL RESEARCH METHODS SERIES

Edited by Leonard Bickman and Debra J. Rog



Matthew B. Miles ■ A. Michael Huberman ■ Johnny Saldaña

Qualitative Data Analysis

A Methods Sourcebook

Edition 3

There are a lot of available reference materials (like the above).

The Bad

Kind of like “I know how to speak so I can interview people and don’t need to learn anything….”

It’s easy to not even know you’re cutting corners. There’s always more to learn about doing a better job at analysis!

https://commons.wikimedia.org/wiki/File:Rebus3.png

The Ugly

- As either a reader or a writer, there is never going to be enough space in a publication to go into enough detail on the methodology that you can have 100% (or 99%, or 90%…) confidence that the research was performed with all due attention to validity and reliability. The best that can be done is to describe enough steps or details to give confidence that the researcher: a) has an idea of ways that reliability and validity can be achieved; and b) applied critical thinking and paid attention to details.

The Ugly

- As either a reader or a writer, there is never going to
be e ugh
det 00%
(or was
per
relia ibe
eno the
rese y and
vali
thinki

There are multiple related issues here.

1. Learning enough/continuing to learn more/taking measures so that **you** are confident/proud of the quality of your analysis

2. Trying to tell from someone's description of their qualitative analysis procedures whether **they** performed a rigorous, careful analysis

3. Writing about the steps you took or choosing which steps you'll take so that you can **convince readers/reviewers** that your analysis was careful, rigorous, and valid

The Ugly

- As either a reader or a writer, there … to
  be e
  deta
  (or … was
  per
  relia … be
  eno … the
  rese … y and
  vali
  thinki

There are multiple related issues here.

1. Learning enough/continuing to learn n… measures so that **you** are confident/proud … of your analysis
2. Trying to tell from someone's descripti… r their qualitative analysis procedures whet… **they** performed a rigorous, careful analysis
3. Writing about the steps you took or choosing which steps you'll take so that you can **convince readers/reviewers** that your analysis was careful, rigorous, and valid

Don't let (3) compromise (1), but they're probably at least somewhat aligned

The Ugly

- As either a reader or a writer, th to
be e
det
(or
per
relia be
eno the
rese y and
vali
thinking

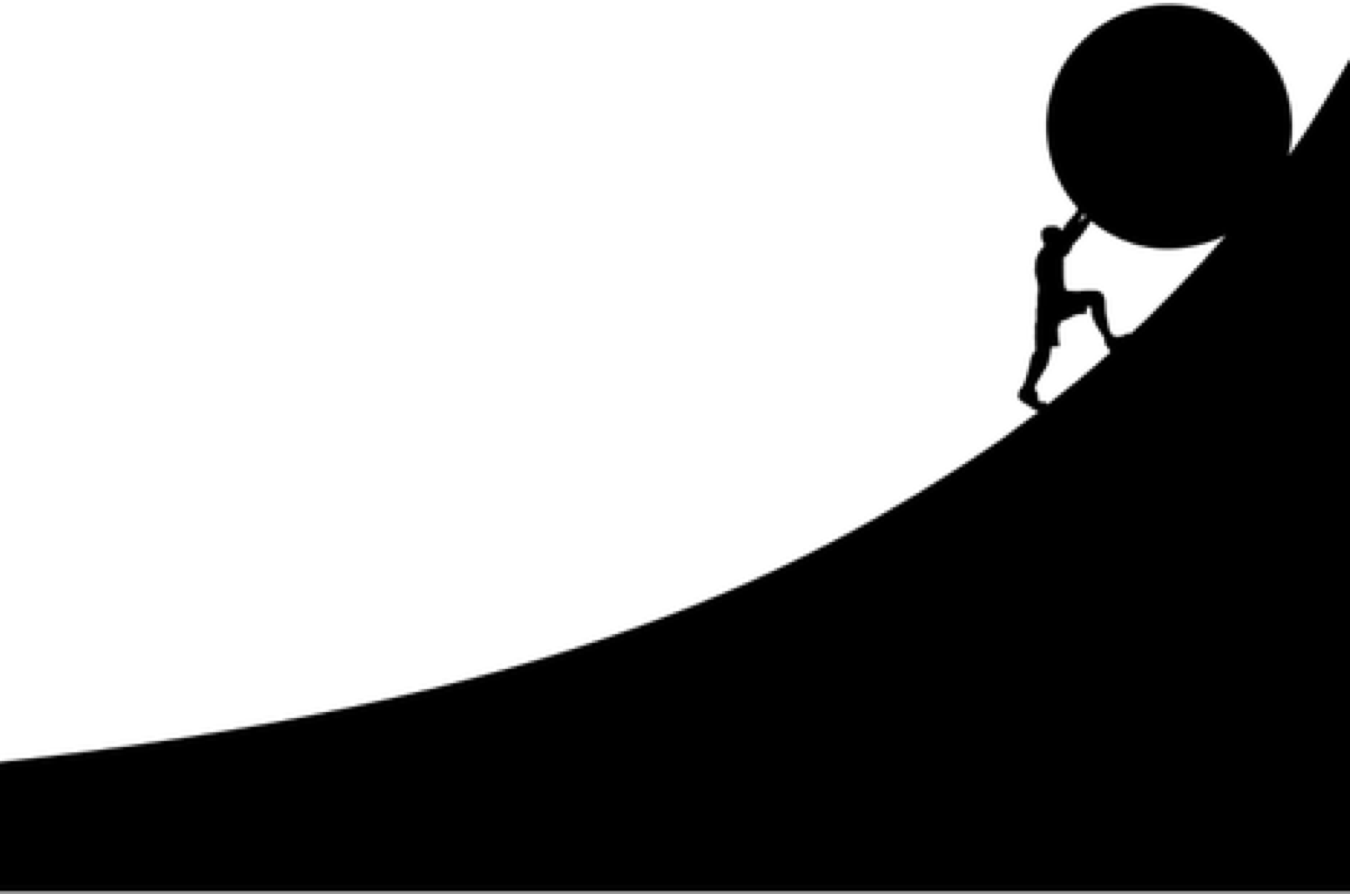
There are multiple related issues her

1. Learning enough/continuing to measures so that **you** are confi of your analysis

2. Trying to tell from someone's description of the qualitative analysis procedures whether **they** performed a rigorous, careful analysis

3. Writing about the steps you took or choosing which steps you'll take so that you can **convince readers/reviewers** that your analysis was careful, rigorous, and valid

Releasing datasets and/or coding manuals and/or inter-rater reliability scores helps, and we shouldn't give up, but….



https://publicdomainvectors.org/en/free-clipart/Sisyphus-overcoming-silhouette/72179.html

This Lecture: Eclectic/“Foundational” Approach

- Attempting to talk about generally useful approaches/advice that are either from different “named” analysis approaches or common to multiple approaches

This Lecture: Eclectic/“Foundational” Approach

- Attempting to talk about generally useful approaches/advice that are either from different “named” analysis approaches or common to multiple approaches

- Named approaches, e.g.
  - Grounded theory
  - Thematic analysis
  - Content analysis

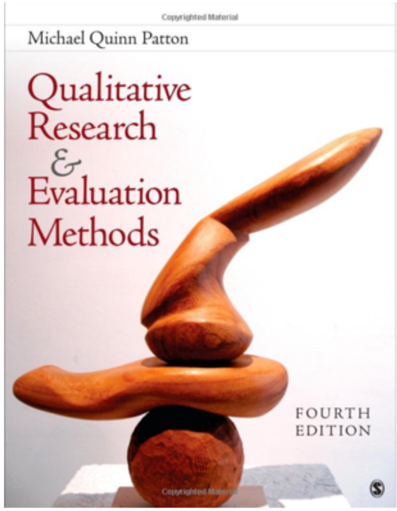
Warning

Not all methods or subcomponents of "named" methods – especially those with rich and specific philosophical backgrounds – can be blindly mixed and matched without compromising the context in which the method is viewed as being valid

Warning

Not all methods or subcomponents of “named” methods – especially those with rich and specific philosophical backgrounds – can be blindly mixed and matched without compromising the context in which the method is viewed as being valid

You may be interested in Chapter 3 (Variety of Qualitative Inquiry Frameworks: Paradigmatic, Philosophical, and Theoretical Orientations) →



Grounded Theory

- Start by coding that's very grounded in the data…eventually progress to theory

Grounded Theory

- Start by coding that's very grounded in the data…eventually progress to theory

This sort of sounds like what you'd always be trying to achieve in qualitative data analysis, but it's a particular set of steps/worldview, so **don't claim you're doing grounded theory just because you're using inductive coding**

Grounded Theory

- Start by coding that's very grounded in the data…eventually progress to theory

This sort of sounds like what you'd always be trying to achieve in qualitative data analysis, but it's a particular set of steps/worldview, so **don't claim you're doing grounded theory just because you're using inductive coding**

You'll probably just convince any readers who know enough to tell that you aren't using grounded theory that you didn't know what you were doing in the rest of the analysis either

Grounded Theory

- Start by coding that’s very grounded in the data…eventually progress to theory

- More particular, structured steps compared to some other approaches

Grounded Theory

- Start by coding that’s very grounded in the data…eventually progress to theory
- More particular, structured steps compared to some other approaches
- At this point, multiple “camps,” so you’d need to specify which you’re using

Thematic Analysis

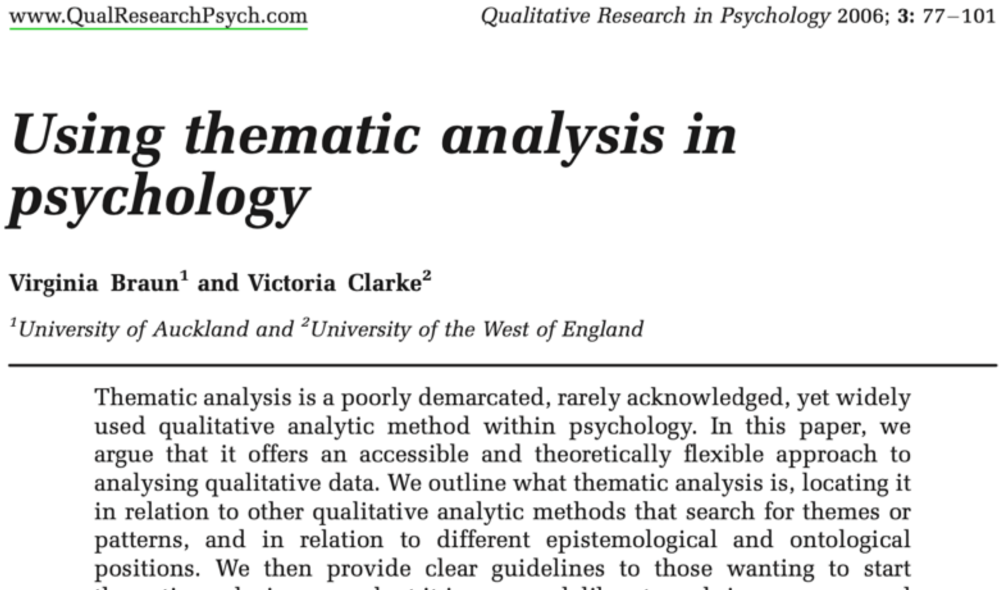
- Again, might mean different things to different people unless you describe or cite exactly what you mean (no, randomly citing a paper or textbook because it says "thematic analysis" doesn't count unless it's a very accurate description of what you did)

Given this, and especially if you're publishing in a venue that doesn't see a lot of qualitative work, maybe you want to explicitly describe as many of your analysis steps/steps to ensure quality as you can for the readers

Thematic Analysis

- Again, might mean different things to different people unless you describe or cite exactly what you mean (no, randomly citing a paper or textbook because it says "thematic analysis" doesn't count unless it's a very accurate description of what you did)

I recommend this read



www.QualResearchPsych.com *Qualitative Research in Psychology* 2006; **3:** 77–101

Using thematic analysis in psychology

Virginia Braun[1] and Victoria Clarke[2]

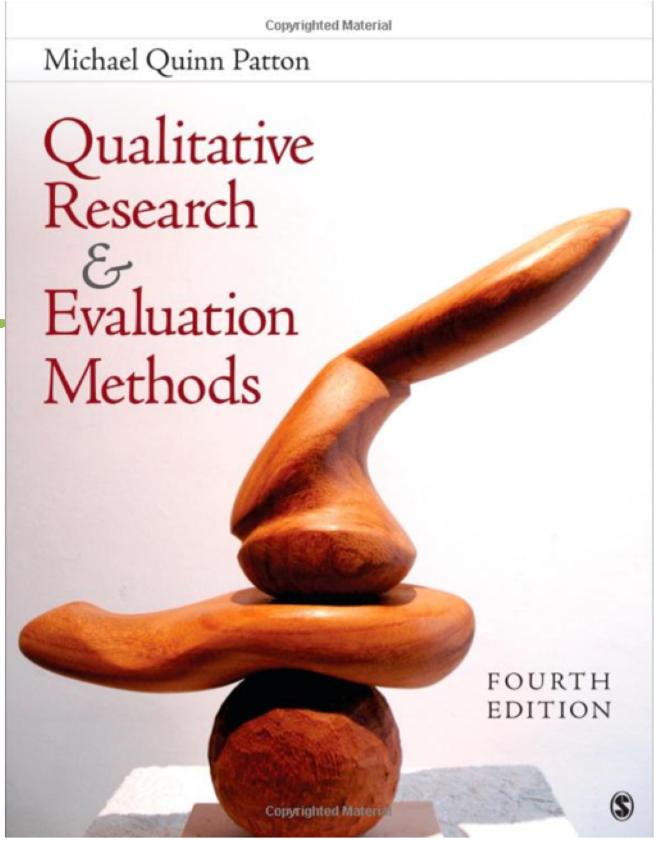
[1]University of Auckland and [2]University of the West of England

Thematic analysis is a poorly demarcated, rarely acknowledged, yet widely used qualitative analytic method within psychology. In this paper, we argue that it offers an accessible and theoretically flexible approach to analysing qualitative data. We outline what thematic analysis is, locating it in relation to other qualitative analytic methods that search for themes or patterns, and in relation to different epistemological and ontological positions. We then provide clear guidelines to those wanting to start

Content Analysis

- Again, it’s challenging finding a single authoritative definition

e.g. this book (which I also recommend) also describes it as meaning different things to different people



Content Analysis

- Again, it’s challenging finding a single authoritative definition

- …but for some people it seems to imply a stronger focus on quantitative analysis of coded data and counts in code categories

codes

- words or short phrases
- summarize, describe, organize, identify concepts in data

codes

- words or short phrases
- summarize, describe, organize, identify concepts in data

What is data?

codes

- words or short phrases
- summarize, describe, organize, identify concepts in data

What is data?

- interview transcripts
- participant diaries
- observer notes
- video recordings
- …

coding manual

- **Coding manual/coding rubric/codebook:** The set of codes (probably organized into categories & subcategories) which you will be applying/have applied to the data and definitions that sufficiently describe what they mean (i.e., when they should be applied).

coding

`coding`

is an overloaded term

coding

is an overloaded term

- inductive **coding** = an approach that is more consciously driven by what things you find in the data, whether or not you were expecting/planning to look for them **(bottom-up)**

- deductive **coding =** an approach that is more consciously driven by trying to apply a pre-determined set of codes to the data **(top-down)** (e.g. from so-and-so’s theory of blah there are five ways to blah, so I’m coding for appearances of 1-2-3-4-5)

coding

is an overloaded term

- **inductive coding** = an approach that is more consciously driven by what things you find in the data, whether or not you were expecting/planning to look for them (**bottom-up**)

I’m going to mainly talk about this, since it’s one of the strengths of qualitative methods (i.e., “let’s try to understand what’s going on in this space instead of imposing our preconceived assumptions”)

coding

is an overloaded term

- inductive **coding** = an approach that is more consciously driven by what things you find in the data, whether or not you were expecting/planning to look for them (**bottom-up**)

Obviously a researcher doesn't operate as a completely blank slate – they have read related work, etc. That's OK – even good! But an inductive approach is about keeping an open mind / prioritizing what shows up in the data and making it fit in a way that works best internally in the dataset, not a way that fits an external framework

coding

is an overloaded term

- **ind** that is more
  cons ngs you find in the
  data, wh re expecting/planning to
  look for th m (**bottom-up**)

If you have a good reason to explicitly do a combination of inductive and deductive coding, that’s fine too, but say as much to the reader (and maybe indicate which are which in the findings)

Obviously a researcher doesn’t operate as a completely blank slate – they have read related work, etc. That’s OK – even good! But an inductive approach is about keeping an open mind / prioritizing what shows up in the data and making it fit in a way that works best internally in the dataset, not a way that fits an external framework

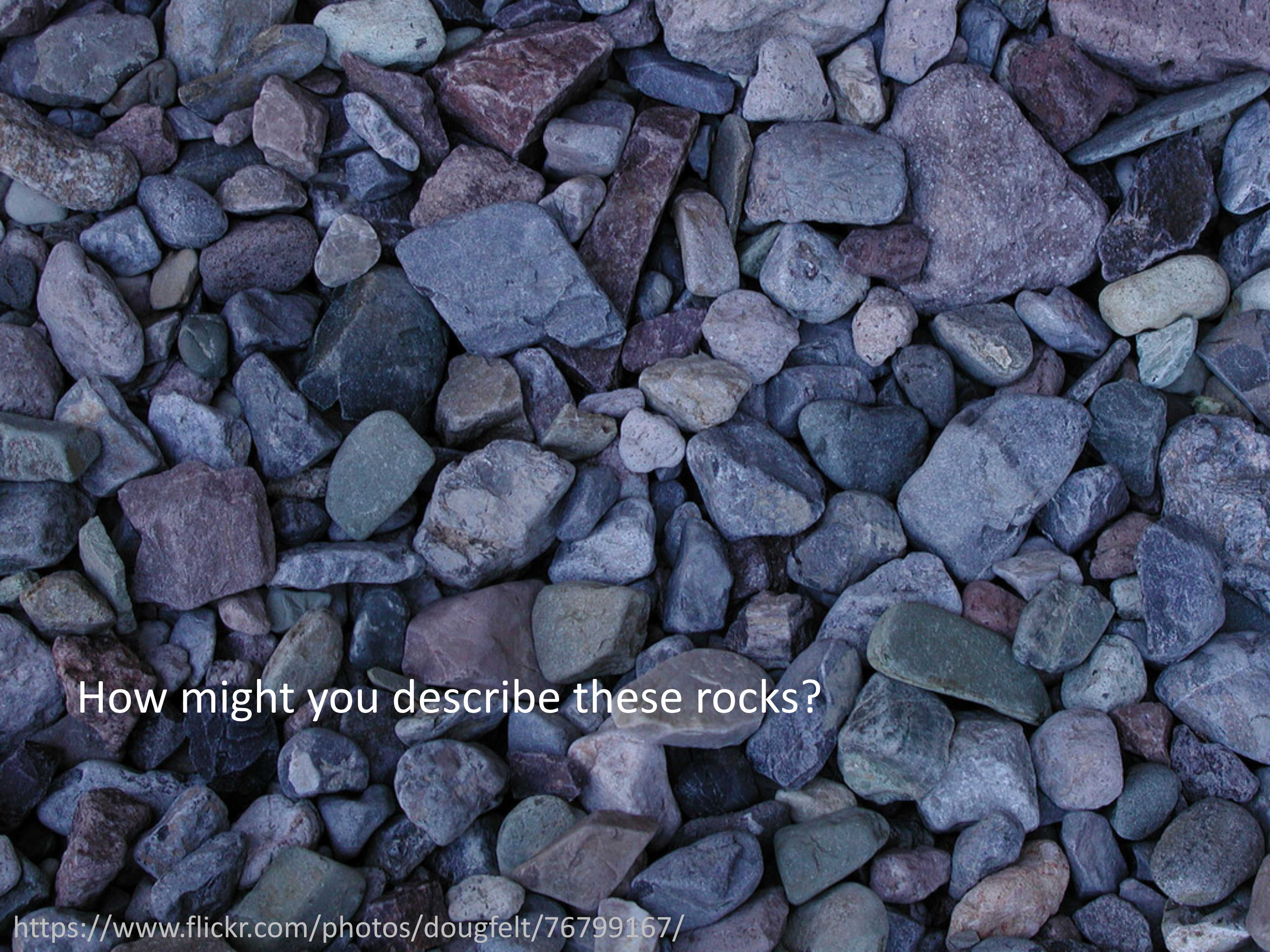
coding

is an overloaded term

- **Open/initial/first-cycle coding:** one of the earlier steps that you take to start to tackle/make sense of the data, which eventually leads to creating the coding rubric/manual

- Coding also might mean the more deterministic (but sometimes still frustratingly interpretative process) of applying the finalized manual to the dataset

So what are example codes?



How might you describe these rocks?

https://www.flickr.com/photos/dougfelt/76799167/

- Size
  - Maximum length?
  - Total volume?
  - Surface area?
- Weight
- Density
- Duration in streambed
- Temperature
- Color
- Shape
- Texture
- Mineral composition
- Vertical height in streambed
- Whether or not it makes a good paperweight
- Flavor (I hope not)

How do you figure out what you should be coding for in initial/open coding?

How do you figure out what you should be coding for in initial/open coding?

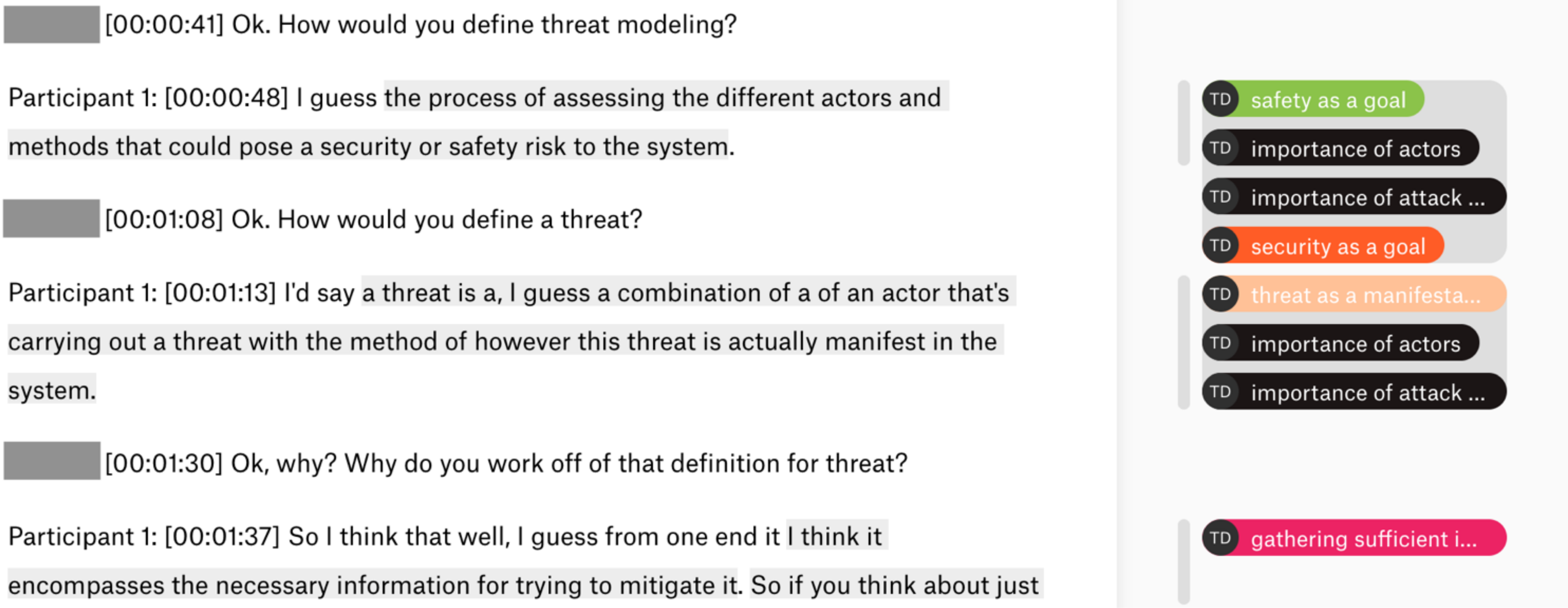
- "What is interesting that is happening in the data (whether or not I was planning to look for it)?"
- "What am I trying to learn about in my research?"
- "How do I want to be able to talk about the data?"

  (If you don't code anything about usage of gendered pronouns you're not going to have that coded data to talk about later….)

How do you figure out what you should be coding for in initial/open coding?

- There is going to be way more information than you’re going to be extracting. You could care about extracting **usage of metaphors**, **intensity of emotions** (with audio transcripts), etc. You are not going to be analyzing everything. Along the same lines, when you present your analysis to readers, it’s by nature an **evidence of presence**. Maybe there was something really interesting that you missed!

One example of first codes applied…



[redacted] [00:00:41] Ok. How would you define threat modeling?

Participant 1: [00:00:48] I guess the process of assessing the different actors and methods that could pose a security or safety risk to the system.

[redacted] [00:01:08] Ok. How would you define a threat?

Participant 1: [00:01:13] I'd say a threat is a, I guess a combination of a of an actor that's carrying out a threat with the method of however this threat is actually manifest in the system.

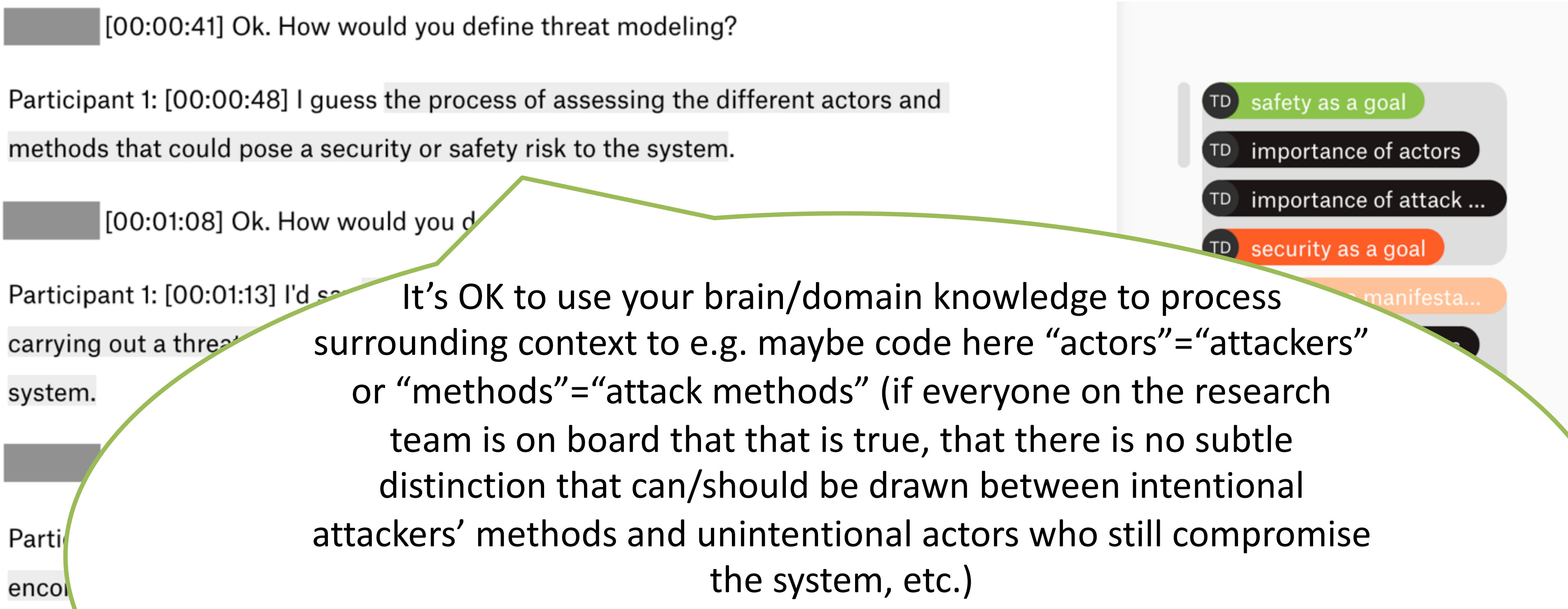
[redacted] [00:01:30] Ok, why? Why do you work off of that definition for threat?

Participant 1: [00:01:37] So I think that well, I guess from one end it I think it encompasses the necessary information for trying to mitigate it. So if you think about just

TD safety as a goal
TD importance of actors
TD importance of attack ...
TD security as a goal
TD threat as a manifesta...
TD importance of actors
TD importance of attack ...
TD gathering sufficient i...

(using cloud.atlasti.com)

One example of first codes applied…



[redacted] [00:00:41] Ok. How would you define threat modeling?

Participant 1: [00:00:48] I guess the process of assessing the different actors and methods that could pose a security or safety risk to the system.

[redacted] [00:01:08] Ok. How would you d

Participant 1: [00:01:13] I'd sa

carrying out a threa

system.

[redacted]

Parti

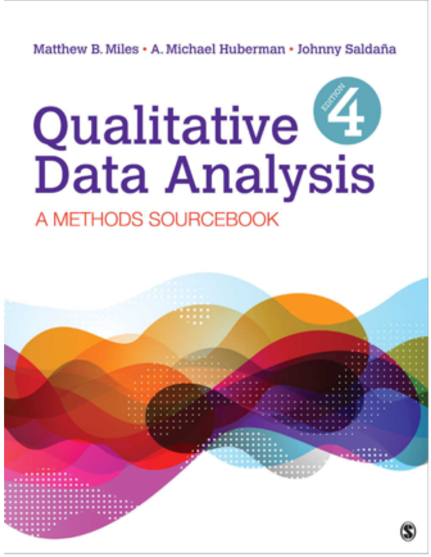
enco

It's OK to use your brain/domain knowledge to process surrounding context to e.g. maybe code here "actors"="attackers" or "methods"="attack methods" (if everyone on the research team is on board that that is true, that there is no subtle distinction that can/should be drawn between intentional attackers' methods and unintentional actors who still compromise the system, etc.)

However, if there is too much ambiguity or arguably multiple interpretations of what someone said, **don't guess and assign meaning that may not be there**. Multiple **researchers working together are incredibly helpful** with this.

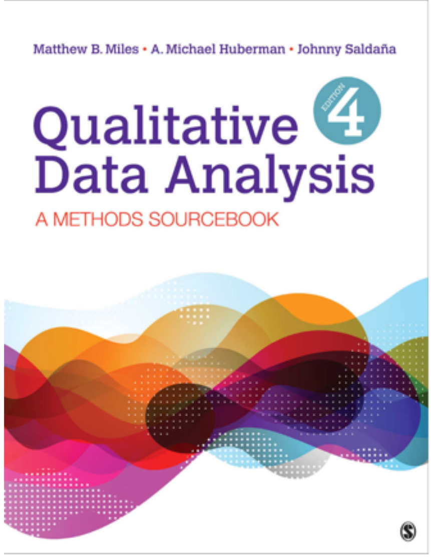
(using cloud.atlasti.com)

Many potential approaches!


Matthew B. Miles • A. Michael Huberman • Johnny Saldaña
Qualitative Data Analysis
EDITION 4
A METHODS SOURCEBOOK

- Descriptive coding
- In vivo coding (in participant's own words)
- Process/"ing" coding ← grounded theory is really into this
- Concept coding
- Emotion coding
- Values coding
- Evaluation coding (e.g., "Residency: Successful")
- Dramaturgical coding
- Holistic coding (i.e. code on large chunks of data at a time)
- Attribute coding
- Magnitude coding
- …

Many potential approaches!



Matthew B. Miles • A. Michael Huberman • Johnny Saldaña

Qualitative Data Analysis

A METHODS SOURCEBOOK

- Descriptive coding
- In vivo coding (in participant’s own words)
- Process/“ing” coding ← grounded theory is really into this
- Concept coding
- Emotion codin
- Values codi
- E

I find this one helpful if you’re falling into a rut of vagueness or excessively literal tagging that doesn’t delve any deeper into content.

“lemons”, “privacy”, etc.

What about “lemons”? “Hating lemons”? “Throwing lemons”? “Using lemons as a metaphor for failure”?

Creating a coding manual is an iterative process

- As you progress through your data (or go back and try new codes on data you’ve been through before), codes will need to be **added**, **merged**, **split**, or **redefined**

Creating a coding manual is an iterative process

- As you progress through your data (or go back and try new codes on data you’ve been through before), codes will need to be **added**, **merged**, **split**, or **redefined**

- There should be a lot of tweaking and arguing/discussing (and you’re supposed to constantly take **notes/memos** about why you choose what you do!)

If a code doesn’t fit well with a piece of data, ask yourself and each other….

- Why isn’t it quite right?
- Look back at data chunks to which you previously assigned this code.
  - How are they similar to the data chunk you’re looking at now? How are they different?
  - Is the ill-fitting code too specific such that it excludes things that seem like they should belong together? Can you articulate why they belong together and use that as a code instead?
  - Is the code so vague that it applies to things that don’t seem like they belong together?
  - Are there maybe 2+ separate, related ideas that sometimes (but not always) appear together? Should you code for them separately?

Computer-Assisted Qualitative Data Analysis Software (CAQDAS) can definitely help this process

Online-access, collaborative versions are surprisingly slow to arrive.

cloud.atlasti.com is recent (June 2018) and still in beta.

It is acceptable (even normal) to start to create the coding manual from collected data while data collection is still underway.

Consider theoretical sampling. Maybe you start to see something new and interesting in your data analysis that suggests that you can't get a full picture unless you also have (doctors, parents, whomever) as participants

https://www.flickr.com/photos/scotiaproductions/7874862018/

Coding is analysis

- It’s an act that describes, coordinates, and/or categorizes the data through the researcher’s lens

How is coding used for *further* analysis?

How is coding used for *further* analysis?

- If you haven’t been, you should try to organize the codes into categories. Is anything meaningful happening with the categories?

Axial Coding (Grounded Theory)

- Creating a new set of codes that explore the relationships between categories:
  - *Causal* conditions
  - *Context* in which it is embedded
  - Action/interactional *strategies* by which it is carried out
  - *Consequences* of those strategies

Thematic Analysis

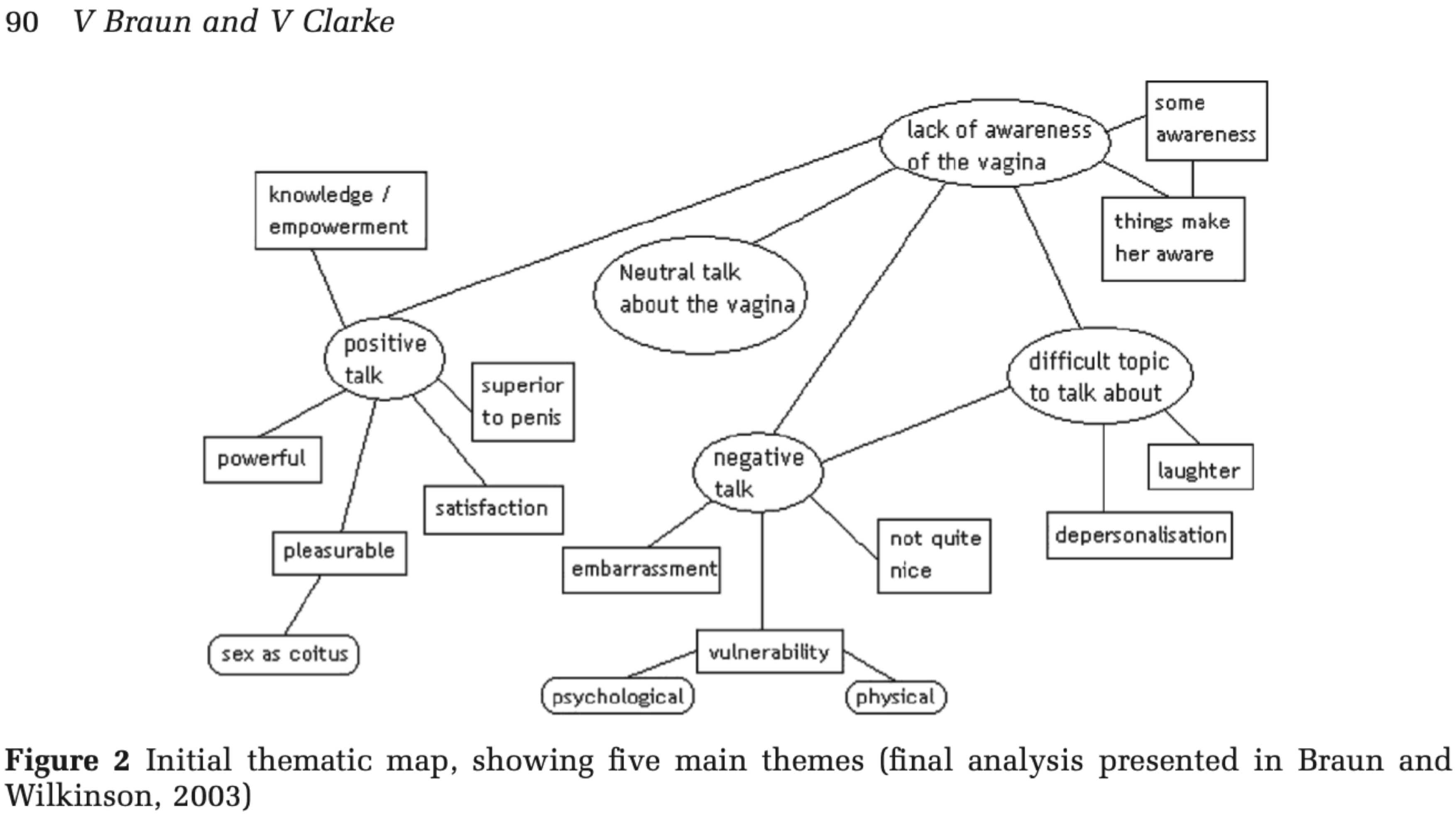
- Are there themes (possibly, but not necessarily equivalent to your code categories) that are supported by the codes and the coded data?

Thematic Analysis

- Are there themes (possibly, but not necessarily equivalent to your code categories) that are supported by the codes and the coded data?

(heads-up: discussion of anatomical body parts on next slides)

Thematic Analysis



90 *V Braun and V Clarke*

Figure 2 Initial thematic map, showing five main themes (final analysis presented in Braun and Wilkinson, 2003)

Thematic Analysis

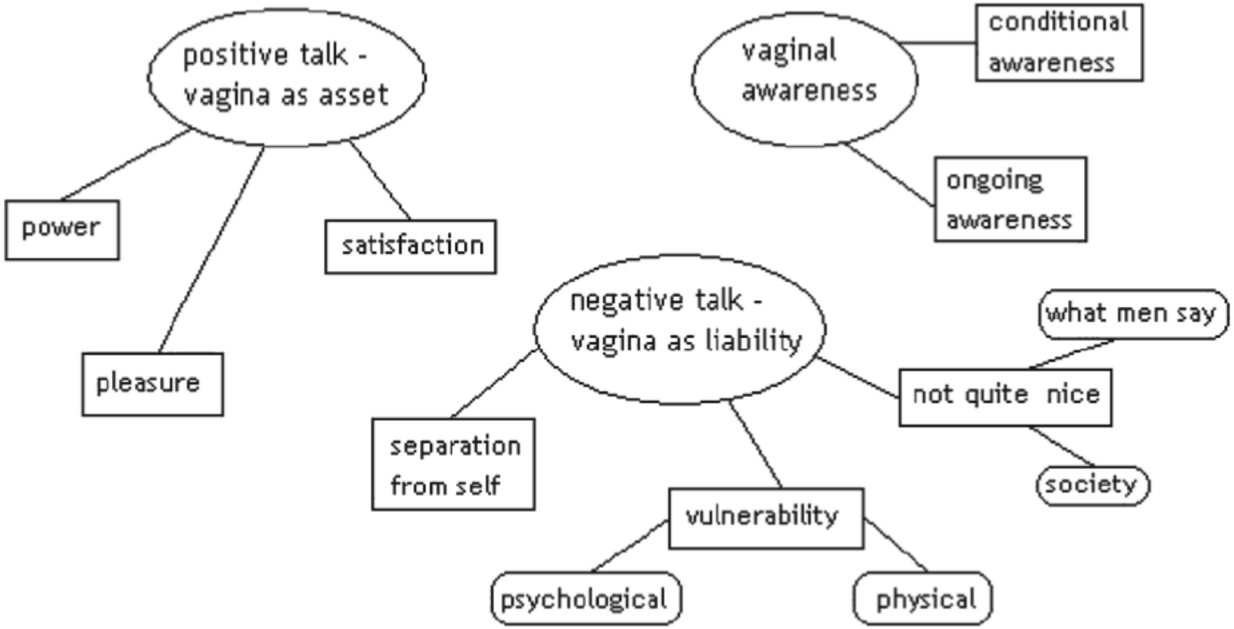
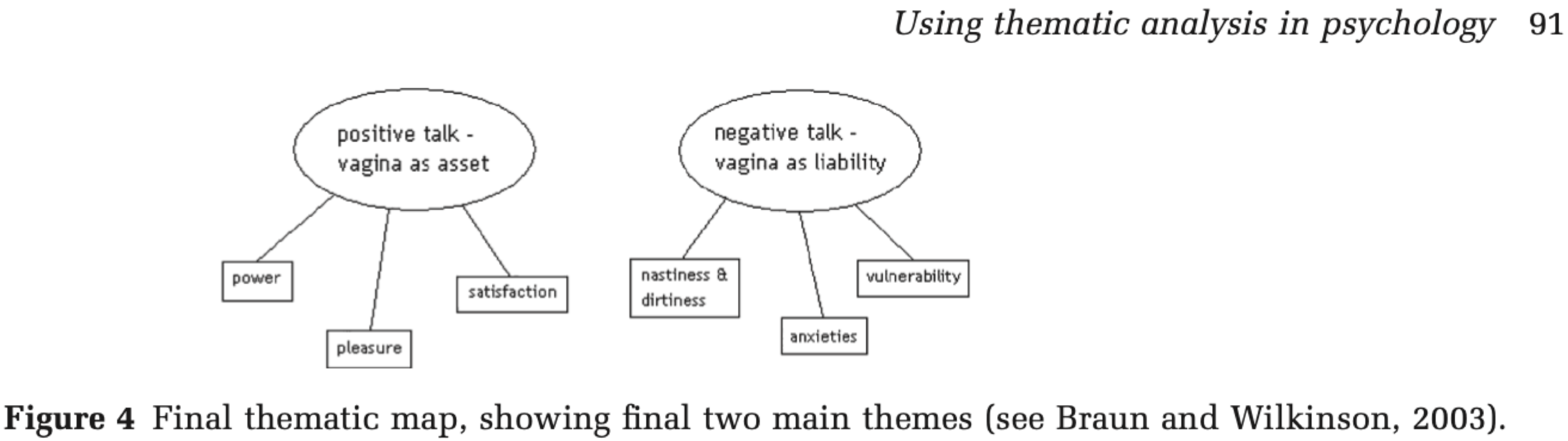


Figure 3 Developed thematic map, showing three main themes (final analysis presented in Braun and Wilkinson, 2003)

Thematic Analysis



Using thematic analysis in psychology 91

Figure 4 Final thematic map, showing final two main themes (see Braun and Wilkinson, 2003).

Code Frequencies/ Code Frequency Tables

- Playing around with patterns in the quantitative representation of the qualitative data may lead you to notice, think through, and look for further evidence of interesting phenomena that were not on your radar

Eventually, in your paper, maybe…

- Themes and new ways of thinking about things people said
- A list of most-salient issues
- A figure modeling necessary preconditions, causes, consequences, sequences of events, etc.
- Frequency counts, e.g.
  - instances that codes appear overall, by participant, location, or other entity
  - instances that unique participants, locations, or other entities raise a code
- etc.

How can we address quality control?

How can we address quality control?

- A lot of quality comes from having more than one researcher involved in the analysis process so that there are multiple viewpoints/interpretations and disagreements are fully talked through until a consensus is reached (for generating codes, for refining codes, for applying final codes)

How can we address quality control?

- A lot of quality comes from having more than one researcher involved in the analysis process so that there are multiple viewpoints/interpretations and disagreements are fully talked through until a consensus is reached (for generating codes, for refining codes, for applying final codes)

- If you memo your whole analysis thought process/decision process and make it available to other researchers as supplementary material

How can we address quality control? (continued)

- If you make your entire codebook (including all definitions) available as supplementary material

- (If you can) if you make your whole dataset and the codes you applied to it (and where) available as supplementary material

…or a more quantitative approach

- **Reliability coding**

Reliability Coding
Can Take Many Forms

- an unrelated researcher, using your coding manual, codes all data as well

What does this help with?
What doesn't this help with?

Reliability Coding
Can Take Many Forms

- an unrelated researcher, using your coding manual, codes all data as well

- different researchers on the team divide up the data and code it individually, but a certain percentage of the data receives overlapping coding

What does this help with?
What doesn’t this help with?

Reliability Coding Can Take Many Forms

- an unrelated researcher, using your coding manual, codes all data as well

- different researchers on the team divide up the data and code it individually, but a certain percentage of the data receives overlapping coding

- a final coding manual is developed using a fraction of the data, and the manual is applied to the remainder of the data

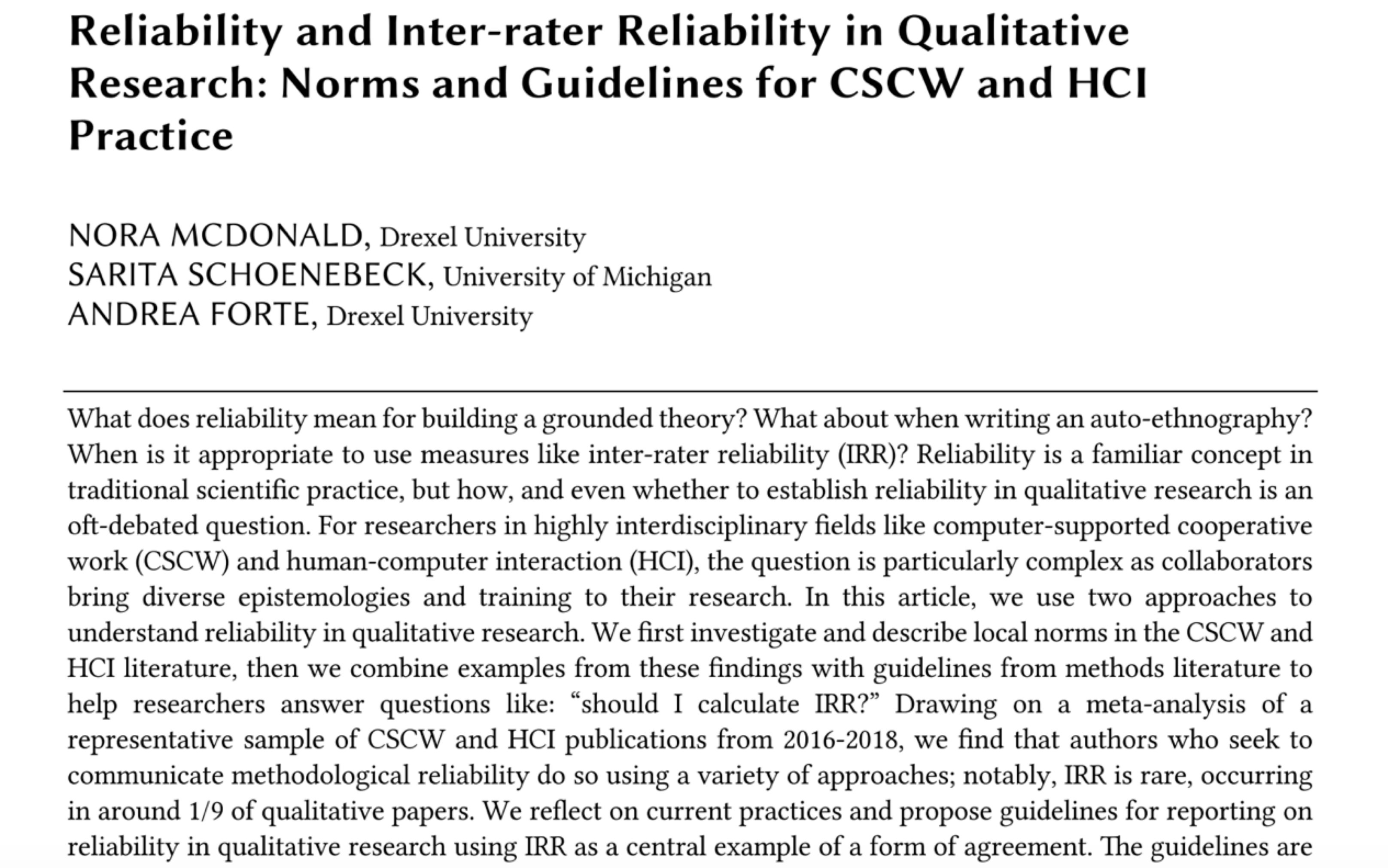
- etc.

What does this help with?
What doesn't this help with?

Inter-rater agreement, e.g.

- **Cohen's kappa**
  - measures the agreement between two raters
  - who each classify N items into C mutually exclusive categories
  - takes into account the possibility of the agreement occurring by chance
- **Krippendorff's alpha**
  - any number of coders
  - each assigning one value to one unit of analysis
  - to any number of values available for coding a variable
  - adjusts itself to small sample sizes of the reliability data
- etc.

Recommended Reading



Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice

NORA MCDONALD, Drexel University
SARITA SCHOENEBECK, University of Michigan
ANDREA FORTE, Drexel University

What does reliability mean for building a grounded theory? What about when writing an auto-ethnography? When is it appropriate to use measures like inter-rater reliability (IRR)? Reliability is a familiar concept in traditional scientific practice, but how, and even whether to establish reliability in qualitative research is an oft-debated question. For researchers in highly interdisciplinary fields like computer-supported cooperative work (CSCW) and human-computer interaction (HCI), the question is particularly complex as collaborators bring diverse epistemologies and training to their research. In this article, we use two approaches to understand reliability in qualitative research. We first investigate and describe local norms in the CSCW and HCI literature, then we combine examples from these findings with guidelines from methods literature to help researchers answer questions like: "should I calculate IRR?" Drawing on a meta-analysis of a representative sample of CSCW and HCI publications from 2016-2018, we find that authors who seek to communicate methodological reliability do so using a variety of approaches; notably, IRR is rare, occurring in around 1/9 of qualitative papers. We reflect on current practices and propose guidelines for reporting on reliability in qualitative research using IRR as a central example of a form of agreement. The guidelines are