

Accept or Address? Researchers' Perspectives on Response Bias in Accessibility Research

JOY MING

INFORMATION SCIENCE, CORNELL, JMING@INFOSCI.CORNELL.EDU

SHARON HEUNG

INFORMATION SCIENCE, CORNELL TECH, SSH247@CORNELL.EDU

SHIRI AZENKOT

INFORMATION SCIENCE, CORNELL TECH, SHIRI.AZENKOT@CORNELL.EDU

ADITYA VASHISTHA

INFORMATION SCIENCE, CORNELL, ADITYAV@CORNELL.EDU

Response bias has been framed as the tendency of a participant's response to be skewed by a variety of factors, including study design and participant-researcher dynamics. Response bias is a concern for all researchers who conduct studies with people — especially those working with participants with disabilities. This is because these participants' diverse needs require methodological adjustments and differences in disability identity between the researcher and participant influence power dynamics. Despite its relevance, there is little literature that connects response bias to accessibility. We conducted semi-structured interviews with 27 accessibility researchers on how response bias manifested in their research and how they mitigated it. We present unique instances of response bias and how it is handled in accessibility research; insights into how response bias interacts with other biases like researcher or sampling bias; and philosophies and tensions around response bias such as whether to accept or address it. We conclude with guidelines on thinking about response bias in accessibility research.

CCS CONCEPTS • Human-centered computing---Accessibility---Accessibility theory, concepts and paradigms • Social and professional topics---user characteristic--- people with disabilities

Additional Keywords and Phrases: Response bias, participant-researcher power dynamics, charity model of disability, accessibility dongle

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper; this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

Accessibility researchers and practitioners are increasingly engaging with people with disabilities to design appropriate, inclusive, and accessible technologies [2,11,66]. However, differences between researchers and their participants, such as background, social status, technical expertise, and perceived abilities, make it difficult for participants to provide critical feedback on new designs and interventions. This tendency of participants to share feedback that deviates from the “correct, honest, and accurate response” [26] is called *response bias*. Prior research has shown high levels of response bias in studies with marginalized communities where participants provided researchers feedback to please them or help achieve their research goals [18]. If unaddressed, response bias can

result in misleading findings or unusable technologies that inadvertently harm the very users or communities researchers intend to serve [16,18,60,62].

Response bias is concerning when conducting studies with people with disabilities because it can be exacerbated by methodology and power differentials. The study design for research in accessibility, including specialized consent processes, communication methods, and recruitment, could amplify response bias [5]. In addition, power dynamics can be influenced by the lack of representation of people with disabilities as researchers due to systematic exclusion [66]. Despite the importance of studying response bias in accessibility research, there is only one study that substantially engages both topics. Trewin et al. 2015 showed that compared to typical student participants, participants with visual impairment gave higher ratings that masked usability issues [60].

We conducted semi-structured interviews with 27 accessibility researchers to understand how response bias manifested in their studies and how they mitigated it. In particular, we investigated what response bias meant to these researchers, how it has shown up in their research, what practices they used to handle it, and whether it influenced their outcomes. Although many of the researchers were initially unsure about the definition of response bias, they were able to pinpoint many instances where they noticed biases in their participants' responses. Their discussions presented a more nuanced understanding of response bias in accessibility beyond simple manifestations and mitigations.

These anecdotes of how the accessibility researchers experienced and handled response bias were influenced by participant-researcher dynamics, the participant's view of their identity and abilities, and the broader context of accessibility technology. Some of the researchers shared instances that aligned with "classic" definitions of response bias from the literature, like when participants gave responses that supported the researchers' goals (i.e. *demand characteristics* [52]) or did not offend them (i.e., *social desirability bias* [26]) to encourage the researchers to "help" people with disabilities. This effect of the "charity model of disability" [27] also influenced how the researchers asked and interpreted questions (i.e., *researcher bias* [3,23]), extending the discussion to other biases beyond "classic" response bias. The researchers also shared strategies they used to mitigate response bias, including learning from the disability community to develop appropriate communication methods and practice relevant etiquette. However, the researchers did not always want to mitigate response bias — in some cases they leveraged it to understand the diversity of their participants' experiences or question their own roles in structural oppression.

Drawing on these findings, we synthesize key takeaways in the form of guidelines for accessibility researchers on how to engage with response bias when working with participants with disabilities. We make the following contributions to the accessibility community and the conversation around response bias:

- **Observations and practices specifically from accessibility technology**, how response bias shows up due to unique power dynamics and methodological considerations.
- **Interactions of response bias with other biases**, how sampling bias and researcher bias could reinforce response bias to influence participants' responses.
- **Perceptions and tensions around response bias**, the researchers' philosophies on how they decide whether to accept or address response bias.

2 RELATED WORK

Response bias is a growing concern specifically within accessibility research due to specialized methodology and unique power differentials. After conducting a literature search¹ on response bias and accessibility, we found only one study that looked at positive bias in Likert scale questions with participants with visual impairment [60]. Due to this gap, we incorporate literature about response bias across many disciplines, including foundational research in psychology [26,50,52,53] and implementations in HCI [10,18,65].

¹ We conducted a literature search using the ACM Digital Library in ASSETS (1994-2020), CHI (1981-2020), and CSCW (1986-2020) Proceedings.

2.1 Types of Response Bias

Response bias has been defined as “a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content” [53]. This “basis” [53] can include a wide range of phenomena, including study design and participant-researcher dynamics. We use the “classic” definition of response bias from the literature to indicate ways in which participant responses can be directly impacted (see Table 1).

Table 1: Select List of Biases That Can Impact Participants’ Responses

“Classic” Response Bias	Other Biases
<p><i>Study Design</i></p> <ul style="list-style-type: none"> ● Study Fatigue ● Task Completion Bias ● Incentive Bias ● Novelty Bias <li style="padding-left: 20px;">Halo Effect <li style="padding-left: 20px;">Legacy Bias <li style="padding-left: 20px;">Unfamiliar Content ● In lab vs. In wild ● Extreme Responding 	<p style="text-align: center;"><i>Participant-Researcher Dynamics</i></p> <ul style="list-style-type: none"> ● Demand Characteristic ● Social Desirability Bias <li style="padding-left: 20px;">Impression Management ● Acquiescence Bias <li style="padding-left: 20px;">Agreement Bias ● Observer Effect <li style="padding-left: 20px;">Hawthorne Effect ● Study Fatigue
	<ul style="list-style-type: none"> ● Selection Bias <li style="padding-left: 20px;">Sampling Bias <li style="padding-left: 20px;">Sample Selection Bias <li style="padding-left: 20px;">Participation Bias ● Researcher Bias

Study Design. One way study design could affect participants’ responses is through *study fatigue* [57], where participants are burdened by the length of the survey, density of information, or consecutive participation [54,59]. Study design also includes *task completion bias*, where participants feel more inclined to finish the tasks that they are assigned [39] or *incentive bias* where participants may give more positive feedback to get compensation [32,58]. Moreover, participants in technology-based studies might experience a *novelty bias* or *halo effect*, inflating feedback because they are simply excited to try something new [8,46].

The type of study also influences participants’ responses. Participants may respond differently if the study is conducted “in the lab” or “in the wild” [10] because it changes how comfortable participants feel or how realistic their responses may be. Quantitative and qualitative research also have different biases. For example, in survey-based research, researchers found that participants with disabilities tend towards *extreme-responding* [26] and gave ratings that were more positive and less correlated with objective metrics [60]. In interview-based research, participant-researcher dynamics may influence responses [20,35,41].

Participant-Researcher Dynamics. Prior research has found that the “social distance” between researchers and participants strongly influence response bias [19,38]. For example, a study showed that participants were more likely to give overly positive responses to a foreign researcher than to a local one because of the large differences in social and demographic factors, even if the participants themselves were not consciously doing so [18]. Types of response bias affected by “social distance” include *demand characteristics*, when research participants want to be a “good participant” and support what they perceive are the researchers’ goals and hypotheses [52]; *social desirability bias* or *impression management*, where participants respond in a way that either projects the image of themselves as a “good person” to the researcher or maintains their own view of themselves as “good” [22,50]; and *acquiescence bias*, where the participant tries to be as accommodating and helpful to the researcher as possible even if it is not their honest opinion [26]. The literal “distance” between the participant and the researcher can also influence their actions and responses, like with the *observer effect*, where the participant may act differently because they know the researcher is watching [36,48].

We contribute to this body of literature by highlighting how response bias from study design and participant-researcher dynamics shows up in accessibility research settings. Moreover, we link previously disconnected literature about other limitations and biases that might influence responses, like *researcher bias* [3,24] and *sampling bias* [25,31,67].

2.2 Handling Response Bias

Psychology research pioneered strategies that reduced the “expected risks and losses” for answering truthfully to address response bias, especially social desirability bias [50]. This included distancing the participant from their answer (i.e., randomized/anonymous answers), decreasing concerns in admitting a taboo (i.e., confidentiality assurances), or heightening the subjective benefit of telling the truth [40]. HCI researchers have used *social proofing*, which reduced expected risk by demonstrating the participants’ responses were similar to their peers [62]. Another example is the adaptation of *indirect questioning* [21] into *role playing*, allowing the participants to answer as someone else [14,47]. There was also the use of the *list experiment*, where participants obscure their exact responses by reporting the number of options they select from a list [1].

Another mitigation strategy is pinpointing when response bias is happening and contextualizing it. Researchers have used tools like the Minnesota Multiphasic Personality Inventory (MMPI), the Response Bias Scale (RBS), and the Lie scale [27] to assess consistent individual differences in response styles [53]. In HCI, researchers have used a *response bias index* to test the accuracy of formulas [12]. Identifying response bias could also include *triangulation* or taking multiple measurements to find where participants contradict themselves [28,61]. Other examples in HCI research include taking additional measurements like mouse movements [33,65] or piloting protocols or prototypes [7].

Adjusting the “distance” between the participants and the researchers could also reduce the potential for response bias. Some studies combatted observer bias with online surveys [17] or social desirability bias with an animated interviewer [37]. Others increased the participants’ familiarity with the researcher and the experiment by introducing participants to the technology ahead of time [49], maintained longer-term interactions with the participants [6], or conducted the study in a location where participants are comfortable [10]. These strategies could build rapport between participants and researchers and reduce the social gap between them.

Handling response bias could also involve the researcher being critical about their own perspective and how it influences the way the study is run. Even a *double-blind study* used to mitigate the placebo effect [30,51] carries the implication that the researchers’ mindset can influence the way in which they conduct the study and, thus, the participants’ responses. One approach is to *interview the investigator* [15], questioning and contextualizing their own biases. Another is to embrace the presence of response bias and to adjust the focus of the study accordingly. For example, the existence of *demand characteristics* could indicate the need to focus more on how participants use technology rather than whether the participants would use it [10].

Our paper builds upon this literature by highlighting accessibility researchers’ experiences addressing response bias when working with people with disabilities. Moreover, the researchers we interviewed also shared reflections on some of their considerations and tradeoffs for whether they handled response bias in the first place.

3 METHODS

The purpose of this study is to understand how accessibility researchers see response bias manifest in their studies and the steps they take to mitigate its effects. We conducted semi-structured interviews with researchers who have directly interacted with participants with disabilities. We use the term “researchers” to refer to the researchers we interviewed and “participants” to refer to the participants in those researchers’ studies.

3.1 Researchers We Interviewed

We recruited accessibility researchers using a variety of sampling methods: snowball sampling, convenience sampling (i.e., posting on social media and listservs), and inconvenience sampling (i.e., direct emailing) [55]. We interviewed a total of 28 researchers; however, one did not self-identify as an HCI or UX researcher and the

interview was dropped from analysis. The 27 researchers we interviewed came from a variety of backgrounds and experiences (see Table 2). A majority of the researchers conducted research in the US, except for two who conducted research in Ireland and India. While we did not explicitly ask for specific demographic information, some researchers shared how their personal identities shaped their role as a researcher, identifying themselves as having disabilities (n=6), being part of the LGBTQIA+ community (n=3), or being part of a racial or ethnic minority (n=3).

Table 2: Summary of Researchers' Background

Category	Description
Position/Role	Doctoral students (n=5) Post-doctoral researchers (n=2) Professors (n=11) Industry researchers (n=7) Other (i.e., retired or industry and academic professional) (n=2)
Background/Fields (keywords)	human computer interaction, computer science, information science, psychology, behavioral science, public health, clinical studies, critical disability studies, disability justice
Populations They Worked with (note: not mutually exclusive per researcher, terminology from the researchers themselves and [5])	Blind or visually impaired (n=18) Mobility-related disability (n=8) Deaf or hard of hearing (n=8) Learning/cognitive disability or neurodivergent (n=11) Mental health condition (n=3) Older adults (n=2) Multiple disability (i.e., deaf-blind) (n=3)
Methods (keywords)	usability studies, mixed methods, contextual inquiry, interviews, field observation, qualitative, quantitative, participatory design, ethnomethodologies, exploratory, evaluation/audits, formative needs assessment, summative evaluation, experiments, biometrics, think-aloud, cognitive walkthrough

3.2 Procedure

We conducted single-session, semi-structured interviews, which lasted 60-90 minutes. All interviews were conducted over Zoom or Google Meet. Two authors attended most of the interviews and one facilitated the interview while the other took notes. Each interview began with an introduction and questions about the researchers' background, the methodologies used, and the populations they have worked with. Then, we talked about response bias in three parts:

1. **Defining response bias.** We asked the researchers what response bias meant to them. If they struggled with the term, we shared our working definition: "when a participant's responses are influenced by other factors such as the relationship/interactions between the researcher and participant." Our intentionally vague definition allowed us to gauge the researcher's interpretation of response bias and build a shared vocabulary.
2. **Recalling past studies.** We asked the researchers to recall the last time they conducted a study that involved direct interaction with participants with disabilities (including remote studies). We inquired about their study purpose and design, asked specific questions tying their experiences to response bias (e.g., "How did response bias show up in your research, if at all?", "What did you do about it?") and used follow-up questions to probe into the researcher's experience.
3. **Reflecting on response bias broadly.** We encouraged researchers to reflect broadly on their thoughts and questions around response bias and how it influenced their study design. Example of questions

included: Does response bias affect outcomes? Have you talked about response bias before? What other questions do you have about response bias?

3.3 Data Analysis

All interviews were recorded through the video conferencing software, processed through an automated system (i.e., Zoom transcription or Otter.ai), and validated by two of the authors. After data collection, these two authors individually coded the same interview using inductive thematic coding [9] and then came together to discuss discrepancies. We avoided using any pre-supposed codes and instead let the codes emerge from our data [13]. We continued to code separately and came together after each interview until we had reached intercoder reliability [45] (i.e., 70% alignment in codes) and codebook stabilization (i.e., new codes accounted for fewer than 10% of our total codes). The coders then re-coded earlier interviews and independently coded the remaining 16 interviews. At the end of the coding process, there were 167 codes (e.g., “leveraging response bias”) in 34 broad categories (e.g., “philosophy of response bias”). The two coders affinity diagrammed the codes on Miro, iterating on code groupings and extrapolating themes (e.g., “response bias is everywhere so you have to accept it to some degree”). Then we re-evaluated our emerging themes into the 7 core themes discussed in our findings.

3.4 Positionality

Our team includes authors with rich experience of working with people with disabilities and an author with lived experience of disability. We carefully considered the potential impacts of response bias on our own study. We recruited a diverse set of researchers to interview and created a procedure that focused on concrete examples and avoided overly general responses that may be biased. Moreover, as the first two authors are new to the field and do not have lived experience with disability, we maintained a posture of learning during the interview and the analysis. As fresh eyes to a growth process that has long been in-progress even before our contributions, we hope to encourage further acknowledgement of response bias by sharing stories from the researchers within accessibility research back to the academic community.

4 FINDINGS

Our interviews revealed how the researchers understood response bias in conjunction with other biases and how they considered whether to accept response bias or use strategies to address it.

4.1 Understanding Response Bias

Although some of the researchers were initially unsure of how to define response bias, all were able to pinpoint specific examples where they may have encountered it. Several expressed interest in learning more about response bias, noting that the interview was generative and helped them reflect on their own work. They described factors that directly (classic response bias) and indirectly (other biases) affected responses.

4.1.1 Classic Response Bias

The researchers in our study referred to response bias as a “*systematic skew in responses*” (R8), attributing it to different factors: participant-researcher dynamics, participants’ view of self, and context of accessibility technology.

Participant-Researcher Dynamics. Similar to previous literature on the influence of power dynamics [18,37], the researchers we interviewed reported that trust and rapport between the researchers and participants influenced the openness and honesty of their participants. The researchers spoke about building trust through a shared identity or knowledge and felt that participants often gave more truthful responses when they had more in common with them. One researcher with a disability, R4, even found the comfort to be bidirectional and said, “[*if*] I know I’m about to go to talk to a disabled person...I just feel more comfortable.” This level of comfort was extended to other intersectional, marginalized identities. R6 noticed that their participants with disabilities more consistently pointed

out the rainbow dinosaur pin on their lanyard than other participants, hypothesizing that the participants felt more comfortable since this pin signaled that R6 was part of another marginalized community, the queer community.

The researchers who did not have a shared marginalized identity connected with their participants through “*disability culture knowledge*” (R7) which included accessibility etiquette. R16 emphasized the importance of accessibility etiquette and how it set the tone of the study:

“It goes back to [the] point earlier of making the user feel comfortable. If you [(the user)] go in and your...first experience is that you can’t get through the door because it’s inaccessible because you happen to be in a wheelchair, you’re immediately [going to be like] okay, what else about this is not going to be...for me.” (R16)

The prior engagements participants had with the researcher reduced some forms of response bias and exacerbated others. Several of the researchers struggled to recruit people with disabilities for their research studies and invited participants from the same set of people (e.g., friends, prior participants). The researchers explained that repeat participants had less of a novelty or halo effect, responding quickly with more elaborate and structured feedback. On the other hand, a pre-existing relationship with participants could exacerbate social desirability bias, since there is a “*danger that they don’t actually want to offend...because they feel they’re actually being negative towards you, not the software*” (R16). Some of the researchers who recruited using their own networks worried that their friends would actively support their project and give only positive feedback. R9 preferred not to recruit friends who are already “*shaping how each other sees the world,*” but R7 saw this as part of an “*on-going dialogue*” about each other and their disabilities.

Feelings of coercion during the study also caused participants to say what they felt the researcher or other people in the room wanted to hear. For example, participants who were brought to their study by parents or recruited through an organization experienced the observer effect. R9 described the impact of the external pressure to participate on the children with autism they worked with:

“There’s this gray area of like assent where...their mommies and daddies brought them here. So, they’re doing it and you can tell like maybe they would rather not. But at the same time...they’re still talking to you. So, it’s weird.” (R9)

Participants’ sense of obligation to complete the study also resulted in biased responses. A few of the researchers mentioned instances where participants still insisted on trying to complete the task even when the technology was broken. Even though the researchers tried to be clear, some participants thought completion was a requirement and those who did not finish “*felt like they didn’t deserve [the compensation]*” (R9). R6 hypothesized that this obligation may be “*disproportionate to folks with disabilities*” because of the income inequality and unemployment in those populations.

Participant’s View of Self. According to the researchers, the participants’ comfort levels depended on how they viewed themselves and their abilities, especially in relation to the researcher. Some participants withheld their feedback due to the power dynamic created by the “*charity model of disability*” (R13), which put the researchers in the position of “*helping*” the participants with disabilities. This caused the participants to respond with gratitude or attempts to not offend the researchers. R27 described the positive feedback they received:

“[There was] that level of encouragement, like, ‘Hey, I think what you’re doing is great. It would be great if you keep working on this.’ I do think there is this...kind of like positive feedback that...people like to provide [so] that it leads to...you are working on more and more solutions.” (R27)

Another reason participants held back or downplayed their comments was because they did not feel as qualified as the tech-savvy, expert researchers. R26 talked about how deferring to the researchers’ expertise could exacerbate demand characteristics, where the participants responded that a technology is a good idea simply because the

expert (the researcher) thought it was worth asking about. Particularly for older adults, their “*self-efficacy*” (R19) was reflected in their confidence in giving critical feedback, like the increase of impression management found in older adults [22]. It was not until R19 commented on technologies the older adults do use (i.e., Zoom or Snapchat) that their confidence in their digital skills increased and they owned their answers as experts of their lived experience.

Additionally, the researchers reported how participants felt like they needed to prove themselves and their ability. For example, R24 recalled an instance when students with visual impairment insisted they had used scissors, even though it became clear later on that they had exaggerated their ability. R4 called this tendency of participants to focus on the optimistic or positive aspects of what they can do as “*performing ability*,” or as they described:

“Something I’ve seen participants...do in interviews, as well as observations, is focus on things that they might be really successful at doing to...show that their disability does not...define who they are and that their life is not...made up of deficits, which is a common stereotype — that people with disabilities cannot do things and that their life is less good.” (R4)

R8 related this to a “*can do*” attitude, a mentality of not giving up easily because they had already overcome so many hurdles in the ableist, built world. This persistence was another reason participants continued to use a broken prototype, because they wanted to overcome the challenge or because they blamed the failure on their ability to use it rather than on the technology itself. R6 noted that this learned persistence due to structural ableism could also make acquiescence a reflex for people with cognitive disabilities:

“[For] folks who have a tough time like reading emotion or just kind of processing complex relationship dynamics, there’s sometimes a tendency toward acquiescence...either because they don’t necessarily have a total understanding that we want their honest feedback or that...it’s okay to say bad things. There’s also kind of the desire to be...liked. Like there’s kind of just a reflex for acceptance after a lifetime of feeling othered.” (R6)

Context of Accessibility Technology. The researchers noticed that some participants gave optimistic or generally positive responses because they saw potential in what technology could offer. Sometimes, this overt optimism was because the baseline was so low — any improvement was an upgrade over the frustrating technology they used daily. For example, R16 shared that even though it only takes 3 seconds for a sighted person to put something in their “Shopping Cart,” participants with visual impairments responded positively to it taking 2 minutes because they were used to it taking 5 minutes. Similarly, R10’s participants called the technology “*enabling*” because it allowed them to do something they could not even imagine doing before. The “*science-fiction*” (R21) nature of some of the proposed technologies, such as self-driving cars, also caused the participants to be excited about how new technology could solve all their problems. In these cases, they often gave positive responses even if the technologies proposed and perceived benefits were still decades away.

In contrast, other participants had more measured expectations for what happened to their feedback, causing them to be more cautious or come prepared with an agenda. This agenda could be from frustrations related to “*accessibility dongles*” (R13), or the history of false promises and technology that does not quite meet the needs of people with disabilities (e.g., sign language gloves or high-tech white canes). Some participants expressed their frustrations by not answering the researchers’ questions or giving pessimistic feedback. R6 talked about how their participants were more measured in their responses because they did not trust large companies to have their best interests in mind:

“With the tech companies...it’s not just wanting to make something new or help...You’re responsible for this stuff that’s out in the world...I think it’s true for everybody, but I think it’s amplified for folks with disabilities because they have to often push so hard to get companies to adopt accessible practices when they’re not legally enforced.” (R6)

On the other hand, participants would also respond in more generic, off-handish ways because they trusted themselves more than the technologists. R21 shared how some of their participants who were older were used to “bootstrapping” technology that did not adapt to their needs. These older participants gave broadly affirmative answers, regardless of how well the technology worked for them:

"It turns out there's a lot of difference with age among the disability community and because people have different expectations and different experiences of the world...[The] older generation, maybe they're used to times when ADA wasn't as stringent and so they're just like bootstrapping something together that gets them to the supermarket like...oh yeah, I can...jam my wheelchair between these two seats and... that's how I strap in the car. And the younger generation is more like no, this is wrong, and I want this to be fixed." (R21)

4.1.2 Other Biases

Beyond response bias, the researchers talked about other biases, including which responses they solicited in the first place (researcher bias) as well as which part of the community was able to respond (sampling bias).

Researcher Bias. The way the researchers viewed and positioned themselves in relation to their participants influenced their research direction, which in turn, influenced how the questions were asked and how the responses were interpreted [3,24].

The researchers' direction was sometimes shaped by assumptions of what participants needed. R27 acknowledged that the technology they were evaluating was often “*designed with the intention that people will like it.*” R23 lamented that many participatory design sessions with participants with autism tended to be “*token*” because of the small sample-size and surface-level questions. This meant that the “*end design of what people are building or how they're using it doesn't necessarily reflect or seem to me like it was that informed by the input that they received*” (R23). This was especially harmful for participants with disabilities because it led to reinforcement of structural ableism. R9 talked about the potential impact of researcher bias on accessibility technology:

"There's sort of an unwillingness to engage with the internalized ableism of participants and how they may be expressing that through their desires.... It's so complicated...Because [researchers] want a fast answer that is basically like...absolution that...what they're doing is fine and that they're good people. And it's like, I'm not saying that you're not good people. I'm saying if you think a little bit more deeply about your research inquiry, you might come up with different projects and it might be very interesting." (R9)

Moreover, the preconceived research directions influenced how the researchers asked questions and how the participants answered. For example, R4 talked about how positioning themselves as someone who knew more than the participants themselves could affect the participants' responses:

"I do think that there could be maybe more thought in terms of the way that we frame our research and the things that we decide to work on, like...whether we're positioning ourselves as saviors who know more than the actual disabled people themselves... And the way that could relate to response bias is...if a participant thinks that a researcher has that attitude...maybe that will change the way that they respond. Or... [it could] affect the questions that are asked, which would then affect their responses." (R4)

How questions were worded led participants to validate the researchers' assumption on what the participant needed. When R21 was trying to optimize the degree of inclination of the on-ramp for a vehicle, they were asking participants “Do you need a ramp at 9 degrees to enter the vehicle?” and participants were responding yes. However, when they switched it to “What is the ideal way for you to get into a vehicle?”, participants responded that they would not prefer a ramp at all, rather, a chair lift.

Finally, the researchers' assumptions of the participants' abilities affected the interpretation of participants' responses and behaviors. For example, R9 talked about how another researcher assumed a child with autism was

not complying because they did not have any interactions recorded by the audio transcript. However, the child was interacting with the therapist nonverbally. That researcher's assumptions on communication and ability inhibited them from seeing a key interaction in the video recording. R15 shared another example where assumptions of the participants' ability resulted in a misinterpretation that blamed the neurodivergent participant rather than the system:

"One of the headsets died [and] ran out of battery. And so [the participant] was not teleporting with the group like we were expecting. And we read it as he can't keep up. That's how we interpreted it and it turned out his system was dead. And we didn't know it until like three minutes in. And so here we made this assumption... [and w]e were attributing the miscommunication to him, rather than to the system. And that led us to... build functionality that makes this stuff clear...So I think it was a really powerful lesson of like [how] we attributed that problem to [the participant] and how we need to do way better." (R15)

Sampling Bias. Sampling bias refers to who is participating in the study due to "decisions by analysts or data processors" or "self-selection by the individuals or data units being investigated" [23,29,66]. The researchers we interviewed were concerned with both sides of the equation for sampling bias: which participants researchers were looking for and which participants were able to and interested in attending.

The researchers noted that the "ability to respond" (R6) is especially important for participants with disabilities not only due to concerns about resources, time, and awareness, but also whether the recruitment or study methodology itself was accessible to them. For instance, to recruit diverse participants within the deaf and hard of hearing community, R7 adjusted their recruitment method to ensure it was accessible:

"We had to do an amendment to our IRB...to say there's structural barriers to doing things like this... [We had to] think about how we are going to do the recruiting message [and]... hire a third party ASL interpreting service...[to] video record our recruitment message so that we're sending the signal right off the bat that we're actually going to be accessible to people." (R7)

The language the researchers used to reach out to participants, especially related to disability identity, was another aspect that determined the types of participants who came to the studies. For example, the researchers postulated that participants who were recruited for their disability identity have a level of confidence with that identity and are more likely to be involved in disability activism. This effect was even stronger if the researchers themselves had a reputation for doing work in disability or their own networks already had disability awareness. For example, R9 had to investigate more obscure corners of the Internet to try to find participants who "aren't already on the disability pride train." R4 explained this phenomenon and how language helped with recruiting diversely:

"I want to do a better job recruiting people that are not who aren't like disabled activists...You know my samples tend to be white and cisgender...I'm trying to do better about being more holistic and recruiting people with the diversity of identities and then also people who don't necessarily use the language that I use." (R4)

This perception of their own disability identity affected participants' responses. For instance, R5 recounted how a participant who led disability justice work in their high school was quicker to start prototyping in the design process because they were "positive about the creative process around [their] disability" (R5).

Another example of sampling bias was the tech-savviness of participants. The researchers explained that participants who self-selected into studies tended to be tech-savvy. While some of the researchers intentionally sought out tech-savvy participants because their feedback was more detailed, more actionable, and "less sugar-coated" (R22), others worried that it skewed results to be overly optimistic about the technology or led to technology design that did not address the needs of less tech-savvy participants. R16 explained the importance of making sure to consider this wide range of technical abilities among the participants:

"Somebody who's highly technical gets the job done because they really know their screen reader inside out and the shortcuts... We [still] have to look at providing another layer of interaction here for the people who don't necessarily know the screen reader as well as the person with the in-depth technical knowledge." (R16)

Lastly, some of the researchers noted how there is diversity even within disability identities, which made it difficult to recruit a potential user of a prototype. For example, R2 explains that *"low vision is a vaguely defined term,"* so the prototype may not be designed for someone with very low vision, while also not being helpful for someone who is *"too capable to evaluate the prototype."*

4.2 Addressing Response Bias

The researchers discussed whether to simply acknowledge response bias (considerations for accepting it) or to further understand and mitigate response bias (strategies for addressing it).

4.2.1 Considerations for Accepting Response Bias

Some of the researchers accepted response bias because they perceived it as an inherent reality of conducting research with human participants. One attitude was that *"everything is biased...because it goes through a human"* (R15). Others echoed this sentiment and added that they *"embrace[d] our humanity and biases"* (R7), taking on a more cultural anthropologist perspective in exploring biases instead of immediately eliminating them. Similarly, some of the researchers acknowledged the influence of innate personalities or disability identities, such as the *"dichotomy between the participants with Down Syndrome who were more likely to please [the] investigator and participants with autism who are honest and say I don't want to use that at all"* (R1). The researchers found it difficult to address these inherent differences and the resulting biased responses.

The researchers also expressed difficulty in identifying response bias and to what degree it is present. During our interview, they provided disclaimers when recalling instances of response bias, emphasizing that this was their interpretation of what happened. Some of the researchers expressed concerns with identifying instances of response bias. R18 explained that they take responses at face value because it's their *"job to believe what participants are saying."* Moreover, R19 was concerned that identifying response bias implied that the researchers knew better than the participants, which could reinforce power differentials and structural oppression:

"It can be hard to say like... this is response bias because that's basically...enacting a form of epistemic violence or injustice, where you're saying that, oh, this person doesn't know what they're talking about. I hesitate to... [say] that one case was definitely response bias." (R19)

On the other hand, some of the researchers accepted response bias because they already did their best to mitigate it. R3 admitted that they were not worried about biases because *"when you design your study, you've already tried your best to remove all the potential possibilities to generate biases and those left are people's personalities."* R6 expressed that *"there's not a clear path to not having [response] bias without very systemic change in how everything works,"* so they focused less on addressing response bias and more on figuring out how much it interfered with the validity and generalizability of their findings.

Lastly, some of the researchers did not address response bias because of the tradeoffs in the *"publish or perish"* (R16) environment in academia and the fast-paced nature of industry. R15 lamented that *"the pressure on me to get a good response is tremendous and it just trickles down to the participants."* R2 felt that in industry, there was more response bias because the teams wanted to confirm planned features to ship the product fast. R6 shared their industry and academia experiences given the limited participant pool and few dedicated resources in accessibility research:

"When I was in graduate school and [doing] academic research there, a lot of the constraints are around like time and resourcing...If you think you're experiencing response bias...you're going to lean really heavily toward keeping

the data because you are so risk averse of throwing away any data no matter how biased it might be, just as a virtue of necessity. And then in industry...you might have more resources to spend on...options to try to reduce bias...but you have to convince someone that it's a good use of resources." (R6)

4.2.2 Strategies for Addressing Response Bias

Others emphasized the importance of addressing response bias because it may lead to inaccurate responses. R24 extended this interpretation saying studies influenced by response bias are *"not valid because the people are trying to appease the intention of the study, as opposed to speaking their truth or sharing their understanding."* Some of the researchers called out how these biased responses may lead to wasted effort and time. R16 reflected on a previous study where they received positive results due to participants being nice, later realizing that the technology wasn't actually helpful. If they had considered response bias, they would *"have abandoned [the study] about six to eight months earlier"* (R16).

Preventing. The researchers talked about the importance of working with the disability community to address some of the potential sources of response bias in the planning stages. The researchers found that working with the disability community helped them build rapport with their participants. For example, R7 shared disability culture knowledge to reassure participants they are not *"judging them and assessing them as a 'sightie.'"* R17 shared their previous research experiences to demonstrate an ability to relate to the participants' stories *"not in an embodied sense, but the fact that I might have heard some of those stories previously."* Others encouraged rapport by working with specific partners in the community. For example, R20 worked with research assistants who could speak ASL to account for cultural aspects of the D/deaf community. R1 worked with collaborators from their partner school since the participants were already familiar with them. While having these collaborators gave expertise, R1 cautioned that having them in the room influenced whether the participant experienced observer effects or if their responses were lost in translation.

Moreover, the researchers also worked with the community to identify methodologies that fit the needs of their participants. R13 modified an activity after learning on Twitter that brainstorming could be overwhelming for those with high anxiety and R9 used online surveys to both reduce the observer effect and allow participants to have more time to respond, extending the *kariotic space* [66]. The researchers also conducted piloting or *"peer testing"* (R18) to catch issues with the study protocol. One potential issue as R16 mentioned is not accounting for the cognitive overload of blind participants keeping track of survey options in their head. The pilots and check-ins with the community also curbed issues related to communication. R25 pointed out that the direct translation in ASL of *"What do you think about the technology?"* would be a leading question: *"Do you like the technology?"*. Even having an interpreter allowed the researcher to be *"flexible with a communication modality,"* so the participant could communicate effectively (R9). Others adapted existing practices like Q-methodology or rapid-fire yes or no questions to elicit more structured feedback. Others used *"plain language"* (R13) in the consent form to ensure participants with disabilities understood they are not expected to complete the study. And some highlighted that *"we [(researchers)] are not testing you [(participants)] ...we're working as a collaborative unit"* (R16) to give participants confidence in their opinions.

While these relationships with the community were essential to the researchers reducing response bias, the researchers felt it was important to ensure the relationship is reciprocal. R7 engaged the concept of *"interdependence"* from disability studies [30] to talk about the importance of reciprocity:

"As researchers, we can kind of go into a community and really extract, take out all this knowledge and take their time and okay, we'll throw a little gift card at you. And then, sometimes, we don't actually end up building robust technology that is based on the...valuable insight they give us. And what's the community left with...There's actually an argument for it being the more ethical thing to do to [and] actually bring your full person into those conversations and... give back in multiple ways." (R7)

Reacting. The researchers talked about the importance of being critical of the overly positive responses from participants. R26 suggested asking follow-up questions that are concrete because “*at the abstract level, lots of things seem...sensible.*” R8 recounted how earlier in their research career, participants older than them gave overly positive responses to encourage them and through follow-up questions, they realized that “*I think this could be really great for some people*” did not mean that the participants themselves would use it. To catch similar contradictions, others suggested asking the question again in a different way or actively soliciting negative feedback.

The researchers also discussed being critical about ways the charity model of disability [27] could influence their mindset and research approach. To counter this influence, R19 emphasized starting each study with a clean slate and “*not com[ing] in with the preconceived notion of... what needs to be created or designed.*” R11 and R1 noted the importance of this in the disability space to get a grasp of the participants’ strong individual needs. During the study, some of the researchers were open to potential off-topic responses that may result from participants’ pre-existing agendas and frustrations with existing technology, building trust with the participants. R21 talked about a time where a participant wanted to talk at length about privacy, even though that was not the intended topic of discussion. R21 wanted to respect the participant’s responses because:

“There aren’t many people in my position with...a really key line into development [that participants with disabilities can come in contact with] So, I have to be like okay this person really has a point to...get across. And...they feel that stronger than any of the questions that I’m asking, and that’s...perfectly valid.” (R21)

Keeping an open mind about feedback also helped some of the researchers think about their own bias. R13 reflected on their own research direction when they received feedback from participants who were frustrated about accessibility technology beyond the study saying, “*people have tried this, and it never gets anywhere*” or accused them of “*scoping too small of trying to fix this small problem rather than addressing a systemic issue.*”

The researchers felt it was difficult but also crucial to reflect and identify biases that may have occurred. For example, R22 used visualizations to identify outliers in self-reported quantitative measures, such as task satisfaction, usability score, and confidence. R12 and R9 video recorded sessions with participants and watched them later to check all the ways (e.g., facial expressions, body language) participants might be communicating. R14 conducted follow-up interviews in ASL to ensure that the survey responses in English matched participants’ responses in ASL. Others did “*member-checking*” (R18) or even “*co-research*” (R24) to ensure proper interpretation of the participants’ responses.

One key way the researchers addressed and reacted to response bias was contextualizing the bias, especially when there is a small sample size. R8 suggested identifying how susceptible a participant is to response bias through objective measures designed for participants with disabilities (e.g., self-efficacy), extending on the MMPI [25]. Others like R24 tried to understand the sociocultural factors that may have influenced participants’ responses, tying in concepts like activity theory and actor-network theory. Several of the researchers also reflected in the reporting stage on how their background and positionality could impact interpretation of participants’ responses. Some of them included a stance statement, indicating “*this is how I’m interpreting my data; and here’s who I am,*” so that other researchers can “*read between the lines and...make their own interpretations*” (R15). In industry, the researchers educated their co-workers about potential response bias and over-generalization of claims made from the data. For instance, R11 warned product teams saying “*80% of blind people prefer this type of technology*” for a 10-person study could be misleading given the denominator.

Table 3: Summary of Findings: Factors That Bias Responses in Accessibility Research

Factors	Manifestations	Mitigations
Researcher-Participant Dynamics	<ul style="list-style-type: none"> ● Shared experiences with disability (through identity or knowledge) ● Prior relationships with the researcher or study 	<ul style="list-style-type: none"> ● Developing reciprocal relationships to both learn/give to the community (disability culture knowledge, accessibility etiquette)

	<ul style="list-style-type: none"> ● Feelings of coercion during the study (how they were recruited, obligation to complete the study, incentives) 	<ul style="list-style-type: none"> ● Working with community partners and being careful of who is in the room ● Making sure the research methods are accessible (cognitive overload)
Participant's View of Self	<ul style="list-style-type: none"> ● Whether the participant feels that the researcher is the expert or in a helping role (self-efficacy, charity model of disability) ● Participants' desire to prove themselves or their ability (performing ability, "can-do" attitude) 	<ul style="list-style-type: none"> ● Framing the study to encourage collaboration or focus on what the participant is bringing (pointing out participants' expertise) ● Including flexible/multiple communication modalities (ASL interpreter, verbal/nonverbal) ● Being clear participants can leave whenever they need to (plain language consent)
Context of Accessibility Technology	<ul style="list-style-type: none"> ● Whether participants feel optimistic about how the technology could improve or what it could achieve ● Whether participants are frustrated with how accessibility technology has not met their needs (accessibility/disability dongles) 	<ul style="list-style-type: none"> ● Setting expectations of what might happen to the feedback ● Being open to what agendas or frustrations participants might want to bring ● Asking follow-up questions for more concrete feedback ● Being clear when presenting takeaways to stakeholders
Researcher Bias	<ul style="list-style-type: none"> ● The researchers' view of the participants and their role affected research direction (charity model of disability) ● How researchers asked questions and how participants answered ● Assumptions of participants' abilities affected interpretation of responses/behaviors 	<ul style="list-style-type: none"> ● Piloting questions/design before conducting the study ● Going in with a clean slate to understand the participant's experiences ● Member-checking interpretations of results or co-research ● Sharing a stance statement
Sampling Bias	<ul style="list-style-type: none"> ● Accessibility of recruitment methods (ability to respond) ● Language researchers used regarding participants' disability identity ● Range in tech-savviness/disability identity 	<ul style="list-style-type: none"> ● Carefully defining inclusion criteria ● Thinking intentionally about language used in study notices and which communities can access them

5 DISCUSSION

While the accessibility field is increasingly bringing in more people with disabilities to evaluate new designs and interventions, there is a need to question how their voices are being represented or distorted in the research. Response bias is one way in which voices can be distorted, because of study design and participant-researcher dynamics. We connected the conversation about response bias [10,18,52] to research in accessibility through the unique power dynamics between researchers and participants [11,66] and research settings that require adapted methods for recruiting participants and running studies [25,64]. Through interviews with 27 researchers about how response bias manifested and how they mitigated it, we shed light on the duality of response bias as a factor that may skew participant's answers further from the truth and a possible method to invalidate or silence the participant's voice. We also discuss response bias more comprehensively with researcher bias and sampling bias. Therefore, our study makes the following contributions:

- **Observations and practices specifically from accessibility technology**, building upon existing work on the types of response bias and how they are handled to see how characteristics specific to accessibility research (i.e., accessibility history, societal ableism) may amplify or change response bias.

- **Interactions of response bias with other biases**, extending the conversation about response bias beyond “classic” factors that directly impacted participants’ responses, including how sampling bias might change the set of responses that is considered or how researcher bias could change how responses are solicited or interpreted.
- **Perceptions and tensions around response bias**, highlighting how accessibility researchers from different disciplines think about response bias and the tradeoffs of accepting or addressing it.

5.1 Tradeoffs and Tensions

In this paper, we present knowledge that would otherwise be shared informally or siloed in different fields. While the researchers we interviewed expressed interest in learning more, many were not sure where to start, saying they wished there were scorecards or checklists that aggregated best practices. Some felt lucky to have worked with and learned from scholars with disabilities, but with the growing popularity of accessibility research [44], it is neither scalable nor fair to put the labor of distributing knowledge on scholars with disabilities [68]. The researchers talked about how their backgrounds influenced their exposure to understanding and addressing response bias, varying across methodologies, settings, and disciplines (see Table 2). Some researchers, including R15 and R23, talked about how they were trying to bridge the gap between the methodology and purposes of study in different disciplines and approaches.

These varied perspectives from disciplines and settings raised philosophical and practical concerns of whether to accept or address response bias [24,42] — some of the researchers tried to eliminate all possible biases to get the objective truth while others were open to new directions the bias may point to. There is also the tension of labelling a participant’s response as biased and leading to epistemic injustice or violence through erasure of their lived experience [68]. Our findings demonstrate the importance of intentionally thinking about response bias early in the research process because of the researchers’ degrees of freedom in considering whether response bias is present and how they would like to address it, if at all.

How the researchers approached response bias depended on the context and tradeoffs around each individual project. These included external pressures to publish or perform as well as resource constraints in time or money, especially for projects in specialized fields like accessibility technology. Given these tradeoffs, it is especially important to consider factors related to response bias in accessibility technology, including but not limited to participant-researcher relationships based on the researcher’s and participants’ perceptions of their own and each other’s identity (i.e., the charity model of disability and structural ableism, length and depth of relationships with the community [2,11,66]) and the specific methodological considerations for the field (i.e., smaller sample size due to difficulty in recruiting, adjusted means of asking questions and communicating [56,64]).

5.2 Guidelines

The implications of response bias are magnified within accessibility, adding to existing technologies that are either unwanted or unhelpful and perpetuating ableist thoughts. We present guidelines in hopes of critically evaluating research and considering response bias holistically and intentionally.

Research Question. Some research questions may inherently allow more room for bias than others. For example, the sensitivity of the topic could result in evasive or misleading answers. In accessibility research, this includes questions about participants’ abilities, which could be influenced by their own perception of their disability identity or whether they feel they need to perform their ability. Another consideration is whether questions are abstract (i.e., “Will the participants use the technology?”) or more concrete (i.e., “How do the participants use the technology?”). Abstract questions could result in participants answering based on their current information, which may contradict with their future actions [10,34]. In addition, the research question could be impacted by the researchers’ view on the charity model of accessibility, which in turn impacted participants’ answers [29]. To reduce that effect, the researchers could “interview” themselves [15] to consider their own positionality and standpoint [4,63].

Methods. Different methods also have varying inclinations towards response bias. Even quantitative measurements could be influenced by researcher bias, both in terms of which metrics are selected and how those metrics are interpreted, especially given a small sample size. Some methods raised in our findings include framing, environment, relationships, and question-framing. For framing, the way the study is presented can take into account power dynamics, the context of accessibility technology, and the charity model of disability (e.g., sending the consent form ahead of time using plain language, setting expectations on whether the technology is exploratory, encouraging participants to recognize their expertise in their lived experience) [66]. Creating a respectful environment, especially around accessibility etiquette, is important for more comfortable participant-researcher dynamics (e.g., asking someone with visual impairment if or how they wish to be guided, considering accessibility of the study space and providing relevant instructions) [11]. This etiquette could be learned through relationships with the disability community reciprocally, from short-term rapport through solidarity to longer-term engagement with specific participants (e.g., member-checking findings, communicating findings back to the community) [2]. Finally, connections with the community also shape how the participant can communicate (e.g., considering cognitive load of questions, providing flexible means of communication like an ASL interpreter) [64].

Analysis. Even if the researchers aim to have a perfectly rigorous study, there may still be response bias. Therefore, it is important to look at whether response bias has manifested during the study, especially given the diversity of individual experiences of participants with disabilities. Some ways in which response bias may manifest include where participants may contradict themselves (e.g., when the participant says the technology is great for someone else but not for themselves) [34]; when the participants answers in the abstract rather than the concrete (e.g., when participants say they like a technology without articulating why); how extreme the participants' responding might be (e.g., when a participant seems to be extremely positive and encouraging) [26]; who else is in the room (e.g., whether a superior from the organization or parent was in the room and how that may have influenced the responses); and which means of communication are being considered from the data (e.g., considering both audio and visual outputs). Moreover, there are external and structural factors that could influence the participants' responses, such as the context of accessibility technology or structural ableism. It is important to both analyze results given these considerations as well as report the context and possible response biases that may occur.

5.3 Limitations and Future Work

We hope this paper is a starting point for further research on response bias in accessibility. In this study, we were limited to the researcher's self-reported experiences with response bias. Future work should contextualize these findings by incorporating participants' experiences in these studies and quantitatively experimenting on specific instances or strategies for response bias. For example, a follow-up experiment could measure to what extent does shared identity such as disability between the researcher and participant affect response bias. We also see value in continuing the conversation in more interactive ways like compiling collaborative resources for the researchers working with participants with disabilities in a living document, like the one on COVID-safe research practices [43].

These conversations about response bias also raise a broader discussion around participant engagement in accessibility research. Response bias is only one of many factors that influences whether participants are meaningfully engaged in the research process [11,66]. The researchers discussed how labelling and mitigating response bias raised ethical considerations (e.g., epistemic violence or injustice, reciprocal relationships over longer-term studies) and made them re-evaluate their own standpoints.

Another area of discussion is the uniqueness of accessibility research and the applicability of the findings to other fields. Many instances of response bias and the mitigation strategies that the researchers used could broadly apply to human-computer interaction research, including practices like asking follow-up questions, getting clear consent, or using multiple methods to triangulate. It would be beneficial to further understand the manifestations and mitigations of response bias that are documented in our study and how it may differ across other marginalized populations [18,62].

6 CONCLUSION

Our research aimed to understand how accessibility technology researchers experienced response bias, how it manifested in their research, and how they mitigated it. Through interviews with 27 researchers, we elicited specific examples of response bias in accessibility research influenced by participant-researcher dynamics, the participants' view of self, and the broader context of accessibility technology. We found that response bias interacted with other biases or limitations to influence responses, expanding the discussion beyond the "classic" definitions of response bias from literature. Moreover, we uncovered the researchers' considerations of whether to accept or address it. We share these stories of the researchers' own experiences and reflection on biases with the intention of starting a conversation. Contributing to what we hope will be a continued conversation on response bias, we present guidelines on how to think about response bias in accessibility research. Future exploration could unpack more specific dynamics between different manifestations and mitigations of response bias and considerations that could be interesting for HCI researchers working with other marginalized populations.

ACKNOWLEDGMENTS

First and foremost, we would like to thank all the researchers we interviewed for sharing their insights and experiences. We would also like to thank those who helped shape this research project and review the paper, specifically Ed Cutrell, John Tang, and Merrie Morris from Microsoft Research's Ability team.

REFERENCES

1. Judd Antin and Aaron Shaw. 2012. Social desirability bias and self-reports of motivation: a study of amazon mechanical turk in the US and India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 2925–2934.
2. Fabricio E. Balcazar, Christopher B. Keys, Daniel L. Kaplan, and Yolanda Suarez-Balcazar. 1998. Participatory action research and people with disabilities: Principles and challenges. *Canadian Journal of rehabilitation* 12: 105–112.
3. Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, 1301–1310.
4. Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 675–684.
5. Brianna Blaser and Richard E. Ladner. 2020. Why is Data on Disability so Hard to Collect and Understand? *Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)* 1. Retrieved from http://respect2020.stcbp.org/wp-content/uploads/2020/08/Research_1_RESPECT_2020_Blaser_Ladner.pdf
6. Susanne Bødker and Morten Kyng. 2018. Participatory Design that Matters—Facing the Big Issues. *ACM Trans. Comput.-Hum. Interact.* 25, 1: 1–31.
7. Mayara Bonani, Raquel Oliveira, Filipa Correia, André Rodrigues, Tiago Guerreiro, and Ana Paiva. 2018. What My Eyes Can't See, A Robot Can Show Me: Exploring the Collaboration Between Blind People and Robots. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*, 15–27.
8. Glenn H. Bracht and Gene V. Glass. 1968. The External Validity of Experiments. *American educational research journal* 5, 4: 437–474.
9. V. Braun and V. Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*. Retrieved from <https://www.tandfonline.com/doi/abs/10.1191/1478088706QP0630A>
10. Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 1657–1666.
11. Emeline Brulé and Katta Spiel. 2019. Negotiating Gender and Disability Identities in Participatory Design. In

- Proceedings of the 9th International Conference on Communities & Technologies - Transforming Communities (C&T '19)*, 218–227.
12. Loïc Caroux, Charles Consel, Lucile Dupuy, and Hélène Sauzéon. 2014. Verification of daily activities of older adults: a simple, non-intrusive, low-cost approach. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility (ASSETS '14)*, 43–50.
 13. Kathy Charmaz. 2011. Grounded Theory Methods in Social Justice Research. In *The SAGE Handbook of Qualitative Research*, Norman K Denzin Yvonna (ed.). SAGE Publications, 359–380.
 14. A. Chavan and E. Schaffer. 2004. The Bollywood method. *E. Schaffer* 5: 129–130.
 15. Ronald J. Chenail. 2011. Interviewing the investigator: Strategies for addressing instrumentation and researcher bias concerns in qualitative research. *Qualitative report* 16, 1: 255–262.
 16. Sasha Costanza-Chock. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need (Information Policy)*. The MIT Press.
 17. Mick P. Couper, Roger Tourangeau, and Darby M. Steiger. 2001. Social presence in Web surveys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, 412–417.
 18. Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. “Yours is better!”: participant response bias in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 1321–1330.
 19. Barbara Snell Dohrenwend, John Colombotos, and Bruce P. Dohrenwend. 1968. Social distance and interviewer effects. *Public opinion quarterly* 32, 3: 410–422.
 20. Katherine Pratt Ewing. 2006. Revealing and concealing: Interpersonal dynamics and the negotiation of identity in the interview. *Ethos* 34, 1: 89–122.
 21. Robert J. Fisher. 1993. Social Desirability Bias and the Validity of Indirect Questioning. *The Journal of consumer research* 20, 2: 303–315.
 22. Rachel L. Franz, Ron Baecker, and Khai N. Truong. 2018. “I knew that, I was just testing you.” *ACM transactions on accessible computing* 11, 3: 1–23.
 23. Christopher Frauenberger. 2015. Disability and Technology: A Critical Realist Perspective. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*, 89–96.
 24. Christopher Frauenberger. 2016. Critical Realist HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*, 341–351.
 25. H. E. Freeman, K. J. Kiecolt, W. L. Nicholls 2nd, and J. M. Shanks. 1982. Telephone sampling bias in surveying disability. *Public opinion quarterly* 46, 3: 392–407.
 26. Adrian Furnham. 1986. Response bias, social desirability and dissimulation. *Personality and individual differences* 7, 3: 385–400.
 27. Roger O. Gervais, Yossef S. Ben-Porath, Dustin B. Wygant, and Paul Green. 2007. Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment* 14, 2: 196–208.
 28. Wayne D. Gray and Marilyn C. Salzman. 1998. Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods. *Human-Computer Interaction* 13, 3: 203–261.
 29. Giampiero Griffo. 2014. Models of disability, ideas of justice, and the challenge of full participation. *Modern Italy: journal of the Association for the Study of Modern Italy* 19, 2: 147–159.
 30. Gordon H. Guyatt and Jason W. Busse. 2006. The Philosophy of Evidence-Based Medicine. In *Evidence-Based Endocrinology*, Victor M. Montori (ed.). Humana Press, Totowa, NJ, 25–33.
 31. James J. Heckman. 1979. Sample Selection Bias as a Specification Error. *Econometrica: journal of the Econometric Society* 47, 1: 153–161.
 32. Gary Hsieh and Rafał Kocielnik. 2016. You Get Who You Pay for: The Impact of Incentives on Participation Bias. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*, 823–835.
 33. Jeffrey L. Jenkins, Joseph S. Valacich, and Parker A. Williams. 2018. Human-Computer Interaction Movement Indicators of Response Biases in Online Surveys. In *38th International Conference on Information Systems: Transforming Society with Digital Innovation, ICIS 2017*. Retrieved from <https://arizona.pure.elsevier.com/en/publications/human-computer-interaction-movement-indicators-of-response-biases>
 34. Colin Jerolmack and Shamus Khan. 2014. Talk Is Cheap: Ethnography and the Attitudinal Fallacy. *Sociological methods & research* 43, 2: 178–209.

35. John M. Johnson and Timothy Rowlands. 2012. The interpersonal dynamics of in-depth interviewing. *The SAGE handbook of interview research: The complexity of the craft*: 99–113.
36. S. R. G. Jones. 1992. Was There a Hawthorne Effect? *The American journal of sociology* 98, 3: 451–468.
37. Sin-Hwa Kang and Jacquelyn Fort Morie. 2013. Users' socially desirable responding with computer interviewers. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*, 229–234.
38. Orit Karnieli-Miller, Roni Strier, and Liat Pessach. 2009. Power relations in qualitative research. *Qualitative health research* 19, 2: 279–289.
39. Timothy Kelley, L. Jean Camp, Suzanne Lien, and Douglas Stebila. 2012. Self-identified experts lost on the interwebs: the importance of treating all results as learning experiences. In *Proceedings of the 2012 Workshop on Learning from Authoritative Security Experiment Results (LASER '12)*, 47–54.
40. Ivar Krumpal. 2013. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & quantity* 47, 4: 2025–2047.
41. Calvin A. Liang, Sean A. Munson, and Julie A. Kientz. 2021. Embracing Four Tensions in Human-Computer Interaction Research with Marginalized Peopl. *ACM transactions on computer-human interaction: a publication of the Association for Computing Machinery*. Retrieved January 23, 2021 from <https://calvliang.github.io/projects/TOCHI2021>
42. Ann Chih Lin. 1998. Bridging positivist and interpretivist approaches to qualitative methods. *Policy studies journal: the journal of the Policy Studies Organization* 26, 1: 162–180.
43. D. Lupton. 2020. Doing fieldwork in a pandemic.
44. Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by “Accessibility Research”? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. *arXiv [cs.HC]*. <https://doi.org/10.1145/3411764.3445412>
45. Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW: 1–23.
46. Alexander G. Mirnig, Magdalena Gärtner, Alexander Meschtscherjakov, and Manfred Tscheligi. 2020. Blinded by novelty: a reflection on participant curiosity and novelty in automated vehicle studies based on experiences from the field. In *Proceedings of the Conference on Mensch und Computer (MuC '20)*, 373–381.
47. Maletsabisa Molapo, Melissa Densmore, and Limpho Morie. 2016. Apps and Skits: Enabling New Forms of Village-To-Clinic Feedback for Rural Health Education. In *Proceedings of the 7th Annual Symposium on Computing for Development (ACM DEV '16)*, 1–10.
48. Torin Monahan and Jill A. Fisher. 2010. Benefits of “observer effects”: Lessons from the field. *Qualitative research: QR* 10, 3: 357–376.
49. Vivian Genaro Motti and Anna Evmenova. 2020. Designing Technologies for Neurodiverse Users: Considerations from Research Practice. In *Human Interaction and Emerging Technologies*, 268–274.
50. Anton J. Nederhof. 1985. Methods of coping with social desirability bias: A review. *European journal of social psychology* 15, 3: 263–280.
51. Lene Nielsen, Lars Rune Christensen, and Anne Sabers. 2017. Do we have to include HCI issues in clinical trials of medical devices? a discussion. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (OZCHI '17)*, 352–355.
52. Martin T. Orne. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *The American psychologist* 17, 11: 776–783.
53. Delroy L. Paulhus. 1991. Measurement and control of response bias. *Measures of personality and social psychological attitudes*. 753: 17–59.
54. Stephen R. Porter, Michael E. Whitcomb, and William H. Weitzer. 2004. Multiple surveys of students and survey fatigue. *New directions for institutional research* 2004, 121: 63–73.
55. Matthew J. Salganik. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.
56. Andrew Sears and Vicki Hanson. 2011. Representing users in accessibility research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 2235–2238.
57. Laure M. Sharp and Joanne Frankel. 1983. Respondent Burden: A Test of Some Common Assumptions. *Public opinion quarterly* 47, 1: 36–53.
58. Eleanor Singer, John Van Hoewyk, Nancy Gebler, and Katherine McGonagle. 1999. The effect of incentives on response rates in interviewer-mediated surveys. *Journal of official statistics* 15, 2: 217.

59. Charlotte G. Steeh. 1981. Trends in Nonresponse Rates, 1952–1979. *Public opinion quarterly* 45, 1: 40–57.
60. Shari Trewin, Diogo Marques, and Tiago Guerreiro. 2015. Usage of Subjective Scales in Accessibility Research. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*, 59–67.
61. Phil Turner and Susan Turner. 2009. Triangulation in practice. *Virtual reality* 13, 3: 171–181.
62. Aditya Vashistha, Fabian Okeke, Richard Anderson, and Nicola Dell. 2018. "You Can Always Do Better!": The Impact of Social Proof on Participant Response Bias. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 1–13.
63. Katherine Welton. 1998. Nancy Hartsock's Standpoint Theory. *Women & politics* 18, 3: 7–24.
64. Mary Wickenden. 2011. Whose Voice is That?: Issues of Identity, Voice and Representation Arising in an Ethnographic Study of the Lives of Disabled Teenagers who use Augmentative and Alternative Communication (AAC). *Disability studies quarterly: DSQ* 31, 4. <https://doi.org/10.18061/dsq.v31i4.1724>
65. Parker A. Williams, Jeffrey Jenkins, Joseph Valacich, and Michael D. Byrd. 2017. Measuring Actual Behaviors in HCI Research – A call to Action and an Example. *AIS Transactions on Human-Computer Interaction* 9, 4: 339–352.
66. Rua M. Williams and Juan E. Gilbert. 2019. "Nothing About Us Without Us" Transforming Participatory Research and Ethics in Human Systems Engineering. In *Advancing Diversity, Inclusion, and Social Justice Through Human Systems Engineering* (1st ed.), Rod D. Roscoe, Erin K. Chiou and Abigail R. Wooldridge (eds.). CRC Press, Boca Raton: CRC Press, 2020., 113–134.
67. Christopher Winship and Robert D. Mare. 1992. Models for Sample Selection Bias. *Annual Review of Sociology* 18: 327–350.
68. Anon Ymous, Katta Spiel, Os Keyes, Rua M. Williams, Judith Good, Eva Hornecker, and Cynthia L. Bennett. 2020. "I am just terrified of my future" — Epistemic Violence in Disability Related Technology Research. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*, 1–16.