ARTICLE

Bias and Debiasing Strategies in Qualitative Data Collection

Julie Zahle

Department of Philosophy, University of Bergen, Sydnesplassen 12-13, 5007 Bergen, Norway
Email: julie.zahle@uib.no

Abstract

A widespread view has it that qualitative research in the social sciences is of poor quality because it lacks effective debiasing methods. In this paper, I zoom in on researcher bias in qualitative data collection. First, I provide a brief outline of qualitative data collection and develop a suitable account of bias. On that basis, I argue that qualitative researchers may mitigate researcher bias through the combined use of two strategies that capitalize on main characteristics of qualitative data collection. Lastly, I defend the two debiasing strategies against possible objections to their effectiveness.

1. Introduction

Bias in qualitative research may come about in a variety of ways. For instance, a textbook on qualitative research notes that "data bias may result from a somewhat unconscious subjective selection process. Researchers are tempted to talk primarily to people they like or find politically sympathetic [. . .] Or it may be that researchers talk to a variety of people, but overidentify with one group. They then hear what this group has to tell them, but less fully what other groups tell them" (Glesne and Peshkin 1992, 99). A different textbook points out that bias may occur in that qualitative researchers may "choose informants who are sympatico with their worldview, may ask leading questions to get the answers they want, or may ignore data that do not support their conclusions. Emotional pitfalls can also contribute to researcher bias" (Padgett 2008, 184).

These observations naturally raise the question of how, if at all, bias in qualitative research may be mitigated. A widespread view has it that there is not much to be done: effective debiasing methods are lacking and, as a result, qualitative research tends to be of poor quality; for example, Galdas (2017) and Montuschi (2014, 135) report this view. The position may be illustrated by a recent editorial by social scientist Walter Schumm (2021). He contends that confirmation bias, i.e., social scientists' tendency to seek confirmation of their positions and prejudices when

conducting research, is a main problem in the social sciences. To solve it, he maintains, the use of qualitative methods should be relegated to a handmaiden's role (Schumm 2021, 291). The methods are not up to the task of testing and establishing scientific claims in an unbiased fashion (only quantitative methods are).

Within philosophy of science, the stance that there are no effective means to mitigate bias in qualitative research has so far not been systematically examined. In this paper, I take a first step in filling this lacuna. I confine myself to examining bias in qualitative data collection and, in line with most discussions of this issue, I focus on researcher bias, that is, bias which researchers introduce into the research process (see, e.g., Padget 2008, 184; Maxwell 2013, 124; Lietz and Zayas 2010, 192). I defend the claim that qualitative researchers may effectively avert researcher bias through the combined use of two debiasing strategies.

Here is how I proceed. I begin by providing a brief introduction to qualitative data collection, including some of its main characteristics (section 2). Next, I present an account of researcher bias that I refer to as the qual-c (qualitative data collection) account (section 3). It diverges from other conceptions by being more precise as to what bias in qualitative data gathering amounts to. Equipped with the qual-c notion of bias, I go on to show that, capitalizing on the main characteristics of qualitative data collection, researchers may use two effective debiasing strategies (section 4). The qual-c strategies, as I call them, go beyond existing proposals for mitigating bias in qualitative research. Further, I examine—and reject—possible objections to the effectiveness of the strategies (section 5). On this basis, I conclude that researcher bias in qualitative data generation may be mitigated: it does not undermine the quality of qualitative research. I end by briefly summarizing my findings (section 6).

Before embarking on this task, though, one more comment is in order. As I use the term, "bias" refers to a negative feature of research: biased research is of poor quality. I think this is most in keeping with standard usage of the term (see, e.g., Crasnow 2017, 193). However, nothing substantial hinges on this terminological choice. In philosophy and the social sciences, "bias" is sometimes used differently to denote a feature of research that need not be negative. Accordingly, a distinction is commonly drawn between positive and negative bias (see, e.g., Antony 2018, 115). From this perspective, I am concerned with negative bias and how to mitigate negative bias in qualitative data collection, and those who prefer may simply rephrase the subsequent discussion in these terms.

2. Qualitative data collection

Qualitative methods of data collection include participant observation, qualitative interviewing, focus group interviewing, and document searches. Among them, participant observation and qualitative interviewing are likely the most frequently used (Bryman 2012, 349). In this paper, I focus exclusively on these two methods as used either on their own or in tandem in a single study. Thus, I have the two methods and their data in mind when I henceforth talk about qualitative methods of data collection and qualitative data.[1]

---

[1] The subsequent characterization of participant observation and qualitative interviewing draws on introductions to the methods I have offered in other papers. See, e.g., Zahle (2018, 2021).

The method of participant observation requires the researcher to take part in the research participants' lives in their own surroundings. The extent of the researcher's participation may range from staying passively in the background to engaging actively in the research participants' activities. Whatever her degree of participation, the researcher should at the same time observe, i.e. note, what goes on. The method should be applied over an extended period of time. Previously in anthropology, for example, researchers were expected to carry out participant observation for at least a year. Today, studies of a much shorter duration are regarded as acceptable.

The method of qualitative interviewing has it that the researcher poses questions to a research participant. In addition, the researcher should permit or encourage the interviewee to digress, expand on her views, exemplify her points, introduce her own concerns, and the like. The interview typically takes place in a setting familiar to the interviewee, who is interviewed, for up to several hours, on one or several occasions.

The two methods may be further characterized by noting several features that they (or their typical use) have in common. One is that they are flexible in the sense that most decisions about data collection are usually made on the go rather than prior to the onset of data collection (see, e.g., Hammersley 2013, 12; Bryman 2012, 403–4, 470). That is, the participant observer may have contacted some people, made arrangements to get access to certain settings, and the like, before she starts generating her data. Still, most of her decisions about what data to gather, including what situations to observe and what to focus on in those situations, will be made during the data collection phase. Similarly, the interviewer may have made some appointments with interviewees, just as she may have prepared and sometimes tested a list of questions to be covered during her interviews. Yet, once she is in the field, she makes most of her decisions about whom to seek out for interviews and exactly how to proceed during each interview. Both in regard to participant observation and qualitative interviewing, the more decisions the researcher makes during data collection, the more flexibly she uses the methods.

Another characteristic of the two methods is that the researcher has little, if any, control over the research environment, the research participants included. The participant observer should go along with the research participants and interfere minimally in how people go about their business. Hence, she should make no attempt to affect the research participants' activities and settings. Likewise, the qualitative interviewer should make the interviewees formulate their own answers, just as she should, to some extent at least, allow them to determine how the interview develops.

The upshot of the application of participant observation and qualitative interviewing is large sets of heterogeneous data. Field data (or field notes) are based on the participating researcher's observations and consist in her descriptions of the research participants, what they did and said, their equipment, the setting, and the like. Interview data (or interview notes) detail what was said during the interview, the interviewee's nonverbal behavior, where the interview took place, and the like. In both cases, the data should be in terms that come as close as possible to those of the research participants and their perspectives (see, e.g., DeWalt and DeWalt 2011, 165–66). Most notably, this is exemplified by verbatim transcriptions of recorded interviews. Thus, field and interview data are heterogenous: they are descriptions of multiple different aspects of the research participants' verbal and nonverbal behavior, their settings, equipment, etc. Moreover, insofar as the data are rather

detailed reports of what transpired during data collection, they add up to relatively large data sets.

3. Bias in qualitative data collection

When data are gathered by way of qualitative methods, researchers may bias the process. In the following, I first present an account, the qual-c account, of researcher bias in qualitative data collection. Then I motivate and expand on the conception by situating it vis-à-vis current discussions of bias in philosophy of science and the social sciences.

According to the qual-c account, bias in qualitative data collection occurs when researchers' commitments influence the data collection such that their data set ends up lacking one or several good-making features.

Unpacking the account, the first part states that, in biased data collection, researchers' data generation is affected by their commitments. The latter include researchers' nonepistemic (moral, political, etc.) values, assumptions, expectations, experiences, feelings, and the like.[2] Since researchers may, or may not, be conscious of their commitments and their impact on the data generation, the account leaves this open.

The second part of the qual-c account specifies under what circumstances the impact of researchers' commitments amounts to bias: this is the case when the collection of data results in a data set that lacks one or several good-making features, i.e., characteristics of a high-quality data set. It is only insofar as a data set has these features that it is fit to serve as evidence base for a true and complete, i.e. nonpartial, answer to the research question under study.[3,4] To render the discussion manageable, I concentrate on three good-making features, namely descriptive adequacy, balance, and sufficiency.[5] They are most directly and obviously in line to be affected by researchers' commitments via the data collection process. I now explicate the three features in turn.

Researchers' commitments may influence their data collection, with the result that some of their field or interview data lack descriptive adequacy, that is, fail

[2] It should be stressed that commitments are meant to encompass factors that cause cognitive bias. Note too that in discussions of qualitative research, researchers' commitments are often referred to as their subjectivity (see, e.g., Peshkin 1988; Roulston and Shelton 2015).

[3] Here I follow Elisabeth Anderson (1995, 37ff) in holding that answers to research questions should be true and complete.

[4] As social scientist Martyn Hammersley also observes, many qualitative researchers are opposed to talking about qualitative analyses being true (Hammersley 1993, 69). The resistance is often motivated by holding that when an answer to a research question is said to be true, it is variously claimed to represent reality as it really is, to be infallible, to correspond isomorphically to reality; to be the only possible answer, and the like (see, e.g., Hammersley 1993; Lincoln and Guba 1985, 294). I make no such assumptions. It should also be noted that, in the qualitative research literature, it is commonly proposed to evaluate answers to research questions in terms of their validity or credibility instead. Both valid and credible accounts are often characterized as accounts that adequately or correctly represent social reality—see, e.g., Hammersley (1993, 69), and Lincoln and Guba (1985, 294) and Bryman (2012, 389) respectively. Thus specified, those who prefer may substitute the term "true" with "valid" or "credible" in the subsequent discussion.

[5] The following account of the three features is based on Zahle (2021). Here, further good-making features of qualitative data sets are discussed too. See also, e.g., Zahle (2018, 2019).

correctly to describe the research participants' verbal and nonverbal behavior, equipment, surroundings, and the like. This should not be taken to suggest that there is always a *single* correct description of what transpired during a session of data collection. Rather, a range of descriptions that correctly detail what happened in words approximating the research participants' and their perspectives may typically be offered. Obviously, if pieces of data fall outside this range and so misdescribe aspects of what transpired, they should not be used in researchers' subsequent analyses. By way of illustration, consider a researcher who conducts a qualitative study on how local politicians in a Danish municipality run their election campaign. As part of it, the researcher carries out participant observation at a political meeting where one of the local politicians is a main speaker. Because she strongly disagrees with the politician, she misperceives several of the politician's points and their reception by the audience. Thus, when the researcher later writes up her field notes, she misdescribes the event in these respects. In this scenario, the data generation was biased resulting in descriptively *in*adequate field data.

Biased data collection may also consist in researchers' commitments causing them to collect qualitative data that fail to add up to a balanced data set. A data set is balanced insofar as its data describe all the (types of) people, activities, settings, viewpoints, and the like that are relevant to providing a true complete answer to the research question. Keeping this in mind, consider a researcher who conducts qualitative interviews to find out how the professors at a college evaluate a new study program. The researcher is very favorably inclined towards the study program and her attitude influences her data collection so that she ends up lacking interview data expressing dissatisfaction with the new study program. In short, the researcher's commitment (her enthusiasm about the new program) biases her data collection in that she ends up with a data set that lacks balance in one important respect: it covers only what the professors like about the new study program, not what they dislike. Accordingly, the researcher is unable to offer a complete (and true) answer to her research question—one that gives the full picture of the professors' reception of the new study program.

Lastly, researchers' commitments may cause them to collect too few field or interview data in which case their data set lacks size-wise sufficiency. A qualitative data set is sufficiently large insofar as it contains enough data about all the relevant (types of) people, activities, settings, viewpoints, etc. to provide a true complete answer to the research question. There is no implication that more data is always better. The point is merely that researchers must gather enough qualitative data to establish a true and complete final analysis. As an example of how researchers' commitments may cause them to gather too few data, imagine a researcher who uses participant observation and qualitative interviewing to determine what interest the students in a high school take in the surrounding society. The researcher strongly expects the young people to lack an interest in politics, one aspect of their surrounding society. In consequence, she wrongly takes this to be confirmed by interviews with only two students. In this situation, the researcher's expectation gives rise to biased data collection resulting in too few data about the students' interest in politics.

The discussion so far has laid out the basics of the qual-c account. Since the account is concerned with bias as a feature of the research process, it exemplifies a

process conception of bias. I now motivate and elaborate on the qual-c account by putting it in the context of other process notions advanced in current philosophy of science and social science discussions. I begin by presenting three ways in which process conceptions of bias may diverge, while noting where the qual-c account stands on these issues.

Process notions of bias have varying scope, that is, they differ in terms of the research and parts of the research process they purport to cover. The most inclusive notions aim to characterize bias in all types of research and stages of the research process (see, e.g., Douglas and Elliott 2022, 202). Other accounts have a more restricted scope. For instance, Hudson characterizes bias in "(statistical) reasoning" (Hudson 2021, 399), whereas Stegenga defines bias in medical research (Stegenga 2018, 153). Clearly, the qual-c account is a process notion with a very restricted scope: it focuses solely on bias in connection with qualitative data collection in social research.

Process conceptions also diverge with respect to their comprehensiveness, as it may be called. Very comprehensive notions purport to characterize *all* forms of bias that may occur in those types of research or parts of the research process that fall within their scope. Instead, less comprehensive ones concentrate on the characterization of *one* or *several* forms of bias. For instance, the view that bias transpires when researchers' nonepistemic values affect the research process has often been put forward as a very comprehensive notion—one that provides an exhaustive account of bias in the research process. By comparison, confirmation bias is typically presented as just one type of bias. According to it, researchers' beliefs, theories, etc. influence the research process by making them look for evidence, or analyze evidence in a way, that confirms these beliefs, theories, etc. (see, e.g., May 2021, 3349; Steel 2018, 897). Returning to the qual-c account, it belongs among comprehensive notions in that it aims to cover all forms of researcher bias in qualitative data collection.

Not surprisingly, process conceptions also differ as to what they take to be constitutive of the occurrence of bias. In this connection, it is informative to distinguish between three families of process notions: source, outcome, and mixed notions, as I shall call them.[6] Source conceptions focus on the origin of bias and hence they take it that bias transpires when certain factors influence the research process. Moreover, most source notions are concerned with researchers as the source of this influence. They are illustrated by the view that bias occurs when researchers' nonepistemic values affect the research process (see, e.g., Kincaid et al. 2007, 5). In addition to nonepistemic values, other influencing factors that are commonly mentioned include researchers' background, identity, and experience (see, e.g., Maxwell 2013, 44ff) and their prior beliefs, assumptions, and expectations (see, e.g., Antony 2018, 114).

Outcome conceptions approach bias from the opposite direction, so to speak. Rather than holding that bias occurs when certain influences impinge on the research process, they maintain that bias transpires when the research process has a certain upshot. Typically, this idea is spelled out by saying that the research process is biased when it (systematically) undermines the truth or accuracy of the research results (see, e.g., Douglas and Elliott 2022, 202; Stegenga 2018, 153).

---

[6] This classification is not meant to be exhaustive. For instance, it does not cover Anke Bueter's epistemic notion of bias (Bueter 2022).

Mixed conceptions are combined source and outcome notions. They maintain that bias transpires when certain influences impinge on the research process such that a certain outcome follows. The social scientists Martyn Hammersley and Roger Gomm mention a notion of bias that illustrates this approach, namely the view that bias "is a tendency on the part of researchers to collect data and/or interpret them, in such a way as to favor false results that are in line with their pre-judgements and political or practical commitments" (Hammersley and Gomm 2000, 152). Another example is Hudson's account which states that a "process of (statistical) reasoning is biased if it is influenced by factors – henceforth, 'nonepistemic' factors – that make the conclusion of this reasoning less likely to be true" (Hudson 2021, 399). The qual-c account, too, is a mixed notion: it states that in order for qualitative data collection to be biased, the researcher's commitments must influence the data collection, *and* this must result in the data set ending up lacking one or several good-making features. It is instructive briefly to consider why the qual-c account may not be reduced to either a source or outcome notion.

To begin with, it will not do to identify bias in qualitative data collection merely with researchers' commitments influencing the data collection process. The reason is that researchers' commitments need not have a negative impact on the quality of their data collection. In fact, their commitments may even positively affect its quality. To see this, consider a researcher who studies the views on migration within a community and whose selection of interviewees is affected by her own attitude to migration (she seeks out people whose views align with her own). This is unproblematic if the influence is limited to the early phases of data collection and the researcher makes sure later to seek out people with different views so that she ends up with a data set that reflects the full spectrum of views on migration within the community. Likewise, imagine a researcher who studies how divorce is experienced (this example is adapted from Anderson 2004, 11ff). Because of her feminist values, she does not share the prevailing traditional family values, and she invites the research participants to reflect not only on the negative, but also on the positive consequences of getting divorced. As a result, she ends up with a data set that, differently from existing studies, allows her to capture multiple aspects to the research participants' experience of divorce. These considerations are in keeping with recent philosophical and social scientific discussions. Here, the prevailing view is that commitments (i.e., nonepistemic values, experiences, etc.) may sometimes affect the research process in acceptable and desirable ways (see, e.g., Bueter 2022, 307; Douglas and Elliott 2022, 202; Holman and Wilholt 2022; Maxwell 2013, 124; Peshkin 1988). In the present context, reflections along these lines explain why the qual-c account cannot be cut down to a source notion of bias: it would then fail to characterize bias as a purely negative feature of research.

It will not work either to define researcher bias in qualitative data collection merely in terms of its outcome. In that case, the account would simply state that bias transpires when the upshot is a data set that lacks one or several of the good-making features. However, other factors besides researchers' commitments may, via the data collection process, result in a data set that fails to realize a good-making feature. For instance, researchers being sloppy, tired, or having poor eyesight may have this effect too. Accordingly, it is necessary to specify that bias only transpires when researchers

end up with a deficient data set due to the impact of their commitments on the data collection.

A few words are also in order about the way in which the qual-c account compares with other conceptions in terms of its specification of the source and outcome of bias. The qual-c account is completely in line with the widely held view that a whole variety of commitments (values, expectations, feelings, and so on) may potentially give rise to a biased research process (see, e.g., Bueter 2022, 308–9; Douglas and Elliott 2022, 202; Resnik 2000, 260ff; Roulston and Shelton 2015).

In contrast, the qual-c account departs significantly from existing conceptions with respect to its characterization of the outcome. Typically, outcome and mixed notions spell out the outcome as the lack of truth or accuracy of the research findings. In the context of qualitative data collection, there is nothing wrong with saying that the upshot of biased data gathering is a false or inaccurate answer to the research question: biased data collection gives rise to a data set that lacks one or several good-making features and this, in turn, has as consequence that the answer to the research question will be (partly) false or inaccurate (and incomplete). However, it is more precise, as the qual-c account does, to point to the immediate outcome of biased data collection, namely a data set that fails to realize one or several good-making features like descriptive adequacy, balance, and size-wise sufficiency. This specification of the outcome is more useful to keep in mind when collecting data: it details what researchers should keep an eye on when trying to steer clear of biasing their data collection.

The foregoing reflections show why the qual-c account is a suitable conception of researcher bias in qualitative data generation. Insofar as it is a mixed notion, it avoids the problems it would run into if reduced to either a source or outcome conception. Moreover, the account stands out by characterizing the outcome of biased data collection in rather precise terms. This renders the account more useful when aiming to avoid researcher bias in qualitative data collection.

4. Debiasing strategies in qualitative data collection

Equipped with the qual-c account of bias, I now show that, capitalizing on key characteristics of qualitative methods of data collection, researchers may avail themselves of two debiasing methods, the qual-c strategies. In combination, they are effective in mitigating researcher bias in qualitative data generation. Also, I briefly compare the qual-c strategies to existing proposals for mitigating bias in qualitative research.

The qual-c strategies have as their sole focus how single researchers may prevent and eliminate bias during the data collection phase. Of course, qualitative researchers may also take steps to preclude bias prior to the onset of data generation. For example, they may conduct pilot interviews (test interviews) that allow them to discover whether their questions will bias the research process (see, e.g., Chenail 2011). Likewise, qualitative researchers may receive feedback from peers on whether their ongoing data collection is biased. While clearly important, I set these sorts of debiasing measures to a side. This clarified, here are the two qual-c strategies:

The *bias prevention* strategy: During data collection, continuously and actively aim for a data set with the good-making features while preventing any commitments from interfering with this effort.

The *bias checking* strategy: During data collection, collect evidence that may confirm and possibly disconfirm that the data set will end up having the good-making features. In case of disconfirming evidence, determine whether the lacking good-making feature was caused by a commitment in order to rectify the data collection.

Both strategies revolve around the good-making features that a qualitative data set should end up realizing. To render their discussion manageable, I continue to focus on the three features of descriptive adequacy, balance, and size-wise sufficiency. With this focus, the bias prevention strategy has it that researchers should constantly monitor and adapt their data collection with a view to obtaining a descriptively adequate, balanced, and large enough data set. At the same time, they should make sure that no commitments on their part interfere with this goal. By way of illustration, return to the example of the researcher who studies how the professors at a college evaluate a new study program. This time around, however, imagine that from the very beginning the researcher applies the bias prevention strategy. In order not to bias her data collection, she starts by asking herself whether she has any commitments that might possibly have a biasing influence, and she quickly realizes that she has a highly positive attitude towards the new study program. Consequently, as she generates her data, she aims for a data set with the good-making features, while attempting to prevent this commitment (her enthusiasm for the new study program) from having any biasing effect: she continuously reminds herself to ask the professors whether they perceive any problems with the program, just as she is particularly attentive to their critical comments.

The bias prevention strategy ensures that researchers do not lapse into biased data collection out of inattention or thoughtlessness, but obviously it cannot stand on its own. Researchers may not be conscious of commitments that bias their data generation, or they may be aware of potentially biasing commitments but not succeed in precluding them from biasing their research. The purpose of the bias checking strategy is to remedy this problem.

The strategy admonishes researchers to collect data that may confirm and possibly disconfirm that their data set has the three good-making features. Further, in case of disconfirming evidence, researchers should work out whether the lacking good-making feature is due to the influence of a commitment on their part. And if it is, they should rectify their data collection. To see how this works, return to the example of the researcher whose political views bias her data collection at a political meeting so that her fieldnotes about a local politician's claims and their reception end up being descriptively inadequate. This time, assume that the researcher pursues the bias checking strategy (in addition to the bias prevention strategy). Accordingly, the researcher asks research participants to comment on her notes about the meeting (this exemplifies the techniques of respondent validation), just as she listens to research participants who discuss the meeting amongst themselves. The resulting evidence makes her realize that her field notes are descriptively inadequate so she

starts pondering whether she has any commitments that might explain her erroneous notes. In this process, she notes that her data about the other local politicians have so far been descriptively adequate and that the local politician in question is the only politician with whom she strongly disagrees. On this basis, she reaches the conclusion that her deep political disagreement with the local politician is likely to have biased her data collection. Hence, she resolves to pay special attention to avoiding this mistake when she applies the bias prevention and bias checking strategies in her subsequent data collection.

The two strategies may be further spelled out in terms of the distinction between first- and second-order evidence (see Zahle 2019 and also Staley 2004). First-order evidence is data that function as evidence for (or against) researchers' answers to their research questions. To play the role of first-order evidence, a data set should realize the good-making features. Against this backdrop, the focus of the bias prevention strategy may be specified as the generation of first-order evidence: researchers should actively aim for a data set with the good-making features that may function as evidence for an answer to their research question. At the same time, they should make sure that their attainment of this goal is not obstructed by their commitments.

Data function as second-order evidence when they play the role of evidence for (or against) other pieces of data being suited to serve as first-order evidence.[7] Second-order evidence should be collected whenever it may reasonably be doubted that first-order evidence realizes, or will end up realizing, a good-making feature. Further, second-order evidence should possess the good-making features too. Accordingly, if this may plausibly be doubted, data in support of their high quality should also be provided. In this perspective, the bias checking strategy admonishes researchers to generate second-order evidence to confirm or disconfirm that their data set will end up realizing the good-making features. Also, they should rectify their data generation in case bias has occurred.

In combination, I submit, the qual-c strategies are effective debiasing methods. As I now explicate, their effective use is made possible by several main characteristics of qualitative data collection (discussed in section 2). Consider first the flexibility of qualitative data generation, that is, the fact that researchers make the majority of their data collection decisions on the go. The flexibility of participant observation and qualitative interviewing enables the employment of the qual-c strategies. It allows researchers to adapt their data collection with the aim of ending up with a data set with the good-making features (the bias prevention strategy), and to check up on whether this goal is being realized (the bias checking strategy). Thus, the more flexibly the qualitative methods are used, the more they leave room for the implementation of the strategies, and vice versa.

Additionally, qualitative data collection is characterized by researchers having little, or no, control over the research environment. Consequently, during data collection the research participants may do and says things researchers did not expect or did not look for. The researcher should take these incidents into

---

[7] Note that a piece of evidence may serve simultaneously as first- and second-order evidence: it may serve as evidence for (or against) a researcher's answer to their research question and for (or against) some other piece(s) of evidence being fit to serve as evidence for such an answer.

consideration when they adapt their data collection to the goal of obtaining a data set with the good-making features (the bias prevention strategy). Further, the incidents they come across may not only confirm, but also disconfirm, that their data set will end up realizing the good-making features (the bias checking strategy). In both cases, the limited control over the research environment prevents researchers from curbing and rigging their data generation so that their data line up with their commitments. In this manner, the limited control over the research environment contributes to the effectiveness of the qual-c strategies in mitigating bias.

Lastly, qualitative data collection involves the gathering of lots of heterogenous data. The large amount of heterogenous data provides researchers with more to go on when deciding how best to adapt the data collection to their goal (a data set with the good-making features), and when testing whether they are moving towards the realization of this goal. Moreover, social scientist Howard Becker notes that the "more observations one makes and the more different kinds of observations one makes, the more difficult it becomes to ignore or explain away evidence that runs counter to one's expectation or bias" (Becker 1970, 56). Adapted to the present discussion, the point is that the larger the set of heterogenous data, the more pieces of data indicative of bias are likely to be generated. As a result, researchers are to a higher degree put under pressure to recognize and take into account this "recalcitrant" evidence when pursuing the two strategies. Thus, these characteristics of qualitative methods, too, underwrite the effective employment of the qual-c strategies.

To complete the presentation of the qual-c strategies, let me briefly comment on how they relate to existing proposals for mitigating bias in qualitative data collection. In the qualitative research literature, at least two main approaches may be discerned. One is the reflexivity approach. It takes as its starting point that, as qualitative researchers go along, they should be reflexive, that is, ponder and get clear on how their commitments (values, assumptions, etc.) affect their research, including their data collection (Bryman 2012, 715; Roulston and Shelton 2015, 333). Sometimes this recommendation is developed into a debiasing strategy by adding that, once aware of their commitments and their impact, researchers should prevent them from biasing their data collection (see, e.g., Olmos-Vega et al. 2023, 241; Palaganas et al. 2017, 432; Sprague 2016, 90).[8] Compare now the reflexivity approach to the qual-c strategies. Though the bias prevention strategy does not explicitly state it, researchers will clearly need to reflect on their commitments and pay attention to their impact on the data gathering in their effort to preclude the commitments from having any biasing influence. Accordingly, the bias prevention strategy may be said to encompass the reflexivity approach. At the same time, the strategy goes beyond the approach by maintaining that researchers should aim for a data set with the good-making features.

The second debiasing approach may be referred to as the evidence approach (see, e.g., Lincoln and Guba 1985, 301; Maxwell 2013, 121ff; Padget 2008, 185ff). Its guiding

---

[8] When the advice to be reflexive is not developed into a debiasing strategy, it is proposed that researchers should register how their commitments influence their data collection so that they may explicitly state this when presenting their research findings (Mantzoukas 2005, 291–92). In passing, note the main problem with this approach: because its proponents oppose the use of any debiasing strategies, research findings are likely to end up being based on data that are descriptively inadequate, unbalanced, or too few in number. But what is the use of such findings?

idea is captured by Joseph Maxwell's statement that "validity threats [like bias] are made implausible by *evidence*" (Maxwell 2013, 121—italics in original; see also Maxwell 2012, 144–45). In this vein, the approach encourages researchers to use a handful of data collection techniques that will allow them to detect bias (and other validity threats). For instance, researchers should look for discrepant evidence or use respondent validation, i.e., ask research participants to assess the quality of their data (Maxwell 2012, 126ff). Returning to the qual-c strategies, the bias checking strategy shares the emphasis on collecting evidence that is indicative of whether or not the data collection is biased. Yet, it differs from the evidence approach in two ways: It asks researchers to generate this evidence with a view to ending up with a data set that has the good-making features, and it leaves it open how researchers should approach the task of providing the evidence, while implying that there are multiple ways to go about it. The latter means that though researchers are not confined to using these techniques, they may well look for discrepant evidence or use respondent validation in their effort to confirm (or possibly disconfirm) that their data set has the good-making features. In this sense, the bias checking strategy may be said to comprise the evidence approach.

These considerations show that, even though the qual-c strategies are not explicitly advocated in the qualitative research literature, two existing debiasing approaches, viz. the reflexivity and evidence approaches, are similar to the strategies in important respects. In other words, the strategies involve ideas and suggestions that are familiar from discussions of qualitative research, while also going beyond these proposals for mitigating bias.

What about philosophy of science discussions? Here, no comprehensive methods *specifically* aimed at mitigating bias in qualitative data collection have been advanced. Yet, some of the insights informing the qual-c strategies have been stated in more general terms or in relation to other forms of research. Most notably, within standpoint theory, a branch of feminist philosophy of science, it is often recommended that researchers in general should reflect on their commitments and prevent those with a negative (i.e. biasing) impact from influencing their research (see, e.g., Crasnow 2008, 1095–96; Harding 2004, 136ff). This proposal is consistent with the bias prevention strategy in the sense that the strategy also invites qualitative researchers to engage in this kind of reflection and make an effort not to bias their data generation. Further, in philosophical discussions of experiments it is sometimes advocated that researchers should generate evidence that may be used to assess the quality of their experimental data (see, e.g., Feest and Steinle 2016, 278ff; Franklin and Perovic 2023, sections 2.1.1.–2.1.2). This suggestion is similar in spirit to the bias checking strategy in that the latter also admonishes qualitative researchers to provide evidence for or against their data set having the good-making features. In these manners, the qual-c strategies involve ideas that are well known from the philosophy of science literature.

5. Possible objections

In this section, I consider four possible objections to the qual-c strategies being effective debiasing strategies. I call them the simple argument, the argument from interpretation, the too-demanding argument, and the argument from critical

scrutiny. The view that qualitative research lacks effective debiasing methods is often motivated by comparison with how bias is mitigated in experimental and quantitative social research. Thus, I adopt a similar approach in connection with most of the objections I examine below. At the same time, I set aside the question of whether the arguments paint a true picture of experimental and quantitative debiasing methods. For my purposes, this is irrelevant in that I am solely interested in the effectiveness of the qual-c strategies.

The simple argument is rarely, if ever, explicitly stated (or I have not come across it). Still, my impression is that something like it often underlies unreflective versions of the view that qualitative research lacks effective debiasing methods. In any case, the argument is instructive to consider. It goes as follows: Debiasing methods within experimental and quantitative social research are effective. For example, consider the method of randomization. In experiments, its use means that the research participants who will receive the experimental treatment are randomly selected. In connection with the quantitative methods of structured interviewing and structured observation, its employment requires the random selection of the research participants to be interviewed and observed, respectively. In all three cases, randomization ensures that researchers' commitments cannot influence, and hence possibly bias, the selection process. Contemplate, too, the use of closed questions in structured interviews, that is, questions with fixed answer options. When using such questions, researchers merely have to circle the response chosen by the interviewee. Hence, their commitments cannot really affect, and so possibly bias, how they record the responses. These kinds of effective debiasing measures, the simple argument continues, are not part of the qual-c strategies. For this reason, it concludes, the qual-c strategies must be deemed ineffective.

Once explicated, the shortcoming of the simple argument is easily seen. It assumes, rather than demonstrates, that experimental and quantitative debiasing methods are *alone* in being effective. Or, differently put, the objection fails to provide any reasons as to *why* the qual-c strategies fall short of being effective in mitigating bias. On this ground, it should be rejected.

In reaction to the simple argument, it may also be noted that experimental and quantitative debiasing methods are mostly inapplicable in the context of qualitative data collection in the sense that they are incompatible with its main characteristics. For instance, the flexibility distinctive of qualitative interviewing and participant observation would be reduced if research participants were randomly selected rather than invited as researchers went along gathering their data. Likewise, recall that qualitative researchers have limited control over the research environment. The use of closed questions would result in a significant increase of researchers' control over the interviewees' responses to their questions as the latter could no longer formulate these themselves. Against this backdrop, it is no surprise that the qual-c strategies fail to include experimental and quantitative debiasing methods. In fact, it may be said to speak in favor of the qual-c strategies that they are consistent with, and even take advantage of, the main characteristics of qualitative data collection.

Be that as it may, the dismissal of the simple argument shows that objections to the qual-c strategies need to challenge their effectiveness in a more direct manner. An argument of this sort may take its inspiration from a recent paper by David Teira (Teira 2021). In the paper, Teira critically examines a debiasing method in

anthropology called the norm of cultural relativism. According to it, field notes should be culturally unbiased, that is, uncontaminated by the researcher's own cultural values and prejudices (Teira 2021, 1079–80). The problem with the norm, Teira contends, is that it is open to interpretation with respect to what constitutes culturally unbiased data (1085). Further, there is no single source of authority on this matter. In consequence, he goes on, anthropologists have failed to reach an agreement on when the norm should be regarded as breached in particular cases (1085). In other words, the norm "makes it impossible to adjudicate whether [...] some] data are actually biased" and hence it is an "ineffective bias correction" (1080).

Similar considerations may now be applied to the qual-c strategies and called the argument from interpretation. The strategies, it might be ventured, are equally open to interpretation. They state that researchers should aim for, and check up on whether they have obtained, a data set with good-making features like descriptive adequacy, balance, and size-wise sufficiency. But surely it is a matter of interpretation what having these features amounts to, just as there is no single source of authority on what constitutes their right interpretation. As a result, if qualitative researchers began explicitly to employ and invoke the qual-c strategies, they would not be likely to agree on when the strategies should be regarded as followed in particular cases. By appeal to the strategies, it would be impossible to adjudicate whether a specific process of data collection was (should count as) biased or unbiased. For this reason, the argument from interpretation concludes, the qual-c strategies are ineffective debiasing methods.

The objection maintains that researchers are likely most of the time to disagree on the interpretation of the good-making features.[9] In my view, this contention is unfounded. In discussions of qualitative research, the good-making features are often, implicitly or explicitly, endorsed as characteristics of high-quality qualitative data sets (Zahle 2021, 106). Accordingly, it may be determined whether their interpretation is continuously in dispute; if that is the case, then it is reasonable to expect this to show in textbooks and papers on qualitative research. Yet, I have not come across any comments along these lines. By way of illustration, consider two examples relating to the feature of descriptive adequacy. In his three-page discussion of the feature (which he calls "descriptive validity"), the social scientist Joseph Maxwell does not mention any problems with its interpretation (Maxwell 2012, 134ff). Similarly, in his widely used introduction to social research methods, the social scientist Alan Bryman simply takes it for granted that field and interview notes should be descriptively adequate, while not in any way indicating that there are often divergent views as to what this amounts to (Bryman 2012, 447ff, 482ff). These observations suggest that qualitative researchers mostly agree on the interpretation of the good-making features. Further, there is no reason to think that this would change just because the features were invoked as part of the qual-c strategies. Thus, if qualitative researchers began explicitly to call upon the strategies, they would likely, for the most part, interpret the strategies (the good-making features) similarly when adjudicating whether an actual process of data collection was (should count as)

---

[9] Since I do not defend the norm of relativism, I shall not go into a discussion of Teira's criticism of it. On my account, biased data collection due to researchers' cultural values and prejudices should instead be mitigated by applying the qual-c strategies.

biased. By implication, the argument from interpretation no longer has any force: the claim that the strategies are ineffective because there would be no consensus on their interpretation may be rejected.

Teira's reflections may also serve as the inspiration for two more possible objections (Teira 2021). In his paper, Teira compares the norm of cultural relativism with experimental debiasing methods like the use of randomization. The latter methods, he maintains, are effective because they are not open to interpretation and are easy to implement: they single out the specific points in the research process where bias may occur and recommend definite actions to preclude this from happening (Teira 2021, 1086–87). With these points in mind, return to the qual-c strategies. The too-demanding argument points out that the strategies require researchers to be constantly on the watch for bias (the strategies do not lay out the specific points in the research process at which bias may occur). Moreover, the strategies leave it up to researchers to decide what actions to perform as part of their employment (the strategies do not point to a few definite actions that may preclude bias from transpiring). Accordingly, the argument maintains, the strategies are too demanding to implement: researchers are unable to keep up their application throughout the data collection phase. In this sense, the qual-c strategies are ineffective in practice.

There can be no doubt that the application of the strategies requires an effort on the part of researchers. However, in this respect the strategies are similar to qualitative methods of data collection. The flexibility of qualitative methods means that researchers must continuously make decisions about their data gathering as they go along, while taking into account the data they have already collected, their evolving understanding of the field, the situation in which they find themselves, etc. Moreover, it is not uncommon for researchers to check up on the quality of their data set by using respondent validation, looking for discrepant evidence, and the like. In this fashion, qualitative data gathering is demanding, yet skilled qualitative researchers manage nonetheless. The employment of the qual-c strategies in connection with qualitative data collection means that researchers have to make many of their usual decisions about data collection. In addition, they must then aim for a high-quality data set, prevent their commitments from interfering with this aim, and check up on whether they are successful in this respect. The latter does not require skilled researchers to make much more of an effort than they already do merely collecting qualitative data. Thus, the qual-c strategies are not too demanding to implement.

The last possible objection to be examined is the argument from critical scrutiny. It takes off from Teira's claim that an experimental debiasing method specifies a definite action that researchers should perform at some point(s) in the research process to control for bias (Teira 2021). In consequence, it may be added, experimenters may easily enumerate all the actions they performed as part of their debiasing methods when reporting their findings. Thus, others may inspect and assess their debiasing efforts. The argument from critical scrutiny points out that the situation is different when it comes to the qual-c strategies. These do not draw attention to a handful of definite actions that qualitative researchers should perform to mitigate bias. Rather, researchers will have to decide on, and carry out, multiple different actions in pursuing the strategies. In fact, it seems, researchers will need to perform so many types of actions that it will be impossible to list all of them in

connection with the presentation of their research results. Accordingly, other researchers cannot "be walked through" and evaluate all their actions directed at mitigating bias. The argument maintains that this is problematic because an effective debiasing method averts bias in a way that is open to critical scrutiny. The qual-c strategies fail to live up to this requirement.

For the present purposes, I shall grant that effective debiasing strategies are such that their implementation is open to critical scrutiny. I want to question whether the qual-c strategies fail to meet this condition. It is plausible to hold that qualitative researchers will be unable to provide a list of *all* their actions involved in the employment of the qual-c strategies when putting forward their final analysis. Still, they may present examples of main ways in which they pursued the strategies. Also, they may relate their general reflections on potential biasing commitments on their part, how they thought the commitments might potentially bias their data gathering, what measures they took to prevent this from happening, and so on. Moreover, if challenged, they may offer further details and defend their decisions and ways of proceeding. This will do, I submit, to qualify the implementation of the strategies as being open to critical scrutiny because it allows others to evaluate researchers' debiasing efforts: enough information is provided to determine whether researchers may reasonably be trusted to have implemented the strategies in a satisfying manner. Accordingly, the argument from critical scrutiny may be dismissed. This means that none of the four possible objections considered in the foregoing provide any reasons to doubt the effectiveness of the qual-c strategies.

6. Conclusion

In this paper, I first introduced two methods of qualitative data collection and developed an account of researcher bias in qualitative data collection. Then I showed that, capitalizing on main characteristics of qualitative data collection methods, single researchers may avail themselves of two strategies that are jointly effective in mitigating researcher bias. Lastly, I defended the strategies against possible objections to their effectiveness. Accordingly, the widespread view that qualitative research is of poor quality due to the lack of effective debiasing methods may be partly refuted: there is no lack of effective strategies to mitigate researcher bias in qualitative data generation.

The refutation of the widespread view is only partial because bias may, of course, be introduced in other ways and in other parts of the research process. Most notably, researcher bias may transpire in the process of data analysis. How qualitative researchers should deal with this threat is the topic of another paper. But perhaps the present paper provides some grounds for optimism: if it is possible to avert researcher bias in qualitative data gathering, why not think the same goes for qualitative data analysis?

References

Anderson, Elizabeth. 1995. "Knowledge, Human Interests, and Objectivity in Feminist Epistemology." *Philosophical Topics* 23 (2):27–58.

Anderson, Elizabeth. 2004. "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce." *Hypatia* 19(1):1–24. https://doi.org/10.1111/j.1527-2001.2004.tb01266.x

Antony, Louise M. 2018. "Quine as Feminist: The Radical Import of Naturalized Epistemology." In *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, 2nd ed., edited by Louise Antony and Charlotte Witt, 110–53. New York: Routledge. https://doi.org/10.4324/9780429502682

Becker, Howard S. 1970. *Sociological Work: Method and Substance*. Oxford: Routledge.

Bryman, Alan. 2012. *Social Research Methods*, 4th ed. Oxford: Oxford University Press.

Bueter, Anke. 2022. "Bias as an Epistemic Notion." *Studies in History and Philosophy of Science* 91:307–15. https://doi.org/10.1016/j.shpsa.2021.12.002

Chenail, Ronald J. 2011. "Interviewing the Investigator: Strategies for Addressing Instrumentation and Researcher Bias Concerns in Qualitative Research." *The Qualitative Report*, 16 (1):255–62. https://doi.org/10.46743/2160-3715/2011.1051

Crasnow, Sharon. 2008. "Feminist Philosophy of Science: 'Standpoint' and Knowledge," *Science & Education* 17:1089–110. https://doi.org/10.1007/s11191-006-9069-z

Crasnow, Sharon. 2017. "Bias in Social Science Experiments." In *The Routledge Companion to Philosophy of Social Science*, edited by Lee McIntyre and Alex Rosenberg, 191–201. London: Routledge. https://doi.org/10.4324/9781315410098

DeWalt, Kathleen M., and Billie R. DeWalt. 2011. *Participant Observation. A Guide for Fieldworkers*, 2nd ed. Lanham, MA: Altamira Press.

Douglas, Heather, and Kevin C. Elliott. 2022. "Addressing the Reproducibility Crisis: A Response to Hudson." *Journal for General Philosophy of Science* 53:201–9. https://doi.org/10.1007/s10838-022-09606-5

Feest, Uljana, and Friedrich Steinle. 2016. "Experiment." In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys, 274–95. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199368815.013.16

Franklin, Allan, and Slobodan Perovic. 2023. "Experiment in Physics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. Stanford, CA: Stanford University Press. https://plato.stanford.edu/archives/fall2023/entries/physics-experiment/

Galdas, Peter. 2017. "Revisiting Bias in Qualitative Research: Reflections on Its Relationship With Funding and Impact." *International Journal of Qualitative Methods* 16 (1). https://doi.org/10.1177/1609406917748992

Glesne, Corrine, and Alan Peshkin. 1992. *Becoming Qualitative Researchers. An Introduction*. New York: Longman.

Hammersley, Martyn. 1993. *What's Wrong with Ethnography?* London: Routledge.

Hammersley, Martyn. 2013. *What is Qualitative Research?* London: Bloomsbury. https://doi.org/10.5040/9781849666084

Hammersley, Martyn, and Roger Gomm. 2000. "Bias in Social Research." In *Taking Sides in Social Research*, edited by Martyn Hammersley, 151–66. London: Routledge.

Harding, Sandra. 2004. "Rethinking Standpoint Epistemology: What is 'Strong Objectivity?'" In *The Feminist Standpoint Reader*, edited by Sandra Harding, 127–40. New York: Routledge.

Holman, Bennett, and Wilholt Torsten. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91:211–20. https://doi.org/10.1016/j.shpsa.2021.11.011

Hudson, Robert. 2021. "Should We Strive to Make Science Bias-Free? A Philosophical Assessment of the Reproducibility Crisis." *Journal for General Philosophy of Science* 52:389–405. https://doi.org/10.1007/s10838-020-09548-w

Kincaid, Harold, John Dupré, and Alison Wylie. 2007. "Introduction." In *Value-Free Science? Ideals and Illusions*, edited by Harold Kincaid, John Dupré, and Alison Wylie, 3–23. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195308969.003.0001

Lietz, Cynthia A., and Luis E. Zayas. 2010. "Evaluating Qualitative Research for Social Work Practitioners." *Advances in Social Work* 1 (2):188–202. https://doi.org/10.18060/589

Lincoln, Yvonna S., and Egon G. Guba. 1985. *Naturalistic Inquiry*. Newbury Park: Sage.

Mantzoukas, Stefanos. 2005. "The Inclusion of Bias in Reflective and Reflexive Research." *Journal of Research in Nursing* 10 (3):279–95. https://doi.org/10.1177/174498710501000305

Maxwell, Joseph A. 2012. *A Realist Approach for Qualitative Research*: Thousand Oaks, CA: Sage.

Maxwell, Joseph A. 2013. *Qualitative Research Design. An Interactive Approach*, 3rd ed. London: Sage.

May, Joshua. 2021. "Bias in Science: Natural and Social." *Synthese* 199:3345–66. https://doi.org/10.1007/s11229-020-02937-0

Montuschi, Eleonora. 2014. "Scientific Objectivity." In *Philosophy of Science. A New Introduction*, edited by Nancy Cartwright and Eleonora Montuschi, 123–44. Oxford: Oxford University Press.

Olmos-Vega, Francisco M., Renée E. Stalmeijer, Lara Vapio, and Renate Kahlke. 2023. "A Practical Guide to Reflexivity in Qualitative Research: AMEE Guide No. 149." *Medical Teacher*, 45 (3):241–51. https://doi.org/10.1080/0142159X.2022.2057287

Padgett, Deborah K. 2008. *Qualitative Methods in Social Work Research*, 2nd ed. Thousand Oaks, CA: Sage.

Palaganas, Elinda, Marian C. Sanchez, Maria P. Molintas, and Ruel D. Caricativo. 2017. "Reflexivity in Qualitative Research: A Journey of Learning." *The Qualitative Report*, 22 (2):426–38. https://doi.org/10.46743/2160-3715/2017.2552

Peshkin, Alan. 1988. "In Search of Subjectivity – One's Own." *Educational Researcher* 17 (7):17–21. https://doi.org/10.3102/0013189X017007017.

Resnik, David B. 2000. "Financial Interests and Research Bias." *Perspectives on Science* 8 (3):255–85. https://doi.org/10.1162/106361400750340497

Roulston, Kathryn, and Stephanie A. Shelton. 2015. "Reconceptualizing Bias in Teaching Qualitative Research Methods." *Qualitative Inquiry* 21 (4):332–42. https://doi.org/10.1177/1077800414563803

Schumm, Walter R. 2021. "Confirmation Bias and Methodology in Social Science: An Editorial." *Marriage & Family Review* 57 (4):285–93. https://doi.org/10.1080/01494929.2021.1872859

Sprague, Joey. 2016. *Feminist Methodologies for Critical Researchers*, 2nd ed. Lanham, MA: Rowman & Littlefield.

Staley, Kent W. 2004. "Robust Evidence and Secure Evidence Claims." *Philosophy of Science* 71 (4):467–88. https://doi.org/10.1086/423748

Steel, Daniel. 2018. "Wishful Thinking and Values in Science." *Philosophy of Science* 85 (5):895–905. https://doi.org/10.1086/699714

Stegenga, Jacob. 2018. *Medical Nihilism*. Oxford: Oxford University Press.

Teira, David. 2021. "On the Limits of Cultural Relativism as a Debiasing Method." *Philosophy of Science* 88 (5):1079–89. https://doi.org/10.1086/714853

Zahle, Julie. 2018. "Values and Data Collection in Social Research," *Philosophy of Science* 85 (1):144–63. https://doi.org/10.1086/694770

Zahle, Julie. 2019. "Data, Epistemic Values and Multiple Methods in Case Study Research." *Studies in History and Philosophy of Science Part A* 78:32–9. https://doi.org/10.1016/j.shpsa.2018.11.005

Zahle, Julie. 2021. "Objective Data Sets in Qualitative Research." *Synthese* 199:101–17. https://doi.org/10.1007/s11229-020-02630-2