Reviewer Critiques (Qualitative Methods) and How to Respond to Them

Author: Jessica Vitak (+ anyone who adds to the document)

About This Document (and a disclaimer)

Reviewing is a highly subjective process. As we all know, each reviewer brings a different background and expertise to their reviews. But it's important to also remember that reviewers are influenced by the amount of time they have to read the paper & complete the review, their current mood, and other factors that may shape what they include (or don't include) in their reviews.

A common problem I have seen over the years is when people have a limited understanding of a specific method (typically qualitative research) and their review includes comments that demonstrate that lack of understanding. You, as the author, need to respond to these comments in such a way that you don't piss off the reviewer while still explaining why their comment is incorrect/a misunderstanding. I started this list based on reviewer comments I kept seeing on my papers, and I think it would be useful to both share this with others and to get others' experiences with how to respond to various methods-focused critiques.

What you should do with this document:

- Read it and add comments/feedback/additional examples.
- Add new critiques and responses based on your experiences.
- Add new critiques that you want people to share their responses to.
- Identify yourself (name/link) when making additions (strongly encouraged).
- Add comments with any questions you have about critiques and responses.

What you shouldn't do with this document:

- Copy example responses verbatim into your response to reviewers – many of these responses have citations, so do the work and engage with them.
- Post non-methods focused critiques (that's outside the scope of this document).

Disclaimer #1: As I mentioned above, reviewing is subjective. The ways I've responded to reviewers have generally worked, but using this approach is no guarantee that your reviewers will also be satisfied. Use this document as a starting point and tailor your response to your paper, methods, and reviewer comments.

Disclaimer #2: I reserve the right to reject or edit any additions to this document.

Questions? Feedback? Email Jessica at jvitak@umd.edu.

List of critiques (jump directly to a critique to read more):

There are other articles and resources out there that address reviewer critiques. I'll link ones I find here (feel free to add others you find!):

- Evaluating Interpretive Research in HCI (ACM Interactions, 2024): https://dl.acm.org/doi/10.1145/3633200
- Telling a story or reporting the facts? Interpretation and description in the qualitative analysis of applied health research data: A documentary analysis of peer review reports (Qualitative Research in Health, 2022): https://www.sciencedirect.com/science/article/pii/S2667321522001287
- Dealing with criticism when publishing qualitative research (book chapter in "Doing Research in Applied Linguistics," 2016): https://www.taylorfrancis.com/chapters/edit/10.4324/9781315389608-21/dealing-criticism-publishing-qualitative-research-xuesong-gao
- H is for Human and How (Not) To Evaluate Qualitative Research in HCI (by Andy Crabtree): https://arxiv.org/abs/2409.01302

Critique: You should quantify your data (e.g., add counts of how many people engaged in X).

From Kumar et al. (2019 CHI): "When reporting our findings, we refrain from using quantitative metrics of how many participants made a given statement because the themes discussed here emerged from more general questions or prompts rather than a narrow accounting where we sought feedback from each participant. Qualitative theorists have provided several arguments against reporting qualitative data numerically, one being that, “[n]umbers can lead to the inference (by either the researcher or the audience) of greater generality for the conclusions than is justified, by slighting the specific context within which this conclusion is drawn” (Maxwell, 2010, p. 479). For this reason, we focus our findings on the themes that emerged across the full dataset rather than a numerical accounting of who said what."

- **Joseph A. Maxwell. 2010.** Using numbers in qualitative research. Qualitative Inquiry 16, 6 (July 2010), 475–482. https://doi.org/10.1177/1077800410364740

From the response to reviewers for Vitak & Kim (2014): “R1 noted it would ‘be useful to have an indication of how frequent each of the cited strategies were.’ We have chosen to not do this, in line with other qualitative studies, largely because the quantification of qualitative findings can be misleading; specifically, if we were to report that 65% of participants engaged in X strategy or 24% of participants referenced having a self-disclosure goal of Y, readers would be naturally inclined to generalize these findings. However, as noted in the Methods section of the paper, the interviews were semi-structured, which means that while there was a general structure to the questions, the interviewer (the first author) did not have a strict set of questions that were asked to each participant.”

From the response to reviewers for Brulé & Bailly (2018): “This study is meant to explore how visually impaired children may use audio-recording technologies during field-trips, and in line with other studies on accessibility. While quantifying the type of actions is useful at this stage to prioritize them in future prototypes, charting the variability and wide range of actions towards learning is more useful as we are trying to delineate a design space for this new family of tools we are proposing.” (edited for concision)

Critique: You should calculate IRR (inter-rater reliability) for your coding.

JV: This depends a lot on the type of analysis you’re doing and the overall research goal, but in general for interview data, IRR is not appropriate to calculate. On the other hand, perhaps you’re using one of the many variations of content analysis to identify themes in tweets. You build a codebook and then you want to code the data for the presence/absence of a given

theme. Then it's probably appropriate to use a second coder and calculate IRR as part of the codebook development.

JV has gotten this a lot when presenting a thematic analysis of interview data and has replied with things along the lines of, "As our goal in this analysis was to identify patterns in the data rather than capture the frequency of certain attitudes/behaviors/themes, calculating IRR was deemed inappropriate."

- Also see this excellent article by McDonald, Schoenebeck, & Forte (2019): Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice: https://dl.acm.org/doi/10.1145/3359174?cid=81100401804
- Can also point to this article: David Armstrong, Ann Gosling, John Weinman, and Theresa Marteau (1997). The place of inter-rater reliability in qualitative research: an empirical study. *Sociology 31,* 3 (1997), 597–606. https://doi.org/10.1177/0038038597031003015
- **Added by Miranda Wei**: Can also point to, and may I recommend as a great read: Janice M. Morse (1997). "Perfectly Healthy, but Dead": The Myth of Inter-Rater Reliability. *Qualitative Health Research*, 7 4 (1997). https://doi.org/10.1177/104973239700700401
- **Added by Eva Hornecker:** This is discussed on pp. 238-242 of Braun & Clarke's Thematic Analysis: A Practical Guide. They distinguish between 'small q' and 'big Q' research, where 'small q' would use things like coding books and inter-rater scores—but they find this an impoverished definition of qualitative research:

  > *"To value a coding process and outcome, wherein different researchers achieve the same outcome (identical coding) using the same measure (codebook), you have to assume that themes are IN the data, waiting to be found by the researcher" (..) " (p.239) ""the idea that coding can be distorted depends not only on a realist idea of a singular truth, but also on idea(l) of 'discovery' of that truth. (...) Furthermore, instead of conceptualising researcher subjectivity - including their pre-existing knowledge of the topic - as valuable, it is viewed as a potential barrier to good coding."*

  Plus:

  > *"The types of codes amenable to use within structured codebook approaches, including measurements of coding agreement are often relatively coarse, superficial or descriptively concrete" (...) richness attained from insight will be lost … there are only so many codes coders can hold in working memory… key problem: "uniformity is valued over depth of insight" "coding reliability procedures also often result in themes that are relatively superficial and underdeveloped" + "one of the things we find most troubling about coding reliability TA is that the values underpinning it are rarely articulated and explicitly acknowledged"* (p. 240)

- **Added by Sohyeon Hwang:** IRR generally works much better with deductive analyses, and it seems worth pointing out that a good chunk of qualitative work leverages inductive

analyses, and that that has distinct strengths. Can point to some of the discussion from grounded theory folks for how inductive coding allows for analysis and flexible development of theory that is rooted in data (e.g., Charmaz). Also, this discussion of how coding evolves and why that can be valuable from a chapter by Muller in “Ways of Knowing in HCI” has been helpful for me: "A code is a descriptor of some aspect of a particular situation (a site, informant or group of informants, episode, conversational turn, action, etc.). When codes are reused across more diverse situations, they gain explanatory power. Each situation becomes a test of the power of the codes to explain an increasing rich set of data. Codes are initially descriptive and tied to particular aspects of the data. Over time, the researcher(s) develop more abstract codes [...]"

- **Added by Laurent Wang**: Sarah Tracy describes when IRR is vs. isn’t appropriate in qualitative research in this paper, which I found to be a clear and helpful explanation to provide to reviewers: Tracy, S. J. (2018). A phronetic iterative approach to data analysis in qualitative research. *Journal of Qualitative Research*, *19*(2), 61–76. https://doi.org/10.22284/qr.2018.19.2.61

Critique: Justify why you used interviews/focus groups vs. other methods.

JV: In our response to this reviewer comment in Lenhart et al. (2023): “We chose semi-structured, virtual focus groups because they allow for a wide variety of perspectives in a setting that allows for immediate follow-up from other participants and facilitators [51]. Focus groups also work well for exploring perceptions and generating ideas [88], which were key goals for this study. In addition, our use of Zoom allowed us to recruit users from across the country. Prior work also suggests that virtual participation—from the comfort of one’s home—can lead to richer conversations [86]; in our case, given we were discussing smart home technologies, this may also have offered a further benefit of reminding participants of their use of SHDs. Finally, virtual participation (versus in-person sessions) may have put some participants at ease and reduced power differentials between participants and researchers [51].”

- Citation [51]: **Richard A. Krueger and Mary Anne Casey. 2014.** Focus Groups: A Practical Guide for Applied Research (5th edition ed.). SAGE Publications, Inc, Los Angeles London New Delhi Singapore Washington DC.
- Citation [86]: **David W. Stewart and Prem Shamdasani. 2017.** Online Focus Groups. Journal of Advertising 46, 1 (January 2017), 48–60. https://doi.org/10.1080/00913367.2016.1252288
- Citation [88]: **Roger A. Straus. 2019.** Mastering focus groups and depth interviews: a practitioner’s guide. Paramount Market Publishing, Rochester, NY.

Critique: Your sample is too small.

JV: There is no “minimum” number of interviews required for qualitative research; rather, researchers should generally focus on establishing saturation (see below for some critiques of this framing). There are multiple forms of saturation; see Saunders et al. (2018) for a detailed discussion of this. If you’re looking to point to a citation to provide some empirical evidence regarding sample size, Kelly Caine (2016) looked at CHI papers published in 2014 and found the average sample size was 12. CHI is a computing-learning venue, so I’d be curious to see if there are evaluations of sample size in other disciplines.

- **Saunders et al. (2018):** Saturation in qualitative research: exploring its conceptualization and operationalization: https://pubmed.ncbi.nlm.nih.gov/29937585/
- **Caine (2016):** Local Standards for Sample Size at CHI: https://doi.org/10.1145/2858036.2858498
- **Sebele-Mpofu (2020):** Saturation controversy in qualitative research: Complexities and underlying assumptions. A literature review: https://www.tandfonline.com/doi/full/10.1080/23311886.2020.1838706

Example of this critique from Sohyeon Hwang: In case it is useful, here's a quote from a CHI review that I got that is directly relevant to this particular critique (and other responses for how to respond seem well-positioned to address it): "The number of interviewees was small. Thus, the authors need to clarify more about what they meant by ‘saturation.’ The data saturated in terms of what? themes? topics under themes? codes? code categories? After reading the Finding section, I suspect that the data had saturated. Please see more below."

- I'm not sure if they had a typo there and but I do think they meant to say "had not saturated." For context, I had 16 interviews and the average length of each interview was 56 minutes. In our response, we also cited Braun & Clarke (2021).

Additional response from Bart Penders: Braun & Clarke argue that data saturation is a concept that is generally coherent for broadly realist, discovery-oriented types of thematic analysis. However, even there, it is very difficult to predict when it will happen and whether it is a requirement to move on. In stark contrast, however, when it comes to reflexive thematic analysis, data saturation is not a particularly useful, or indeed theoretically coherent, concept at all. They write: "meaning is generated through interpretation of, not excavated from, data, and therefore judgements about ‘how many’ data items, and when to stop data collection, are inescapably situated and subjective". In other words: depending on your circumstances, there is no minimum.

- Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative research in sport, exercise and health*, 13(2), 201-216.

Additional response from Miranda Wei: Braun & Clarke also expand on data saturation in a different 2021 article (see pages 15-17 in particular). Though data saturation has been regarded as a “gold standard” for qualitative research, I generally agree with all of the arguments

presented above that data saturation is not always appropriate. The focus on saturation and IRR may reflect the encroachment of quantitative and/or positivist ways of knowing that do not always align with qualitative research. One alternative I have used in responses to reviewers is that the goals of the work were to prioritize meaning sufficiency instead of saturation, in order to "emphasize meaning-richness as key to the validity of the (size of the) dataset."

- Braun, V., & Clarke, V. (2021). Conceptual and Design Thinking for Thematic Analysis. *Qualitative Psychology,* 9(1), 3-26.

Additional response from Eva Hornecker: Braun & Clarke write, "the concept of data saturation …. Is also deeply problematic … We recommend avoiding claims of saturation, instead, we find concepts such as **information power** useful" (Braun & Clarke, 2021, p. 28)

Additional response from Émeline Brulé: it seems to me that the question of saturation and sample sizes is tied to that of sampling, and by extension, what we can learn from interviews. I found this paper useful on that topic: https://link.springer.com/article/10.1057/ajcs.2012.4

Suggested article on sampling from Torkil Clemmensen: Gentles, S. J., Charles, C., Ploeg, J., & McKibbon, K. A. (2015). Sampling in qualitative research: Insights from an overview of the methods literature. The qualitative report, 20(11), 1772-1789. https://nsuworks.nova.edu/tqr/vol20/iss11/5/

Critique: But this is not generalizable!

From Aakash Gautam:

> *I have received this even when I have explicitly stated that our findings are not generalizable. I do not have a good response; I need help. I have tried different ways to make an argument against this. One, which aligns with my research approach, by emphasizing transferrable learning (e.g., the methodological considerations, pragmatic issues). Another attempt I have tried is placing qualifiers in many places (e.g., findings from Nepal, in the context of this particular <setting description>).*

Related example of this from Rebecca Scharlach:

> *Results could be considered not representative because of the lack of quantitative methods.*

From Sunyup Park:

> *How does this study's limitations impact the generalizability of the findings?*

JV: This is an excellent addition to the list and something I have been on the receiving end of frequently, particularly paired with comments about quantifying the data and/or concerns about the sample size. Here are some ways I've responded:

- In response to reviewer comments about generalizability in **Kumar et al. (2017)**, we responded with: “Our goal with this study is not to establish generalizable findings, and the methods we used are not designed for generalizability.”
- In response to reviewer comments about generalizability in **Vitak & Kim (2014)**, we responded with: “Regarding generalizability (R3), we did not intend for the findings to be generalizable. The primary purpose of (most) qualitative research is understanding and describing human behavior, which is why we employed the criterion sampling method rather than a random sampling method, which would have allowed for some degree of generalizability.”
- In the limitations section of **Zimmer et al. (2018),** we wrote: “As with all qualitative studies, our goal was not generalizability to the wider population of fitness tracker users, but to provide a deep and rich accounting of how users managed their device and their PFI [personal fitness information].”

It’s really important to stress that generalizability is not your goal and to connect back to your research questions, which are (presumably) focused on depth of understanding of a particular group/community/set of users. In addition, you may want to check out Daniel (2019), where he proposed the TACT framework (trustworthiness, auditability, credibility, transferability) as an approach to develop “good” qualitative research. In particular, he contrasts transferability with generalizability.

- **Daniel, B. K. (2019).** What constitutes a good qualitative research study? Fundamental dimensions and indicators of rigour in qualitative research: The TACT framework. In *Proceedings of the European conference of research methods for business & management studies* (pp. 101-108). https://doi.org/10.34190/JBRM.17.3.002

[Added by Kwame Porter Robinson and suggested for more designerly and socio-technical work assessed by qualitative methods]

One response I’ve given is to largely point to the importance of knowledge generation, over a specific method, and within the particular context of work. My work tends to involve socio-technical application co-design with ongoing qualitative assessment. In my responses against generalization or “not-scientific-enough” critiques, I’ve included the following:

- Made *Paradigmatic Commitments* as an upfront sub-section in my *Introduction*.
- Note that there are paradigms that span research and practice that are just as crucial to practitioners and designers that do not include scientific generationalization as found in the natural sciences.
    - See: Beach, Dennis. 2016. “On Marxist Critical Ethnography.” Pp. 1–7 in Encyclopedia of Educational Philosophy and Theory, edited by M. Peters. Singapore: Springer.
    - See: Bertelsen, Olav, Susanne Bødker, Eva Eriksson, Eve Hoggan, and Jo Vermeulen. 2018. “Beyond Generalization: Research for the Very Particular.” Interactions 26:34–38. doi: 10.1145/3289425.
- That I take knowledge production—not scientific generalization—as a paradigmatic commitment, aiming for a saturated cumulation of subjective, particular, and local co-design experiences that repeatedly demonstrate how socio-technical agency is

informed, in specific cases ... Here, knowledge is design knowledge, useful not only to academics but to designers as well—where designers include our participants and communities that may not otherwise consider …

- Beyond how knowledge is produced we must consider for what purpose and how it can be applied for "generative, evaluative, inspiration, descriptive, critical, and other concrete purposes" (Höök et al. 2015; Oulasvirta and Hornbæk 2016).
    - Höök, Kristina, Peter Dalsgaard, Stuart Reeves, Jeffrey Bardzell, Jonas Löwgren, Erik Stolterman, and Yvonne Rogers. 2015. Knowledge Production in Interaction Design.
    - Oulasvirta, Antti, and Kasper Hornbæk. 2016. "HCI Research as Problem-Solving | Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems." CHI.
- Taking a reflexive stance like this, refocusing the value of knowledge production itself over methods used to create it, helps reframe work away from techno-solutionism—by centering the particular and place we increase accountability to participants, and that which is particularly useful to them in the here and now (Lindtner, Bardzell, and Bardzell 2016).
    - Lindtner, Silvia, Shaowen Bardzell, and Jeffrey Bardzell. 2016. "Reconstituting the Utopian Vision of Making: HCI After Technosolutionism." Pp. 1390–1402 in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16. New York, NY, USA: Association for Computing Machinery.

Additional Response from Avikshit Pratap: Smith (2018): "Statistical types of generalizability that inform quantitative research are not applicable to judge the value of qualitative research." (pp. 1) Instead, Smith argues that qualitative studies might claim analytical generalization (concept and theory) or intersectional generalizability ("'work that tracks patterns across nations, communities, homes, and bodies to theorize the arteries of oppression and colonialism" pp.5).

- Smith, Brett (2018) 'Generalizability in qualitative research: misunderstandings, opportunities and recommendations for the sport and exercise sciences', *Qualitative Research in Sport, Exercise and Health* 10(1), 137–149. Routledge

Additional response by Émeline Brulé: I've found alternative framings for generalization useful, e.g., analytical/theoretical generalization, transferability. See also https://pubmed.ncbi.nlm.nih.gov/20598692/

Critique: Participants were mostly women.

This critique and response is provided by Katie Siek, who also suggested it could be a subset of the generalizability critique.

Even when we broadly advertise, we typically get more women—from asking researchers about their experiences with fraudulent participants [Panicket et al, 2024] to how older adults would use e-textiles [Jelen et al, 2021].

- Aswati Panicker, Novia Nurain, Zaidat Ibrahim, Chun-Han (Ariel) Wang, Seung Wan Ha, Yuxing Wu, Kay Connelly, Katie A. Siek, and Chia-Fang Chung. 2024. Understanding fraudulence in online qualitative studies: From the researcher's perspective. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3613904.3642732
- Ben Jelen, Olivia K. Richards, Samantha A. Whitman, Tom Ongwere, K. Cassie Kresnye, and Katie A. Siek. 2021. Exploring the Use of Electronics to Customize Pervasive Health Technologies with Older Adult Crafters. In Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '20). Association for Computing Machinery, New York, NY, USA, 166–178. https://doi.org/10.1145/3421937.3421976

Eva Hornecker added that this is an interesting critique given that nobody bats an eye if you have 60-70% male participants in studies.

Critique: You need to add a control group.

From Teja Maddali (hmaddali@umd.edu):

This isn't from a review that I've personally received. But I have seen talk of it in reviewer groups for qualitative papers. For example, when our study is about older adult perspectives on a specific kind of technology and the reviewer says the findings feel generic and could apply to any group without actually comparing with another group (e.g. young adults) as a control.

The suggestions I've seen to address this kind of comment are two-fold. First, we could argue that we are usually focussing on experiences/perspectives of a certain group in our qual analysis. It doesn't make sense to interview people who don't have the experience that we are interested in studying. The usual qual analysis reference/or whatever you cite in your methods section could be used to back this. The second is that it might be valid to compare the experiences of two groups. But it doesn't make sense to consider one group a "control group" since the definition seems heavily intertwined with a quant setting where you are trying to test for the effect of an intervention in an experiment. I've also seen reviewers suggesting that past findings from a literature review where the focus is on another group can be used in such a comparison. So if these kinds of findings for another group are already present in your related works section, you could highlight these without doing a separate study yourself. Feel free to correct me if either of these arguments feel flawed or could be better expressed.

From Qian Wan https://llewynwan.github.io/:

I've received similar feedback that asked for a baseline or confirming quantitative results when submitting to a very tech-heavy venue. I think there are many problems that don't

> *have a baseline condition and can't be quantified (e.g., design for mindfulness). Even if some surveys were validated, administered, and statistically significant results obtained, it might well be the case that participants had fundamentally different understandings of the survey question from each other, or even the researchers. There's also an argument to be made that how can a number or survey question represent the richness of an experience e.g., of mindfulness. Like Einstein said, "Not everything that counts can be counted and not everything that's counted truly counts."*

JV: Thanks for sharing this example Teja! Control groups are particularly important in experimental design, where you want to explore whether some kind of experimental intervention impacts outcome variables you're measuring. You will nearly always be able to measure the impact of the intervention and use statistical tests to assess whether any differences are statistically significant.

This could apply in a study that is relying largely on qualitative data. Perhaps you're working with a local community to determine whether certain types of messaging about vaccines increase likelihood to get a vaccine. You do focus groups and interviews and use your analysis to develop two potential messages. The best next step in your project is probably **not** to go back to those groups and simply ask them which they like better. Instead, you may want to design an experimental study to test the effectiveness of the messages. You could have a control group + two experimental conditions, one for each message. Participants from the community are recruited and randomly assigned to one of the three groups. Some see Message A, some Message B, some no message or something generic ("control"). Then you assess the impact using quantitative measures (e.g., likelihood to get the vaccine), and you compare the findings across the three groups. However, if your RQ in this project is to identify community needs or make policy recommendations, you likely don't need to run an experiment like this.

As you note, comparing two groups is relatively common in qualitative research, but to go back to my description of experimental methods and your astute point, unless you are measuring the effect of an intervention, one should not be treated as a "control." In fact, unless you have an intervention that you are testing, it's impossible to have a control group. Beyond that, there are some important ethical considerations that need to be considered; for example, random assignment is very important in experimental research studies because, in part, of the ethical principle of justice, which argues for fair distribution of costs/benefits to all potential participants.

All of this is to say that **your methodological choices depend fully on your research questions and research goals!** I also think it's important for researchers to familiarize themselves not just with the basics of the methods (how to design an interview protocol, how to analyze data), but also the research paradigms underlying them. This will help you to understand the goals of research through various paradigms, align your study design with that paradigm, and feel more confident in defending your choices – in this case, not to collect data from a control group.

Additional response from Émeline Brulé: I've had a somewhat similar critique on case study

type research (that any case study paper should include several case studies as a point of comparison). It has been useful to reframe it as a theoretical sampling problem—while this sample is adequate for the question being asked. Conversely, I tried to bypass this critique by including and referring to past literature comparing groups (in a study comparing the UK and FR educational settings and its impact on participatory design) but faced critiques by reviewers that because they were not expert on the educational literature, they couldn't evaluate how well it was covered and the differences mentioned should be empirically studied instead. Difficult to address given that positioning by reviewers.

Critique: What if your respondents are lying?

This critique and response was provided by Bart Penders:

Some fields or disciplines radically distrust humans and see them as a source of bias, contamination, and as something to be distrusted. Statements by respondents are often seen as subjective, mistaken, wrong or incomplete. How does one respond to the claim that respondents might be lying—for whatever reason?

I have received these types of comments a few times in very veiled terms, but also very explicitly, in reviews for Penders, B. (2017). The response can be rooted in what Harry Collins (2019) writes in "Forms of Life"—namely that researchers usually cannot detect if respondents are lying. Especially so, if they lie really well. Does this then disqualify all empirical material ever collected, because it might be corrupted through lying respondents? Collins argues 'no'. His argument is that to lie well, the lie has to make sense. It has to fit the frame of reference that others convey too. If the researcher is sufficiently immersed in that frame, a good lie cannot change that frame. Collins refers to this as the "antiforensic principle"—the cultural sociologist or anthropologist is interested in the meaning as it is given and as it travels in cultures, not in single events. Lies harm our views/frames of single events, **but not** those of cultures. In other words, since we are usually not after generalisability in the empirical sense (see above), the risk this poses is minimal, even when it occurs.

- Collins, H. (2019). *Forms of life: The method and meaning of sociology*. MIT Press.

At the time of the actual review response, Collins' work was not yet published, and we had to make due with a response that was not rooted in literature. We wrote: "As part of this study, we build relationships with respondents. Part of those relationships are formal, allowing them anonymity and the right to withdraw at any time. This reduces any need they may feel to be dishonest. The other part of those relationships are informal, and we try to convey deeply why we do this study and how their contribution is going to help us, regardless of the content of their answers. We cannot know for certain whether respondents have lied or attempted to deceive us, but it is unlikely that all of them did, which would be the alternative explanation for the internal consistency of their narratives." [edited for brevity]

Additional response from Émeline Brulé: This is a little tangential, but I'm thinking about the controversy over Margaret Mead's being lied to by her Samoan interviewees. It seems to me that there's been media coverage of lies in qualitative research, which underpins this critique (note: the consensus is, broadly, that Mead didn't get lied to). Triangulation and consistency with past literature helps address this critique, although the fact that we are also asked to generate new and 'surprising' insights can be at odds with that.

Critique: Don't include "discussion" material in the findings section.

From Eva Hornecker: This is a tricky one and I think it is discipline-specific. In quantitative work, this is the rule to not go into any discussion, but there it's a different type of data. It also depends on what is meant with 'discussion' - quantitative reviewers often interpret any analysis/interpretation as 'discussion', since for them, the interpretation of findings belongs into discussion, and findings are the pure fact (but for qualitative. work, analysis is akin to quant research doing the stats and outputting numbers).

Braun & Clarke book on TA (p.131): Analysis section "combines data extracts and analytic narrative - the things you write, giving the reader your interpretation of the data and their meaning." "a more qualitative reporting model here, where the 'results' and 'discussion' section are combined, so your analytic narrative contains connections with, and develops the analytic points in relation to, other literature…"

Braun & Clarke, pp 132-133, they discuss the 'traditional model of report' that separates the 2, which also describes the disadvantages of separating findings and discussion: "your manuscript will likely be a bit repetitive". "An integrated approach works particularly well when strong connections exist with existing research, and when the analysis is more theoretical or interpretative. Combination can also avoid repetition"

I don't have access to most of my books (office moving), but I'm pretty sure that Heath and Luff in their video interaction analysis book recommend a similar approach of having a vignette and then interpreting it (thus analysis/discussion of the incident and connecting it to more abstract terms), but they also still will have a full discussion section.

A good way to argue could be that the parts of discussion that help to avoid repetition are integrated in the 'findings' sections, and where the discussion directly ties to specific aspects of findings. And that the actual discussion section then brings it all together and goes one step further.

Critique: Justify use of a particular theoretical framing.

This critique and description was provided by Émeline Brulé:
It's happened to me a couple times to be asked about using a certain theoretical framing instead of an inductive, grounded theory-like approach. I've always justified it by "study aims to contribute to theory in X, Y, Z ways", but I wonder if others have faced it, and if so how they've gone about addressing it.

JV: I had a variation on this happen recently! I come from a social science (communication) background, where it is standard practice to structure a study around a theory, so I have published many articles that are motivated by a particular theory. We recently had a paper published in CSCW that evolved significantly over several years and ended up not being a standard interview study but rather a roadmap for qualitative researchers who want to use a particular theory (Nissenbaum's contextual integrity or CI). I was a bit baffled to see a reviewer write that we needed to "identify the suitability of CI for this work and better situate the CI framework among other privacy theories (using a comparative discussion) in the context of this work."

Basically, the reviewer felt the best way to "prove" that our chosen theory was "best" was to include comparisons to all other major privacy theories, which is not the goal of the paper, nor does doing so serve to help achieve the goal of the paper. In our response, we reiterated what the goal of the paper was, then added this:

> *We recognize that CI is one of many privacy theories, and we cite existing work that discusses other approaches (e.g., added a bunch of cites). However, the question of which privacy theory a researcher should select is a separate question from the one our paper addresses and is out of scope for this paper. Rather, our aim is to provide guidance for researchers who have opted to use CI as their privacy framework.*

I personally believe that most of the research that frames itself as grounded theory is not, in fact, grounded theory. Grounded theory is a form of pure inductive research where the researcher goes in without pre-existing knowledge or conceptions of what is happening and focuses on letting the data tell the emerging story. That is rarely the case – we usually have prior work/theories that can help guide us. So if you're being asked why you use a theory vs. inductive approach, think about the RQ and the population and the phenomenon of interest. What can be gained by going in assuming we know nothing about this? How might starting from existing theory/knowledge and building on that (in your study design, protocol, analysis, etc.) potentially benefit or bias your research?

Additional response from Qian Wan:
I share the same sentiment with Jessica here, that most Grounded Theory are not indeed "grounded" *tabula rasa*. Empirical data are inherently "theory-laden" [source]. Even a grounded theory-like, inductive approach implies an epistemic judgment that this method arguably produces better answers to a research question than existing theories if any. I really doubt if any

studies, even quantitative or natural scientific experiments, can be done without presupposing any theories. In HCI research, theories are often embedded in the null hypothesis, survey questions, or even research questions.

Critique: This is not objective. This is not scientific. This is just anecdotal.

This critique and description was provided by Qian Wan: The first critique comes from a colleague, the second was brought up in an oral defense of a PhD thesis, and the third comes from a reviewer. I found not a few colleagues that said this. The criticism of other methodologies as "not-scientific" presupposes that science (most likely natural sciences in English) is a better or more reliable source of knowledge. It often falsely treats scientific knowledge as the only "transcendental" or "objective truth", and therefore begs the question against other types of knowledge such as philosophy, humanities, and social science.

It's kind of scientism, and all counter-arguments against scientism therefore apply here:

- https://blog.apaonline.org/2018/01/25/the-problem-with-scientism/

A simple counter-argument would be: the statement "being scientific is conducive to more reliable knowledge" itself is a value judgment, the truth of which cannot be determined within the scope of science. Therefore scientism is self-defeating.

A challenge of the traditional "view-from-nowhere" objectivity can also be found in the discourse of feminist HCI.

- Towards a Feminist HCI Methodology: Social Science, Feminism, and HCI https://dl.acm.org/doi/pdf/10.1145/1978942.1979041

Put simply, the feminist epistemology fundamentally rejects the traditional definition, and even desirability of "objectivity" in academic research. Knowledge seeking is value-laden, and a "transcendental", view-from-nowhere knowledge is unattainable (except by God). https://plato.stanford.edu/entries/scientific-objectivity/

I don't know how to respond to the third yet. It argues against the validity of qualitative research altogether. I don't think "anecdotal" ought to be used as pejorative. It is just a different way of knowing.

Critique: Your interview length was too short! You should have talked longer.

This critique comes from Julie Kientz, and I admit I (JV) have never seen such a critique before! I think the underlying concern is probably similar to what we see in critiques about the number of interviews – reviewers wonder if you can reach saturation if you only talk to X number of people or if your interviews "aren't long enough." But this is a silly critique in my opinion. The length of the interview will depend in large part based on how many questions you ask of participants. If I have one focused topic/set of questions and I only want to focus on a specific experience/attitude/belief, my interviews may only be 15-20 minutes. More normally, I have multiple topics I want to cover, which generally means more questions and longer interviews. We need to move away from this problematic attitude that length is correlated with quality/validity/rigor.

So how to respond to a reviewer offering this critique? I have looked around a bit and only seen very generic comments in qualitative methodology books (e.g., "interviews typically last 30 minutes to more than an hour"). I hesitate to suggest authors try to quantify the data they do have, as that is usually a bad idea. Instead, I would reiterate how you determined you reached saturation, which is the metric through which we decide when to stop collecting data. You could also provide a detailed response on how you built your protocol and (presumably) that you obtained rich anecdotal descriptions from participants related to your questions.

From Jasmine Linabary: Gist-Mackey and Kingsford in their article on linguistic inclusion state, "As scholars, we are often trained that verbose, rich interview data is ideal for all qualitative organizational communication research, even as these standards implicitly privilege white-collar communication norms. In this study, scholars are called to reimagine the most commonly used method of qualitative data collection, interviews, in order to address implicit classed communication biases in qualitative organizational communication scholarship." They provide an argument based on class that can be cited in response to questions about interview length.

- Gist-Mackey, A. N., & Kingsford, A. N. (2020). Linguistic inclusion: Challenging implicit classed communication bias in interview methods. *Management Communication Quarterly*, *34*(3), 402-425. https://doi.org/10.1177/08933189209341

Critique: Why did you conduct only qualitative analysis? Discuss this limitation!

This critique was provided by Hyungjun Cho. I think there's two ways to interpret this (given the brief description) and both ways of interpreting it are problematic in my opinion. If the reviewer is implying the researchers should have also conducted quantitative analysis on their interview data, I would first look at the above section on quantifying qualitative data, as there are very clear and strong reasons why this is generally bad practice.

The alternative interpretation is that the reviewer thinks the person shouldn't be limiting their data collection to just qualitative methods. This speaks to the benefits of triangulation, which may include qualitative and quantitative data collections. Triangulation is usually a good thing, as it allows you to get different perspectives or take different approaches to answer your research question. But it's not **necessary** to triangulate a qualitative-driven study. Remember, you should always pick the methods that make the most sense for your research question, and it's good practice to have a clear rationale for your methodological choices, either directly stated in your paper or provided in response to critiques like this.

In terms of limitations, I honestly cannot think of any limitations to only conducting a qualitative analysis of qualitative data because there are significant problems when trying to run quantitative analyses of a small sample that has been purposefully selected. If your reviewer wants you to explicitly state that the goal of your study is not to generate generalizable knowledge, then you can do that, although I'd probably keep it to my response letter and point to some citations that highlight what the goals and outcomes are for qualitative research and how that differs from quantitative research. For example, in "What Constitutes Good Qualitative Research," Daniel (2019) describes how qualitative research can have a goal of *transferability*, which is a bit different a concept from generalizability.

[add new critiques here]