Participatory Design and Power in Misinformation, Disinformation, and Online Hate Research

Joseph S. Schafer*
Human-Centered Design &
Engineering, University of
Washington; Center for an Informed
Public, University of Washington
schaferj@uw.edu

Kate Starbird
Human-Centered Design &
Engineering, University of
Washington; Center for an Informed
Public, University of Washington
kstarbi@uw.edu

Daniela K. Rosner
Human-Centered Design &
Engineering, University of
Washington
dkrosner@uw.edu

ABSTRACT

As a research tradition, participatory design (PD) tends to focus on power dynamics where researchers hold greater power than participants. This paper uses design fiction to consider what this tendency overlooks by examining settings where participants may exist in multiple power relationships simultaneously implicated by the research, specifically focusing on the contexts of misinformation, disinformation, and online hate (M/D/OH). Drawing from existing literature in M/D/OH, we present a series of imaginary method abstracts that prompt questions for researchers to reflect on as they adapt PD techniques for new, different contexts. We highlight three value tensions—authenticity, reciprocity, and impact—integral to sustaining a concern for responsibility in PD scholarship. We end with reflections and potential considerations for responsibly applying PD and design fiction methods in M/D/OH settings.

CCS CONCEPTS

• **General and reference**; • **Cross-computing tools and techniques**; • **Design**; • **Human-centered Computing**; • **Collaborative and social computing**;

KEYWORDS

Participatory Design, Misinformation, Value Tensions, Design Fiction

1 INTRODUCTION

Participatory methods for interactive systems design reflect rich traditions of informing impactful research, and can serve as important pathways to broaden our understanding of challenging societal issues. When deployed with care, these methods have long allowed

*Corresponding author.

academics to cede power to impacted communities, and elevated the perspectives of those communities within academia [55, 56]. Design scholarship has also begun to probe the limits of participatory design (PD) for centering equity [57], pointing to important gaps when researchers wield power over participating groups and ultimately control whose voices matter. Trying to engage Chicago residents in a conversation about smart city development, for example, Christina Harrington, Sheena Erete, and Anne Marie Piper [57] have discussed how PD workshops are often structured to further marginalize underserved and overburdened groups, such as through the use of particular ideational materials (colored pencils, markers) that can feel infantilizing. Across this body of work, scholars tend to apply PD in settings where researchers hold more power than participating groups.

Separately, a long tradition of ethnographic research has examined a wide set of power relationships, including what some call "studying up" — or the process by which researchers closely examine powerful actors, institutions, and structures as their locus of inquiry [86, 91, 112]. From high-tech technology firms to hate groups[1], scholars have outlined the varied threats to researcher safety and well-being, as well as difficulties related to the possibility of elevating the very voices researchers aim to challenge or hold to account [5, 86, 91, 112]. Though some design scholars have found methods of participant observation productive, techniques of PD remain an important but under-studied tool in settings where participants hold significant power, especially in revealing whose perspectives matter and how.

What would it mean to apply PD techniques to interactive system design within a wider set of relationships to power? Drawing from these distinct conversations, this paper turns to the multifaceted complexities of misinformation, disinformation, and online hate (M/D/OH) as a problem space to examine the tradeoffs of PD. This area of research is especially notable given the important societal ramifications of misinformation and disinformation for public health [87], democratic processes [28], and climate change [117] (among many other areas), and the widespread prevalence of online hate [7, 66]. Design scholarship has an important role to play in addressing these issues, as the structure of online platforms and systems significantly impacts the prevalence of these phenomena, and there is growing awareness of the value of community-engaged design methods in these contexts. In a recent (2022) roadmap for "information integrity research" [64], the U.S. National Science and

---

[1]"Studying up" as applied to hate groups could be contested. Scholars have traced how such groups claim to have been marginalized for some time, even as they ascend in their cultural currency and connections to infrastructures of power (e.g. [100]).

Technology Council recommended using participatory methods to both understand how communities are impacted by "corrupted information" and to design interventions to help communities protect themselves against these harms. However, M/D/OH researchers are just beginning to explore the use of participatory research methodologies in these contexts. As PD methods grow in popularity and reach [43, 99, 111], as scholars increasingly attach an implicit positivity to community-engaged work [111], and as participatory design is applied to an increasingly broad set of contexts [16], we hold a greater responsibility to interrogate and question our assumptions as design scholars.

To understand the potential impacts of PD in these settings, we turn to design fiction, a method of envisioning what particular contexts might be like in the future through speculative scenarios and stories. Here, we employ and adapt imaginary method abstracts to anticipate methodological and ethical challenges in pursuing this work. Imaginary method abstracts (based on work by Mark Blythe and Elizabeth Buie [21, 22]) is an adapted approach to assess the consequences of potential design research by writing in the style of academic abstracts about fictional prototypes and their future study. We combine this technique with insights from a stakeholder analysis to sketch possible futures for PD in M/D/OH research and foreground value tensions in using these methods in the M/D/OH context [47]. This paper explores some of the benefits and risks of applying PD methodologies to this context, and lays out potential paths for combining these areas of research.

This work makes three main contributions to scholarship on interactive systems design. First, it expands conversations on power by examining what participatory techniques might open and foreclose in contested settings such as M/D/OH. In particular, we outline three salient value tensions for sustained responsibility in PD engagement—authenticity, reciprocity, and impact—and urge interactive system design scholars to consider the varied axes along which participatory approaches might be adapted as they move and scale. In this way, we see this paper as adding to conversations around conducting effective, responsible PD research, similar to the guidelines for participatory equity-centered design outlined by non-profits such as The Creative Reaction Lab [33].

Second, we illustrate how adapted imaginary abstracts that focus on fictional methodological developments (rather than fictional prototypes) help scholars consider the tradeoffs of design research approaches. The expanded technique helps scholars of interactive systems design imagine possible outcomes of a future method application. This work complements existing examples of imaginary abstracts to consider a wider set of potential interventions premised on participation.

Lastly, our analysis helps bolster the imaginative dimensions of empirical inquiry and analysis (trace ethnography, social media analysis, etc.) connected with M/D/OH research. By examining some of the potential benefits and risks of using participatory approaches, and by outlining tensions between the goals of liberal democracy and justice-oriented scholarship, we help scholars imagine what participatory techniques might involve within M/D/OH contexts and how M/D/OH scholars might navigate their use, negotiation, and adaptation.

2 BACKGROUND

Our work builds on existing research across two areas of study: (1) the goals and outcomes of PD and (2) the role of PD in M/D/OH contexts. We review core ideas within both conversations in the sections that follow.

2.1 Participatory Design

Many disciplines have a history of using participatory approaches to research. Within the field of human-computer interaction, common early examples of PD arise from the tradition of Scandinavian design, through projects like the UTOPIA project [42] and the early work of the Norwegian Iron and Metal Workers' Union [52]. With additional attention to equity, PD research allows research participants to have a more authentic voice in the research process, beyond what they might have in other design workshop formats or through other qualitative methodologies, like interviews or surveys. As Harrington and colleagues (p.20 [57]) note, "[e]quitable PD considers and centers those who have been historically underserved, communities that have not been in positions of power." Alongside recognizing power differentials, this process can help cede some of the researchers' and/or designers' power to the participants [56]. This approach also aligns with the slogan popularized by the disability rights movement, "nothing about us without us," which Sasha Costanza-Chock has connected to PD and other methods of community-involved research and design [31]. When conducted with care, PD can embody community-centered, democratic values which allow those traditionally excluded to have meaningful input into designing systems.

However, saying that PD has the *potential* to involve communities and historically excluded populations does not mean that it *inherently* represents and incorporates these perspectives authentically. Even deciding what community means, how it gets defined, and by whom shapes the problems designers define and address. Recognizing the complexity of this participatory turn, our work builds on decades of critical anthropological scholarship [4, 59, 63, 110] and decolonial thought [121, 122, 129] focusing on the particularities of research participation within uneven power relationships. Lilly Irani, for example, critiques PD research as seeking to surface evidence of predetermined conclusions of funders or researchers, rather than actually representing the authentic perspectives of participants [37]. Similarly, Noura Howell and colleagues reflect on their own work and how it has failed in the past by primarily representing the creations of designers, even in the context of a co-designing, participatory framework [61], and Suzanne Bødker and Morten Kyng reflect on how participatory design scholars often choose to engage participants in ways that allow them to avoid political stances [23]. More broadly, PD brings up as a question whether design is the most effective or important way to work with communities, especially those already harmed by societal systems. Even Anthony Dunne and Fiona Raby, critiqued for their universalizing gaze on what futures are preferable [89, 115], recognize its participatory limits. They write: "the power of design is often overestimated. Sometimes we can have more effect as citizens than as designers. Protests and boycotts can still be the most effective ways of making a point" (p. 37 of [38]). This range of scholarship observes how PD processes often give researchers more power to

frame the terms of study and greater latitude to focus on their own perspectives in communicating the results of the work. As Randi Markussen (p. 63 of [78]) reminds, these researchers are "neither innocent nor all powerful," but hold their own invested interests.

Despite these shortcomings, recall from Harrington and colleagues above, PD has the potential to effectively challenge harmful managerial structures and forge new relationships to decision-making. This is especially true if, as Sarah Fox et. al. argue, PD researchers are reflexive throughout the research process and actively evaluate the power dynamics present within a research scenario [44]. It is this tension between the benefits of community-driven processes and the risks of over-representation that we seek to probe within settings of M/D/OH research.

2.2 Existing PD work in M/D/OH-related contexts

Misinformation, disinformation, and online hate are three important subcategories of harmful content online. We adapt our definitions for misinformation and disinformation from the work of Caroline Jack, defining misinformation as false — but not necessarily intentionally false — information, and disinformation as false or misleading information deliberately spread with the intention to deceive [65]. Though definitions of hate speech (and therefore online hate) are contested [6], we define online hate here as online content which targets an individual or group with hostile and harmful content (including racist, misogynistic, anti-Semitic, Islamophobic, transphobic, or homophobic content, among others), often based on identities the targeted person holds. We group these three areas — misinformation, disinformation, and online hate (M/D/OH) — together as they are often deeply entangled, co-occurring phenomena, and thinking of them collectively allows us to consider more of the nuances that different aspects of harmful content research bring to PD studies in these areas. Examples of this entanglement include anti-Semitic conspiracy theories based on centuries-old hate-based disinformation [46], and disinformation campaigns that opportunistically amplify false rumors (what would otherwise be misinformation) [108].

Recently, scholars have criticized the fields of misinformation and disinformation research for having too narrow of a focus on certain kinds of mis- and disinformation. These include J. Khadijah Abdurahman and Andre Brock, Jr.'s critiques that M/D/OH research centers predominantly white communities [1], Sarah Nguyễn et. al.'s writings that M/D/OH research has been predominantly U.S. and Anglo-centric [85], and Alice Marwick et. al.'s critical disinformation syllabus which shows how the field has tended to neglect examples beyond social media [80]. Though there has been some work on M/D/OH in other national contexts, including Brazil [19, 73, 101], Indonesia and the Philippines [90], India [3, 13], and Nigeria [8], and others have recently outlined how disinformation research can move beyond these more narrow conceptualizations [40, 74, 93, 120], as a whole the fields of M/D/OH research have tended to focus on white, English-speaking, Global North-based contexts. It is worth noting that the reason for this focus is not because other communities do not face challenges, which is clear from responses by minoritized communities to these problems (for examples, see [53, 62]). This overrepresentation is likely at least

partially attributable to the fact that these research communities, and academia as a whole, are white-dominated spaces where Black and Brown voices disproportionately face elision and silencing [88]. The reproduction of this inequitable recognition within M/D/OH scholarship could be for two related, but distinct, reasons. First, as J. Khadijah Abdurahman and Andre Brock, Jr. note [1], white misinformation and disinformation researchers could be focused on the issues that they see as most prevalent or impactful to their communities, such as the QAnon conspiracy movement, and focus their research on those topics with which they have the most experience. Slightly differently, white researchers may be disinclined to study communities of color as outsiders, perhaps because they do not have the cultural or linguistic knowledge necessary to conduct research in these communities, or due to concern that studying them through this 'outside' framing could serve to problematize or stigmatize already marginalized groups.

Partially as a response to these critiques of the narrow scope of the field, and alongside calls from US federal agencies like the White House Office of Science and Technology Policy to more broadly involve community-engaged methods in M/D/OH-related research [64], researchers have begun to use community-engaged, participatory, and interactive system design-based methods in M/D/OH contexts. Jason Young and colleagues wrote about the need for work with librarians to combat misinformation resulting from a series of community-engaged workshops, calling for the design of effective anti-misinformation resources for librarians and for support in addressing the social and political contexts limiting librarians' abilities to address these issues [128]. Zahra Ashktorab and Jessica Vitak used PD methods to research teen cyberbullying, which falls under a broad definition of online hate, and shows how PD can be used to help victims to address online harms by describing several system prototypes to mitigate harm [12]. Ahmer Arif proposes using design-based research methods to research both how people react emotionally to the information they are consuming and to develop strategies to be more thoughtful about the information they are consuming [9]. Tamar Wilner et. al. organized a series of workshops to co-design methods for increasing digital literacy in BIPOC and rural communities [124]. Sukrit Venkatagiri and colleagues designed an interactive system to support crowdsourced misinformation investigations [123]. Participatory game design (PGD) is also a method which has already gained traction in these research contexts, such as Carlos Roberto Torres-Parra et. al.'s *Poder Violeta* project, a participatorily designed video game used to combat sexual harassment (though focused more on offline rather than online hate) [116], the work of Ioana Literat et. al. on using PGD to create games to combat the spread of fake news [76], and the "Loki's Loop" project led by Chris Coward, Jin Ha Lee, and Lindsay Morse to build escape rooms to raise awareness of misinformation [32]. Within a related area of online conflicts, Kavous Salehzadeh Niksirat et. al. used participatory design methods to create an interactive system for resolving privacy conflicts [105]. Collectively, these studies have demonstrated the viability of using these research methodologies for topics broadly in M/D/OH contexts. We look across these studies to consider the breadth of M/D/OH research and its impacts to probe how PD could be used in M/D/OH research, and the value tensions this approach would surface. Considering a wide array of

these potential paths forward for PD research would help provide a roadmap for careful scholarship in these areas.

3 METHODS

Our analysis builds on a deep and long-term engagement in research on misinformation, disinformation, and online hate by Joseph S. Schafer (three years) and Kate Starbird (ten years). During the course of this research, both authors have relied primarily on mixed-methods social media data analysis, qualitative interviews and text analysis, and considered the limits of these methods for understanding the contextual realities of navigating M/D/OH spaces, as well as for envisioning alternative system designs to address these issues. Below we describe our attempt to better understand the potential of PD developments in M/D/OH contexts, a path that took us from stakeholder analysis to imaginary abstracts and value tensions.

3.1 Stakeholder Analysis

First, the authors performed a modified version of the stakeholder analysis process described in the work of Batya Friedman and colleagues [47]. Drawing from their extensive experience in these contexts, the authors developed a table of various kinds of communities which are impacted by M/D/OH phenomena, and have been studied by researchers working in these fields. The authors also developed a list of the possible designable artifacts that are relevant to M/D/OH, again informed by our observations from our experience working in these contexts. As the specific context of a particular artifact and community are highly relevant to whether a community is a direct or indirect stakeholder as defined by [47], we do not differentiate between these kinds of stakeholders in our tables. We also do not intend these stakeholder lists to be systematic, only to represent a list of several communities and artifacts potentially relevant for PD work in M/D/OH contexts.

3.2 Imaginary Method Abstracts

Design fictions broadly have a rich tradition of surfacing ethical issues and other considerations with technologies and designed artifacts [26, 69]. As a form of design fiction, imaginary abstracts are a technique developed by Mark Blythe and Elizabeth Buie based on science fiction author Stanislav Lem's prefaces and introductions written as if in the future, which allows, in their words, "the author to play with ideas, plots and character without having to write an entire novel or short story" [21, 22]. For Blythe and Buie, the imaginary abstracts focus on prototypes — designs that might or might not exist and research settings have not happened [21, 22]. They describe the fictional registers as multiple — designs, fields of study, and findings. Across these fictional dimensions, the authors suggest that this writing technique prompts useful questions such as whether failure is a useful finding and whether a deployment is "worth doing?" Here, we adapt this method to focus on PD research studies, rather than prototypes—opening an array of questions around the stakes of methodological adaptation and expansion. Attending to the adversarial nature of the D/OH contexts, we specifically chose to add watermarks to each of these imaginary method abstracts, to prevent them from being recontextualized in deceptive ways — for example to falsely claim that these studies

have been recommended or already conducted by these authors or other researchers in the field.

3.3 Value Tensions Analysis

After developing our imaginary method abstracts, we thought about the benefits and risks of using PD methods for M/D/OH research, combining all of these together across abstracts through successive rounds of close-reading and conversation between the authors. These discussions included iteration and elaboration on the tables and abstracts, such as refining the vocabulary used in row 3 of the table to describe those affected by M/D/OH as manipulated targets and denigrated targets[2], rather than as target vectors and target victims, and adding Abstract 4 after noticing additional tensions that our prior abstracts did not sufficiently illustrate. We then regrouped these implications based on similarities in the underlying values that these benefits and risks implicated. As each of these groups included both benefits and risks, these are best understood as value tensions, and specifically tensions *within* values. Value tensions are also an important part of the process developed by Batya Friedman and colleagues [47], though their work focuses more on tensions *between* values. Focusing on tensions *within* values allows us to specifically think about when these values are useful in service of the broader goals of the research we pursue, or if they can actually frustrate or harm those goals, rather than implicitly assuming these values we seek to embody in our research are always positive.

Combining elements of value-sensitive design analysis with speculative approaches has proven fruitful for past research, such as in the work of Stephanie Ballard et. al. engaging designers with these methods to surface ethical concerns for hypothetical technologies [15]. Our related method of combining the stakeholder analysis component of VSD analysis with the use of imaginary abstracts to elicit value tensions has multiple advantages. First, the stakeholder analysis allows us to consider a wide array of the potential approaches PD research could take in studying M/D/OH topics, emphasizing breadth. Second, the creation of imaginary method abstracts allows us to focus on a small number of possibilities, and think through the particular implications of doing those kinds of work. By concretizing the hypothetical research directions and helping clarify our methodological critique, the approach also allows us to more quickly consider an array of the methodological value tensions arising from a PD research approach, as the imaginary method abstracts allow us to envision studies in sufficient detail but do not require us to conduct the full project.

While imaginary method abstracts can be a powerful design fiction tool, there are a few limits to the insights this approach can generate, and to the value tensions they can then illuminate. First, the level of detail within these imaginary method abstracts is necessarily low; knowing more about specific methodological intricacies would potentially allow for a more nuanced exploration of the value tensions raised by the study. Furthermore, both the abstracts and the illuminated value tensions are also limited by the positionality of the authors, and so other value tensions that are relevant to this research may not have been perceived by our team.

---

[2]Target is a common rhetorical framing within M/D/OH research conversations, though this kind of language evokes more militaristic imagery than is common in the care-oriented field of PD research. We reflect on these language tensions in our discussion section.

Table 1: A table outlining some potential participant communities for hypothetical PD studies in the context of online misinformation, disinformation, and hateful content. This is not intended to be comprehensive, only to outline some potential options. Also, these rows are not mutually exclusive, and have some overlap.

Participant community groups	Example of non-PD research/writings on these communities in M/D/OH contexts
1. Platform users, broadly	Understanding how misinformation can be mitigated by platform user news outlet quality assessments [94]
2. Content moderation workers	Research on the psychological state of those working in content moderation [109]
3. Targets of M/D/OH, including both 3a: manipulated targets as vectors for persuasion and propagation, and 3b: denigrated targets as enemies or scapegoats	3a. Research on activist communities infiltrated by agents of disinformation campaigns [10] 3b. Survey of prevalence of people who have experienced online hate and harassment in video game communities [7]
4. Online Influencers	Understanding the impact of influential accounts on spreading misinformation in the 2020 election [68]
5. Creators of harmful online campaigns	Using machine learning methods to identify hydroxychloroquine misinformation, inspired by a video from the America's Frontline Doctors group [77]
6. Former members of online vector communities	Interviewing current and former believers in chemtrail conspiracy theories to understand their sensemaking processes [126]
7. Journalists reporting on M/D/OH	Understanding rapid collaborations between journalists and academics both working on misinformation [58]
8. Fact-checking organizations	Unpacking varying views on fact-checking responsibilities between citizens, journalists, and fact-checkers in Bangladesh [54]
9. Community organizations educating against misinformation (such as librarians or teachers)	A literature review unpacking the collective response from librarians and LIS researchers to misinformation after the 2016 US elections [113]
10. Governmental actors spreading mis/disinformation	Researching whether Russian information operations directly impacted voting behaviors [39]

We do not claim that the resulting insights are representative or comprehensive, only that this analysis reveals potentially significant value tensions worth further consideration in this research context.

4 WHAT COULD PARTICIPATORY DESIGN RESEARCH LOOK LIKE FOR STUDYING MISINFORMATION, DISINFORMATION, AND ONLINE HATE?

First, informed by a partial stakeholder analysis and drawing on the authors' prior experiences in studying M/D/OH topics, we develop two tables below. The first table suggests a list of possible communities to engage in PD research on this topic, and the second suggests the kinds of artifacts which could be (re)designed as part of the research process. The communities we highlight are informed by both the authors' experiences working in these areas, as well as by existing literature focusing on these communities in non-PD frameworks, and we highlight relevant work on examples of these community types for each row of this table.

Though many of these community stakeholder groups are relatively straightforward to understand, it is worth explaining that row 3, targeted communities, can mean multiple kinds of recipients of M/D/OH content. First, there are communities which are

manipulated targets, or communities which are utilized by disinformation purveyors to spread disinformation or hateful content to a wider audience. There are also *denigrated target* communities, or communities which are being bombarded with online hate and disinformation vilifying or scapegoating them.

For our next table, we outline possible artifacts that could be the focus of a PD study on M/D/OH. The artifacts we highlight represent categories that often figure prominently in academic and popular discourse around these topics. Just as several communities are involved in various aspects of misinformation, disinformation, and online hate, so are many kinds of artifacts, meaning that PD research on M/D/OH topics could consider designing a variety of objects.

As we can see from the above tables, there are a variety of settings where PD could be applied to research into misinformation, disinformation, and online hate. We do not claim this list to be comprehensive; rather, we offer the tables as generative artifacts laying out several possible avenues for future research. To ground our discussion of PD research in these contexts, we next present our imaginary method abstracts [21, 22] which could potentially be conducted to study aspects of misinformation, disinformation, and online hate, loosely inspired by the above tables. Our abstracts are similarly grounded in prior contexts of literature and real-world events, as they draw from both combinations of the rows in our two

Table 2: A table outlining some potential artifacts for hypothetical PD studies in the context of online misinformation, disinformation, and hateful content, alongside examples of what current examples of these artifacts might look like. This is not intended to be comprehensive. These rows are also not mutually exclusive, and artifacts can represent a combination of different rows.

Design artifacts	Examples of existing artifacts in these categories
A. Platform moderation infrastructure	These are mostly internal to platforms, one prominent example in the news was Facebook's XCheck program [60]
B. Official platform policies	Facebook's community standards [82]; Twitter's former Covid misinformation policy [71]
C. Informational campaigns/PSAs	CDC messaging campaigns around the coronavirus vaccines [84]
D. Platform affordances	The quote-tweet feature as a platform-supported mode of interaction, which exists on Twitter but not on Mastodon; the ability to comment on YouTube videos
E. Community norms/guidelines	The code of conduct enforced on the 2022 CSCW Discord server, available at [2]
F. Support systems for harmful content targets	Resource guides for targets of online harassment [48]; anti-harassment Twitter plug-ins [20]
G. Curricula and educational materials for combating misinformation	The SIFT curriculum for information verification [27]; the "Cranky Uncle" game for combating climate misinformation [30]; A High School "MisinfoNight" curriculum [119]
H. Legal policies around M/D/OH	France's former anti-misinformation and hate speech law [25]; German laws requiring removal of pro-Nazi and Holocaust-denying social media content [41]

Fictional Abstract 1: Support Through Trauma: Co-designing Workplace Support Systems for Content Moderation Workers

Prior research has shown that content moderation workers often face extremely poor working conditions, including low pay and insufficient mental health protections given the hateful, traumatizing content they must view for their job. Through a series of participatory design workshops, we worked with contractors in the Philippines for a major social media company to design novel workplace support systems for mitigating the mental traumas these workers face. We present three prototypes of internal support systems, along with methodological notes on working with vulnerable employee groups across national and linguistic barriers.

Figure 1: Our first imaginary method abstract, overlaid with a watermark to prevent recontextualization.

tables and from important M/D/OH research which used methodologies other than participatory design.

The first abstract helps us ask: how might PD be used to directly engage with and support groups integral to countering M/D/OH, but that have been primarily studied in other contexts?

Abstract 1 is informed by the work of Sarah Roberts [103], and is a combination of Row 2 (content moderation workers) from Table 1, and Row A (platform moderation infrastructure) from Table 2. Sarah Roberts' work [103] showed the horrible working conditions and lack of support that content moderation workers are often given at major social media platforms. Designing alternative support systems could be a productive use of PD methods.

However, what would it look like to do PD work with a population that are the direct denigrated targets of disinformation and online hate, and to design less technologically-oriented artifacts?

Abstract 2 is built off of a recognition of the increasing threats that election workers face [130], and is a combination of Row 3b (denigrated targets) from Table 1, and Rows B (platform policies) and H (legal policies) from Table 2. Here, we wanted to imagine the vulnerabilities denigrated target groups face, and the risks of focusing a research spotlight on these communities. We also wanted to specifically think about what the implications of having policy as a participatory design artifact would be in the M/D/OH context.

Thus far our imaginary method abstracts have focused on those harmed or victimized by M/D/OH. The next abstract helps us ask:

Fictional Abstract 2: Protecting the Polls: Participatory Policy Design to Prevent Violent Online Threats and Harassment

Violent threats against election workers in the United States due to ill-informed beliefs about elections being stolen have disturbingly become more common in recent years. However, adequate policy frameworks and systems do not exist to currently address these harms. In this paper, we describe two series of participatory workshops we conducted to design and implement new policies which would protect and support election workers facing threats and harassment. The first series of workshops took place specifically with election workers who have been specifically targeted with death threats and online harassment for their role in serving our democracy, to co-design beneficial policies for addressing the harms they face. The second series of workshops included both the election workers, as well as local politicians and community leaders, to facilitate the implementation of these policies. We present our designed policy artifacts, as well as notes on the challenges of co-designing when participants have unequal capacities to affect change.

Figure 2: Our second imaginary method abstract, overlaid with a watermark to prevent recontextualization.

Fictional Abstract 3: Counter-Participatory Design: Designing to Disrupt neo-Nazi Digital Infrastructure Use

With the rising prevalence of white nationalism online, design scholars have studied the way people use existing digital infrastructures to spread hate-based beliefs. However, access to these contexts can be difficult or impossible for outsiders. To address this hurdle, we ran a study where one of the authors embedded themselves in an online neo-Nazi community for a period of 4 months, during which time they facilitated multiple participatory workshops to both understand their current uses of technical infrastructure and the features which currently hinder their further growth. We then present the findings from these workshops, to recommend counter-designs which can reduce the current pathways these groups use to spread hate. We also provide methodological reflections on "participatory" co-design activities which are later repurposed to counter a community's goals, and the attending ethical considerations.

Figure 3: Our third imaginary method abstract, overlaid with a watermark to prevent recontextualization.

What happens if we instead consider applying PD methods to studying groups that are causing the spread of M/D/OH?

Abstract 3 is informed by prior work by Joan Donovan, Becca Lewis, and Brian Friedberg on digital infrastructures used by white nationalist and far-right groups, such as [36], and is a combination of Rows 3a (manipulated target communities) and 5 (creators of harmful online campaigns) from Table 1, and Row A (platform moderation infrastructure) and Row D (platform affordances) from Table 2. We use this abstract to think through the challenges of doing PD work with participants that have goals in opposition to the researchers, as well as the methodological and ethical issues with doing PD work involving deception in the research process — issues that grow especially complicated in the M/D/OH context.

Finally, we use our last abstract to examine PD within a community that is both a target of M/D/OH as well as highly marginalized in other contexts. In particular, we ask: How can researchers use PD to understand these situations in a more nuanced way?

Abstract 4 builds off recent work by Sarah Nguyễn et. al. studying disinformation in Asian diasporic communities [85], and is a combination of Row 3a (manipulated targets) from Table 1, and Row C (informational campaigns) from Table 2. Here, we wanted to think through studying a possible vector target community, and what doing PD work in that context would mean. We also wanted to understand through this fictional study how to conduct PD work in M/D/OH without problematizing the community with which we are collaborating.

Fictional Abstract 4: Building Resilience to Election Disinformation in Asian Diasporic Communities

In recent years, election disinformation has become a significant issue in the United States, undermining faith in the democratic process. While this issue has also had significant impacts in Asian diaspora communities, research into these communities and contexts has been comparatively limited. In a series of five participatory design workshops, we worked with members of Vietnamese-American diaspora communities to design informational resources and educational materials to address the lived experiences and concerns of Vietnamese-Americans navigating election information environments. We present the designed objects from our workshops along with reflections from community members on the experiences of navigating misinformation- and disinformation-heavy online environments, and strategies to inform contextual methods for sensitively addressing misinformation and disinformation in these communities and contexts.

Figure 4: Our fourth imaginary method abstract, with a watermark to prevent recontextualization.

5 DESIGN FICTION VALUE TENSIONS

Next, we turn to our thematic groupings of the benefits and risks we perceived from the hypothetical studies into value tensions which PD research foregrounds in M/D/OH research contexts. In particular, we describe how value tensions pertaining to *authenticity, reciprocity*, and *impact* are especially salient when evaluating the implications of using this research method. Below, we elaborate on each of these value tensions revealed by imagining PD research in these contexts.

5.1 Authenticity

The first value tension that became salient in thinking through these fictional studies relates to authenticity. By this, we mean genuine representation of the scope, contexts, and experiences of a topic. In the M/D/OH context, this means authentically representing the concerns, needs, and issues facing communities involved in or impacted by M/D/OH, rather than solely relying on researcher- or designer-imposed frames. This increased authentic visibility could have significant upsides, such as increased attention to community concerns or access to resources, but authentic visibility might also increase the potential for harm.

One way that authenticity can be promoted by PD in M/D/OH contexts comes from the potential to significantly improve the representation of groups currently excluded from mainstream M/D/OH research conversations. PD methods could allow researchers to cede their power and perspective to participants, to more authentically represent the diversity of perspectives on M/D/OH topics. For example, even if the researchers have not themselves been contractors working in content moderation, using PD for a study of content moderation contractors like that described in Abstract 1 could be useful to gain an authentic understanding of the issues these communities face. Alongside efforts to make structures within the academy more equitable, the study of these communities and their struggles with online misinformation and disinformation is important. Using PD to study these aspects of M/D/OH may help

to address omissions in our understanding of these phenomena, both regarding our knowledge of the issues and our abilities to confront M/D/OH. PD frameworks could be a way to conduct research so as to more authentically understand the perspectives of these communities.

Though engaging authentically with the communities and contexts is necessary for doing PD research in M/D/OH, this process can be traumatic for researchers. Even in less directly-engaged methods, researchers working in M/D/OH contexts come across graphic, upsetting or disorienting content, much like the communities that the researchers are studying. For example, Ahmer Arif et. al. write that when studying disinformation about Russian information operatives on Twitter engaging in discussions about the Black Lives Matter movement, "we found ourselves experiencing doubt when linking some of these accounts with pejorative terms like 'trolling' and 'propaganda'. This was especially true when we immersed ourselves with RU-IRA data in the ways that most closely resemble how an ordinary social media user would encounter their content," indicating that this research can be emotionally challenging and disorienting, especially when engaging in a way closer or more similar to that of participants [10]. Additionally, writing about a more directly engaged — though not based in PD — research study, Francesca Tripodi writes that "Once I began watching, reading, and listening to the news that my conservative respondents cited as trustworthy, it became increasingly difficult to discern the truth. . . I started applying [conspiratorial frames popular in these communities] to my own understanding of current events without even realizing it" (p. xv of [118]). Trauma for researchers from engaging with stressful research content is something that has been grappled with in fields outside of M/D/OH as well, such as in domestic violence research [45]. However, the more hands-on nature of PD methods means that the level of engagement with potentially traumatizing and/or disorienting content and topics could be exacerbated. The level to which these traumas are experienced are also dependent on the researcher's positionality. For example,

though doing research that engages directly with neo-Nazi or white nationalist communities, such as described in Abstract 3, would be traumatic for many researchers, this might be particularly harmful for researchers of color or others targeted by hate.

Furthermore, as several of our abstracts make clear, authentically representing other perspectives through conducting PD research within M/D/OH contexts might cause harm to those communities. Harms related to visibility, both from academic research (e.g. [83, 97, 104]) and in online settings (e.g. [35, 96]) have been seriously explored in previous research. The imaginary abstracts we produced also show ways that these visibility risks could manifest for PD research in M/D/OH contexts. By spotlighting a group targeted by misinformation, disinformation, or online hate campaigns, like the research on election officials receiving death threats for alleged malfeasance described in Abstract 2, this increased attention could put a larger target on these groups, and subject them to further harassment or attacks. Further, increased visibility in the context of M/D/OH research may be perceived as problematizing the studied community — i.e. increasing perceptions that the community is vulnerable to spreading misinformation and/or disinformation — and could lead to stigmatization. This is especially true for studies with *manipulated target* communities. For example, increasing attention on the Asian-American diasporic community, as the imaginary study in Abstract 4 would do, could lead to a misperception that they are more vulnerable than other communities, causing harm to the community and possibly resulting in additional distrust of researchers and scientific research more broadly (exacerbating the very problems that the study seeks to help mitigate). Similarly, the increased visibility brought to the content moderation contractors through the study in Abstract 1 could cause their employer to fire these workers for being critical of their work conditions.

5.2 Reciprocity

Another value tension which we found salient for considering PD research in M/D/OH contexts is that of reciprocity. For this context, we define reciprocity (in line with prior work [18]) as ensuring participants and researchers have mutual respect, understanding, and concern for each other, rather than emphasizing independence or the ossification of roles and stratifications. As Cynthia L. Bennett and Daniela K. Rosner describe, "[o]rienting designers to expect reciprocation might shift empathy building toward opportunities for mutual sense making, multiple first person narratives, and shared accountability" (p. 9 of [18]). Reciprocity can help to promote more empathetic respect and understanding, but with some groups involved in M/D/OH, especially with creators of harmful online campaigns, this empathetic respect or understanding may not be possible or desirable.

PD research can embody a value of reciprocity with some communities in that the method can be used to develop and demonstrate respect and care. While we describe the previous risk of authenticity related to problematizing groups by increasing their visibility in M/D/OH contexts, embodying reciprocity and mutual understanding could mitigate this threat. For example, in the case of Abstract 4, PD methodologies could help legitimize the real concerns of Asian-American diasporic communities where misinformation and

disinformation are often prevalent, and push back against conflations of mis/disinformation and the community in which it spreads.

However, reciprocity might not always be possible to establish for PD studies in M/D/OH contexts. Reciprocity requires mutual respect, both from the researcher toward the participant community *as well as* from the participants toward the researcher [18]. Doing research in these contexts is already risky for researchers, who face harassment, threats, and legal attacks on their work. PD is a method which far more directly engages with participants than many common M/D/OH approaches like large-scale social media analysis [68], digital ethnography [125], or qualitative analysis of online content [17], which means that these sorts of risks are likely to be amplified if engaging directly with those who are spreading M/D/OH. While less of a problem in establishing reciprocity with denigrated target communities, such as Abstract 2, the risk dramatically increases when engaging with participants who are actively spreading harmful content, such as Abstract 3. Engagement with hostile or targeting communities even in traditional research methodologies is already something contested in research ethics circles (e.g. [49, 67, 81]), and that level of interaction with these communities is far less than is required for PD. The positionality of researchers is also important here, as researchers holding different racial, sexual, disability, or gender identities would be differentially vulnerable to these attacks, for at least two reasons. First, many groups that believe in conspiracy theories or spread online hate do so toward specific identities and directed toward specific *denigrated target* communities, such as anti-Semitic or anti-feminist conspiracy theories, so researchers belonging to these identities are likely to face even more vicious attacks or threats from the groups they are studying. Secondly, the identities that researchers hold may make them more vulnerable to attack, less likely to be believed when reporting that they are being attacked, and less likely to be taken seriously as rigorous scholars. Though the risks of direct contact exist for many researchers in studies of hostile communities like the neo-Nazi communities in Abstract 3, these are heightened when the researcher holds identities directly targeted by the group they are studying.

5.3 Impact

The third and final primary value tension we found salient in our fictional abstracts concerns the notion of impact, or doing research that has real-world ramifications. PD is a methodology which prioritizes having real impacts for its participants, and for attempting to create real change. Similarly, recent calls for misinformation research to treat itself as a "crisis discipline" akin to climate research have circulated, underscoring that impact is also important for research in M/D/OH contexts [14]. Conducting impactful research can help to address the problems of M/D/OH, but may also have unjust results (anticipated or unknown).

One way that PD practices for M/D/OH research can be beneficially impactful for community members is by using the research method to empower communities and help them become more resilient against misinformation and disinformation. Recent work has argued that many people turn to misinformation because they are feeling disempowered [127] and that community empowerment

can reduce belief in false conspiracy theories [98]. PD, by facilitating direct engagement with participants, could not only support empowerment but also support education and outreach efforts to counter M/D/OH. Researchers have already begun to explore how community-engaged work with libraries, for example, can help to counter misinformation [128], and PD could continue to advance this work. This form of community engagement could help community members grapple with the harms caused by participating in spreading hate and provide fact-checking/countering content to promote community resilience.

Impactful PD research for studying M/D/OH could, on the other hand, present broader risks to the research community by platforming false or hateful ideas and allowing them to gain impact. This is already an issue acknowledged in the M/D/OH research context, under the framing of giving content the "oxygen of amplification" [95]. If we were to conduct PD research with communities that are actively causing harm through spreading M/D/OH, such as of the neo-Nazis studied in Abstract 3, the ceding of power to communities which occurs in PD may serve to amplify these views, giving them a greater reach than they might otherwise have. Even beyond direct amplification of these views, applying PD in this context could serve to further center these particular harmful communities, rather than the perspectives of those surviving the harms, perspectives that disinformation studies have largely neglected. This is not to say that the centering problems in M/D/OH research, as previously critiqued in [1], would dramatically worsen with PD research, but rather that using this research methodology does not automatically address these centering problems. Instead, PD could elevate those problems if exclusively applied to the same subset of M/D/OH research topics. These concerns about amplifying and platforming harmful ideologies expand beyond the realm of academia for researchers working in this space. Currently, given the relevance of online misinformation, disinformation, and hateful content to major societal issues like the January 6 insurrection [106], the coronavirus pandemic [24], and climate change [70], research in these spaces often migrates quickly into public conversations with policymakers, platform officials, and journalists. As a result, these problems of platforming and ceding power to groups actively causing harm could ripple out beyond the realm of academic discussions of M/D/OH.

The use of PD for impactful M/D/OH research also has the risk of generating backlash effects which could hinder future work in these contexts and researchers' ability to address these harms. In all four of our abstracts, it is possible to imagine how they could be weaponized against the field. In Abstract 1, the act of participatory research with content moderation workers could be criticized as researcher interference into platforms, along the same lines as the public backlash to the misleading criticisms of the "Twitter Files" [75]. Similarly, researchers using their positions to advocate for policies to protect harassment victims of election disinformation as described in Abstract 2 could also be subject to claims that researcher support of government (even local government such as election officials) is synonymous with harmful "collusion", a line of criticism which has previously been applied to research in M/D/OH [29]. Abstract 3's use of a 'counter-participatory' design framework, which in essence relies on deceiving the white nationalist participants, is also very easily criticized, just as the "New Knowledge"

research project was criticized for using deception during an active political race, problematically employing disinformation in an attempt to understand it [107]. In a study like Abstract 4, researchers studying M/D/OH topics in manipulated target communities, especially those which have historically been ill-treated by researchers, may find their work being reframed and used as evidence that researchers are not to be trusted by the very groups targeting these communities in the first place. Depending on the specifics of each of the studies described in our abstracts and the conduct of the research teams in these imaginary studies, these criticisms may be made in good or bad faith — and have varying levels of credibility. Given the already highly contested, politicized nature of M/D/OH research, it's easy to imagine that work using PD methods could increase backlash to both individual research teams and to the field as a whole, which researchers should consider when organizing and presenting on this research.

Additionally, researchers working for public impact can also face increased mental strain and pressure, such as feelings that they are failing if the real-world impacts they want their work to have fail to materialize. This is something that researchers like Natascha Klocker have grappled with in related participatory methodologies, such as participatory action research, concluding that researchers should think carefully about what kind of participation is appropriate for the impact their work will have, and that researchers should emotionally prepare for the challenges of attempting impactful research [72]. The implications of these research topics for major world events can mean that researchers feel pressure to have their research impact these issues. Further, in the PD context, researchers may feel even more pressure (compared with other methodologies) to ensure their participants are positively impacted by the research they are conducting. For example, depending on the depth of connection formed between researchers and partners, researchers may feel increased responsibility to improve the employment conditions for the content moderation contractors in Abstract 1 or to protect election worker participants from future death threats in Abstract 2. This possibility is particularly true for researchers who are themselves targets of online hate or are members of communities targeted by mis- and disinformation, as they are members of the communities they are hoping to impact. With amplified stakes, they may feel obliged to perform impactful research for their participants and themselves.

6 DISCUSSION

We have so far seen how speculative approaches such as imaginative abstracts help scholars outline some of the stakes and possibilities of established participatory methods. Participation can take many forms and invite different levels of participation by research partners, depending on the setting. When thinking about whether and how to apply PD methodologies to M/D/OH research topics, our analysis suggests that scholars consider the positionality of both the researcher and the research partners when evaluating the power dynamics at play. We find that PD research in these contexts that undertakes a reflexive approach, as previously argued in [44], might productively expand this reflexivity to incorporate reflection on the kinds of participant voices being foregrounded. This accounting for varied positioning around participation recalls

a form of Arnstein's Ladder [11], a framework originally used to describe levels of engagement with communities in urban planning (from non-participation to citizen control). Our analysis does not offer specific prescriptions for how to navigate these tensions when combining participatory design methods with the M/D/OH context; instead, we argue that what is most critical is to reflect on these tensions before and throughout engaging with participants in these contexts. Though an inherent feature and strength of participatory methods is that they can lead to outcomes unanticipated by researchers, it is nevertheless prudent to think through how to scope a participatory project to minimize the potential harms of this work.

In closing, we "zoom out" from our analysis of specific value tensions to show how these tensions reflect three central methodological risks with combining the research areas of participatory design of interactive systems, design fiction, and M/D/OH research: (1) considerations for using PD for M/D/OH, (2) the fraught character of designing "with" (or against) hate, and (3) the challenges of using design fiction in M/D/OH contexts. Finally, we reflect on some directions for future work that could build off of this area.

6.1 Participatory Design in Misinformation, Disinformation, and Hate Contexts

From our analysis we see how PD may provide important advantages to the fields of M/D/OH research by expanding the kinds of questions and insights proposed and the variety of solutions generated. But these insights and solutions vary widely depending on the kind of community PD engages. When working with manipulated target communities, whether or not the researchers are members of these communities, it is important to reflect on the ways that our work might add to stigmatizations of these communities (as prior qualitative research ethics papers have argued, e.g. [51, 92]), and to provide these participants agency in the PD process. When working with denigrated targets, we believe it is worth reflecting on how to perform work that does not enlarge the targets on their backs, and make them more vulnerable to future harm through increasing their visibility (as previously described in e.g. [83, 97, 104]). We also believe it is important for PD researchers in M/D/OH contexts to surface and consider how engaging a community in participatory design for research might compromise their work either now or in the future — e.g. working with moderators in such a way that would allow critics (whether in good faith or bad) to conflate all moderation with censorship, in something akin to the "Twitter Files" [75] — and to weigh and/or mitigate these risks in the research design. Sometimes, the benefits of doing the research might not outweigh those risks.

When we consider adapting PD methods to study spreading communities of M/D/OH, this research methodology becomes particularly fraught. As Théophile Lenoir and Chris Anderson surface and problematize in their introductory essay to the "What Comes After Disinformation" volume [74], research on M/D/OH topics in recent years has often built, either implicitly or explicitly, atop assumptions that truth is normatively good and that manipulation is normatively bad. This grounding can make tensions surrounding deception and power in research methods especially salient. From this position, sustained deception in a research study — such as

Abstract 3's fictional study with white nationalists — would violate this normative ground. Utilizing deception could undermine trust in the research team's work, as well as the broader field, and potentially science as a whole. Prior work where researchers used sockpuppet accounts to attempt to measure the impact of online influence operations was subjected to extensive criticism from both good-faith and bad-faith actors, and undermined confidence in the field [107]. Along these same lines, Thomas Rid, a scholar of Soviet active measures, has argued that countries that ascribe to democratic values cannot use active measures such as disinformation without undermining the very values they are ostensibly defending [102]. Researchers seeking to uphold liberal democratic values against misinformation and disinformation similarly cannot utilize sustained deception in understanding these phenomena without undermining their commitments. However, these value commitments and prioritizations (for example, truth as a primary commitment over justice as a primary commitment) are not necessarily universal in M/D/OH research. In the wake of Russia's invasion of Ukraine, some M/D researchers and journalists on the misinformation beat (e.g. [114]) grappled with how to approach reporting on "the Ghost of Kyiv" — a likely false story of Ukrainian heroism, but one that was motivating Ukrainian resistance. A recent volume, "What Comes After Disinformation" [79] articulates a range of value tensions that complicate the "truth-first" grounding. All of this underscores how important it is that PD researchers in M/D/OH contexts who intend to use deception surface their value commitments and carefully think through how the use of deception, or even the appearance of deception, could be weaponized to delegitimize a specific study — and/or the broader field.

In writing this paper as scholars from distinct (PD and M/D/OH) scholarly communities, we additionally reflected on the kinds of language taken up to describe M/D/OH activities and partners. For example, M/D/OH scholars frequently rely on militaristic terms such as "perpetrator," "target," "information operation," "bombarded," and "infiltrate" to describe the varied actions and actors in these settings. Through our analysis, we began to notice these terms challenge the care politics of PD scholarship, prompting new reflections on what those concepts assume and how greater nuance and accountability can be brought to our interpretations through writing.

6.2 The Fraught Character of Designing "with" (or against) Hate

Our analysis has raised important questions around not only the risks of participation but also the risks of this very thought experiment. To what extent is our analysis of PD in complex and harmful situations of M/D/OH drawing more attention to the very patterns we hope to upend? What more do we expect from a PD study focused on hate, from turning design scholars' gaze toward harm? What is the cost and to whom?

These questions begin to trouble how we scope our speculative project. To probe the study of structural harm is to ask whose voices design researchers elevate and obscure, and to reveal inherent problems with treating hateful contexts as the focus/target of design. A range of feminist and anti-racist scholarship has convincingly argued for not centering the actions and perspectives of those already

wielding power, including the hate groups described in Abstract 3. Denise Ferreira da Silva [34], for example, astutely suggests that the unpacking of the philosophy behind racism can also work to reinforce racist beliefs. Within our analysis of PD, particularly Abstract 3, we see a similar potential. Explanations of the philosophy behind hate groups or other hostile settings may serve to further entrench hateful ideology. It may also make way for the kind of vitriolic analysis of pain, disillusionment, and terror transnational feminist and Black studies scholars have long worked to dismantle.

In partial response to these concerns, PD scholars may assume a position of 'participatory design against' to actively undermine harmful activity. Such an approach might involve researchers using findings from PD activities to counter or push back against harmful activity within a study setting. These possibilities trouble the underlying values and commitments of participatory work, identifying tensions between egalitarian and justice-oriented outcomes (discussed above). In light of previous work that has turned to PD and other design-based practices as ways of generating new insights in M/D/OH contexts (e.g. [12, 32, 76, 116, 124, 128]), our analysis invites scholars to look beyond the limited scope of M/D/OH scholarship to consider new and emergent PD techniques not as solutions but as methodological horizons to which we might hold ourselves accountable.

6.3 Design Fiction and M/D/OH Research

A final central concern that surfaced in our analysis concerns the imaginary method abstracts and the risk of unintended harm. We worried in particular about the risks of their recontextualization — a particularly acute concern in "adversarial" research spaces such as disinformation and online hate, where the individuals and organizations implicated in the study of these phenomena can also turn those tactics onto the researchers studying them. In terms of both the value tensions of authenticity and impact, thinking about how the findings of research in this area might be reconstrued or represented in other settings were significant concerns we surfaced when thinking about applying PD methods to M/D/OH contexts.

With further reflection, our fears of recontextualization extend not just to these hypothetical studies, but to this very paper. Near the end of the manuscript preparation phase, we decided to add the watermarks to our imaginary abstracts to prevent them from being recontextualized and used to support bad faith criticisms of the field as though the fictional studies had already been done. We do not endorse the actual doing of the research described in any of the imaginary abstracts we produce, and indeed we explicitly argue Abstract 3 is likely incompatible with normative assumptions common in M/D/OH research. However, we also felt that without watermarking, it is all too easy to imagine that a screenshot of one of the abstracts could be weaponized as 'evidence' to attack adjacent community-based research.

But our analysis has also shown how hypothetical studies such as imaginary method abstracts allow us to chart a variety of value tensions arising from potential PD studies, without running the ethical risks of conducting a study before fully evaluating its implications. Additionally, we see that the imaginary methods abstracts allow for enough specificity to consider some of the nuances of

cases while allowing for flexibility to think through multiple potential implementations. In this sense, design fiction and speculative design have the ability to fulfill important roles in the M/D/OH space. Whether enabling a less risky form of exploring the ethics of a research space, as is done in this paper, envisioning speculative solutions to the challenges of M/D/OH, or imagining evolutions of these fields 'after disinformation studies' [74], these methods can be quite illuminating. Much like the challenges we outline around backlashes, however, it is important to think about possible recontextualizations when applying design fictions to M/D/OH topics, and presenting the outputs of this work in such a way to mitigate those threats.

6.4 Future Work

While we are considering participatory methods in this paper, we did not conduct this study in a participatory manner. Future work might revisit and expand our analysis in partnership with groups already engaged in meaningful community-based practices around M/D/OH. For example, we might use participatory workshops with community partners to reveal the range of reflection process the fictional abstracts elicit and how the reflections map to the value tensions we identified. Such research might focus on further refining the application of participatory approaches in a similar manner to Verena Fuchsberger et. al.'s engagement of participants with fictional job postings to reflect on future work conditions [50], or Stephanie Ballard et. al.'s work with designers combining value-sensitive design and design fiction to surface ethical concerns with technology [15]. Regardless of the form, we urge scholars to thoughtfully consider how to make sure that fictional abstracts would not be misinterpreted as existing studies within these engagements. Additionally, as community-engaged research gains traction within M/D/OH research and in designing systems to address these phenomena, reviews of existing literature should be done to see how these tensions are considered and addressed by existing researchers and practitioners.

7 CONCLUSION

In this paper, we have used design fiction to create imaginary method abstracts to explore the viability of using PD methodologies for studying topics of misinformation, disinformation, and online hate. We find three value tensions—*authenticity*, *reciprocity*, and *impact*— particularly salient when evaluating whether and how to conduct PD research on these topics, and outline some of the opportunities and challenges associated with each tension. We conclude that PD designers and researchers working in M/D/OH contexts should reflect carefully on the kinds of participation that best suit their specific research questions and context, both involving the positionality of the researchers and participants, as well as the degree to which PD participants have the capacity to participate.

REFERENCES

[1] J. Khadijah Abdurahman. 2021. (dis)Info Studies: André Brock, Jr. on Why People Do What They Do on the Internet. *Logic Magazine*. Retrieved September 28, 2022 from https://logicmag.io/beacons/dis-info-studies-andre-brock-jr/

[2] ACM SIGCHI. 2022. Code of Conduct. *CSCW 2022*. Retrieved December 7, 2022 from https://cscw.acm.org/2022/code-of-conduct/

[3] Syeda Zainab Akbar, Anmol Panda, and Joyojeet Pal. 2022. Political hazard: misinformation in the 2019 Indian general election campaign. *South Asian History and Culture* 13, 3 (July 2022), 399–417. DOI:https://doi.org/10.1080/19472498.2022.2095596

[4] Sareeta Amrute. 2016. *Encoding Race, Encoding Class: Indian IT Workers in Berlin*. Duke University Press, Durham, NC.

[5] Sareeta Amrute. 2020. Immigrant Sensibilities in Tech Worlds: Sensing Hate, Capturing Dissensus. *Cultural Anthropology* 35, 3 (August 2020), 374–403. DOI:https://doi.org/10.14506/ca35.3.02

[6] Luvell Anderson and Michael Barnes. 2022. Hate Speech. In *The Stanford Encyclopedia of Philosophy* (Spring 2022), Edward N. Zalta (ed.). Metaphysics Research Lab, Stanford University. Retrieved February 8, 2023 from https://plato.stanford.edu/archives/spr2022/entries/hate-speech/

[7] Anti-Defamation League. 2022. Hate Is No Game: Hate and Harassment in Online Games 2022. Retrieved December 8, 2022 from https://www.adl.org/resources/report/hate-no-game-hate-and-harassment-online-games-2022

[8] Oberiri Destiny Apuke, Bahiyah Omar, Elif Asude Tunca, and Celestine Verlumun Gever. 2022. Does Misinformation Thrive With Social Networking Site (SNS) Dependency and Perceived Online Social Impact Among Social Media Users in Nigeria? Testing a Structural Equation Model. *Journal of Asian and African Studies* (June 2022), 00219096221108738. DOI:https://doi.org/10.1177/00219096221108738

[9] Ahmer Arif. 2018. Designing to Support Reflection on Values & Practices to Address Online Disinformation. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ACM, Jersey City NJ USA, 61–64. DOI:https://doi.org/10.1145/3272973.3272974

[10] Ahmer Arif, Leo Graiden Stewart, and Kate Starbird. 2018. Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (November 2018), 20:1-20:27. DOI:https://doi.org/10.1145/3274289

[11] Sherry R. Arnstein. 1969. A Ladder Of Citizen Participation. *Journal of the American Institute of Planners* 35, 4 (July 1969), 216–224. DOI:https://doi.org/10.1080/01944366908977225

[12] Zahra Ashktorab and Jessica Vitak. 2016. Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, San Jose California USA, 3895–3905. DOI:https://doi.org/10.1145/2858036.2858548

[13] Sumitra Badrinathan. 2021. Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. *American Political Science Review* 115, 4 (November 2021), 1325–1341. DOI:https://doi.org/10.1017/S0003055421000459

[14] Joseph B. Bak-Coleman, Mark Alfano, Wolfram Barfuss, Carl T. Bergstrom, Miguel A. Centeno, Iain D. Couzin, Jonathan F. Donges, Mirta Galesic, Andrew S. Gersick, Jennifer Jacquet, Albert B. Kao, Rachel E. Moran, Pawel Romanczuk, Daniel I. Rubenstein, Kaia J. Tombak, Jay J. Van Bavel, and Elke U. Weber. 2021. Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences* 118, 27 (July 2021), e2025764118. DOI:https://doi.org/10.1073/pnas.2025764118

[15] Stephanie Ballard, Karen M. Chappell, and Kristen Kennedy. 2019. Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (DIS '19), Association for Computing Machinery, New York, NY, USA, 421–433. DOI:https://doi.org/10.1145/3322276.3323697

[16] Liam Bannon, Jeffrey Bardzell, and Susanne Bødker. 2018. Reimagining participatory design. *interactions* 26, 1 (December 2018), 26–32. DOI:https://doi.org/10.1145/3292015

[17] Andrew Beers, Sarah Nguyễn, Emma S. Spiro, and Kate Starbird. 2021. Rejecting Science With Science: Boundary-Work In Anti-Mask Twitter Reply Threads During COVID-19. *AoIR Selected Papers of Internet Research* (September 2021). DOI:https://doi.org/10.5210/spir.v2021i0.12143

[18] Cynthia L. Bennett and Daniela K. Rosner. 2019. The Promise of Empathy: Design, Disability, and Knowing the "Other." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–13. DOI:https://doi.org/10.1145/3290605.3300528

[19] Priscila Biancovilli, Lilla Makszin, and Claudia Jurberg. 2021. Misinformation on social networks during the novel coronavirus pandemic: a quali-quantitative case study of Brazil. *BMC Public Health* 21, 1 (June 2021), 1200. DOI:https://doi.org/10.1186/s12889-021-11165-1

[20] Block Party. 2022. Consumer tools for online safety and anti-harassment | Block Party. Retrieved December 7, 2022 from https://www.blockpartyapp.com/

[21] Mark Blythe. 2014. Research through design fiction: narrative in real and imaginary abstracts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Toronto Ontario Canada, 703–712. DOI:https://doi.org/10.1145/2556288.2557098

[22] Mark Blythe and Elizabeth Buie. 2014. Chatbots of the gods: imaginary abstracts for techno-spirituality research. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, ACM, Helsinki Finland, 227–236. DOI:https://doi.org/10.1145/2639189.2641212

[23] Susanne Bødker and Morten Kyng. 2018. Participatory Design that Matters—Facing the Big Issues. *ACM Trans. Comput.-Hum. Interact.* 25, 1 (February 2018), 4:1-4:31. DOI:https://doi.org/10.1145/3152421

[24] Shannon Bond. 2021. Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows. *NPR*. Retrieved December 1, 2022 from https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes

[25] Aurelien Breeden. 2020. French Court Strikes Down Most of Online Hate Speech Law. *The New York Times*. Retrieved January 23, 2023 from https://www.nytimes.com/2020/06/18/world/europe/france-internet-hate-speech-regulation.html

[26] Ryan Calo, Batya Friedman, Tadayoshi Kohno, Hannah Almeter, and Nick Logler (Eds.). 2020. *Telling Stories: On Culturally Responsive Artificial Intelligence*. University of Washington Tech Policy Lab. Retrieved April 19, 2022 from https://techpolicylab.uw.edu/telling-stories/

[27] Michael Caulfield. 2019. SIFT (The Four Moves). *Hapgood*. Retrieved June 2, 2022 from https://hapgood.us/2019/06/19/sift-the-four-moves/

[28] Center for an Informed Public, Digital Forensic Research Lab, Graphika, and Stanford Internet Observatory. 2021. *The Long Fuse: Misinformation and the 2020 Election* (v1.3.0 ed.). Stanford Digital Repository: Election Integrity Partnership. Retrieved December 12, 2021 from https://purl.stanford.edu/tr171zs0069

[29] Jeff Cercone. 2022. Election Integrity Partnership targeted election misinformation, not conservatives. *Poynter*. Retrieved February 9, 2023 from https://www.poynter.org/fact-checking/2022/election-integrity-partnership-targeted-election-misinformation-not-conservatives/

[30] John Cook. 2021. Cranky Uncle: a game building resilience against climate misinformation. *Plus Lucis* 3, 2021 (2021), 13–16.

[31] Sasha Costanza-Chock. 2020. Design Practices: "Nothing about Us without Us." In *Design Justice* (1st ed.). Retrieved from https://designjustice.mitpress.mit.edu/pub/cfohnud7/release/4

[32] Chris Coward, Jin Ha Lee, and Lindsay Morse. 2022. Loki's Loop Escape Room. Retrieved January 19, 2023 from http://lokisloop.org

[33] CRL. Creative Reaction Lab. *Creative Reaction Lab*. Retrieved May 8, 2023 from https://crxlab.org

[34] Denise Ferreira Da Silva. 2022. *Unpayable Debt*. Sternberg Press, London.

[35] Michael Ann DeVito. 2022. How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility Via Folk Theorization. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (November 2022), 1–31. DOI:https://doi.org/10.1145/3555105

[36] Joan Donovan, Becca Lewis, and Brian Friedberg. 2018. Parallel Ports: Sociotechnical Change from the Alt-Right to Alt-Tech. In *Post-Digital Cultures of the Far Right*. transcript Verlag, 49–66. DOI:https://doi.org/10.1515/9783839446706-004

[37] Dylan Doyle-Burke and Jessie "Jess" Smith. 2020. *Labor and Innovation: Exploring the Power of Design and Storytelling with Lilly Irani*. Retrieved from https://podcasts.apple.com/us/podcast/labor-innovation-exploring-power-design-storytelling/id1505229145?i$=$1000474434129

[38] Anthony Dunne and Fiona Raby. 2013. Design as Critique. In *Speculative Everything*. MIT Press, 33–46. Retrieved from https://muse-jhu-edu.offcampus.lib.washington.edu/chapter/1057618

[39] Gregory Eady, Tom Paskhalis, Jan Zilinsky, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. 2023. Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nat Commun* 14, 1 (January 2023), 62. DOI:https://doi.org/10.1038/s41467-022-35576-9

[40] Kirsten A. Eddy. 2023. Centering Identity and Morality in Disinformation Studies. *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill* (January 2023). Retrieved February 9, 2023 from https:

//citap.pubpub.org/pub/6oi63cc7/release/1

[41] Melissa Eddy and Mark Scott. 2017. Delete Hate Speech or Pay Up, Germany Tells Social Media Companies. *The New York Times*. Retrieved January 23, 2023 from https://www.nytimes.com/2017/06/30/business/germany-facebook-google-twitter.html

[42] Pelle Ehn. 1988. Work-oriented design of computer artifacts. Doctoral Thesis. Umeå University, Umeå, Sweden. Retrieved November 28, 2022 from http://www.diva-portal.org/smash/get/diva2:580037/fulltext02.pdf

[43] Mette Agger Eriksen, Seppe De Blust, Oswald Devish, Christian Dindler, Carl DiSalvo, Anna Seravalli, and Majken Toftager Larsen. 2022. Expanding Learning in Participatory Design: Mapping the Field of Learning Theory and Practice in PD. In *Proceedings of the Participatory Design Conference 2022 - Volume 2*, ACM, Newcastle upon Tyne United Kingdom, 233–235. DOI:https://doi.org/10.1145/3537797.3537842

[44] Sarah Fox, Catherine Lim, Tad Hirsch, and Daniela K. Rosner. 2020. Accounting for Design Activism: On the Positionality and Politics of Designerly Intervention. *Design Issues* 36, 1 (January 2020), 5–18. DOI:https://doi.org/10.1162/desi_a_00571

[45] Sílvia Fraga. 2016. Methodological and ethical challenges in violence research. *Porto Biomedical Journal* 1, 2 (May 2016), 77–80. DOI:https://doi.org/10.1016/j.pbj.2016.04.005

[46] Brian Friedberg. 2020. The Dark Virality of a Hollywood Blood-Harvesting Conspiracy. *Wired*. Retrieved January 24, 2023 from https://www.wired.com/story/opinion-the-dark-virality-of-a-hollywood-blood-harvesting-conspiracy/

[47] Batya Friedman, Peter H. Kahn Jr., and Alan Borning. 2008. Value Sensitive Design and Information Systems. In *The Handbook of Information and Computer Ethics*, Kenneth E. Himma and Herman T. Tavani (eds.). John Wiley & Sons, Incorporated, Hoboken, UNITED STATES, 69–101. Retrieved January 21, 2023 from http://ebookcentral.proquest.com/lib/washington/detail.action?docID$=$353290

[48] Jaclyn Friedman, Anita Sarkeesian, and Renee Bracey Sherman. 2018. Speak Up & Stay Safe(r): – A Guide to Protecting Yourself From Online Harassment. Retrieved December 7, 2022 from https://onlinesafety.feministfrequency.com/en/

[49] Christian Fuchs and Graham Meikle. 2018. "Dear Mr. Neo-Nazi, Can You Please Give Me Your Informed Consent So That I Can Quote Your Fascist Tweet?": Questions of Social Media Research Ethics in Online Ideology Critique. In *Routledge Companion to Media and Activism*. Routledge. Retrieved from https://westminsterresearch.westminster.ac.uk/download/3771d54accbad32ff3aad192b97046846bf33526a21984caf3c25b140ef88c26/160940/Routledge%20Companion%20to%20Media%20and%20Activism.pdf

[50] Verena Fuchsberger, Thomas Meneweger, Daniela Wurhofer, and Manfred Tscheligi. 2017. Apply Now! Fictional Job Postings as an Instrument to Discuss Interactive Futures of Work. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS '17), Association for Computing Machinery, New York, NY, USA, 581–586. DOI:https://doi.org/10.1145/3064663.3064750

[51] Kemesha Gabbidon and Tiffany Chenneville. 2021. Strategies to minimize further stigmatization of communities experiencing stigma: A guide for qualitative researchers. *Stigma and Health* 6, 1 (February 2021), 32–42. DOI:https://doi.org/10.1037/sah0000280

[52] Judith Gregory. 2003. Scandinavian Approaches to Participatory Design. *International Journal of Engineering Education* 19, (January 2003).

[53] Rachelle Hampton. 2019. The Black Feminists Who Saw the Alt-Right Threat Coming. *Slate*. Retrieved February 9, 2023 from https://slate.com/technology/2019/04/black-feminists-alt-right-twitter-gamergate.html

[54] Md Mahfuzul Haque, Mohammad Yousuf, Ahmed Shatil Alam, Pratyasha Saha, Syed Ishtiaque Ahmed, and Naeemul Hassan. 2020. Combating Misinformation in Bangladesh: Roles and Responsibilities as Perceived by Journalists, Fact-checkers, and Users. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (October 2020), 1–32. DOI:https://doi.org/10.1145/3415201

[55] Jean Hardy, Caitlin Geier, Stefani Vargas, Riley Doll, and Amy Lyn Howard. 2022. LGBTQ Futures and Participatory Design: Investigating Visibility, Community, and the Future of Future Workshops. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (November 2022), 1–25. DOI:https://doi.org/10.1145/3555638

[56] Christina Harrington and Tawanna R Dillahunt. 2021. Eliciting Tech Futures Among Black Young Adults: A Case Study of Remote Speculative Co-Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–15. DOI:https://doi.org/10.1145/3411764.3445723

[57] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing Community-Based Collaborative Design: Towards More Equitable Participatory Design Engagements. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (November 2019), 1–25. DOI:https://doi.org/10.1145/3359318

[58] Melinda McClure Haughey, Martina Povolo, and Kate Starbird. 2022. Bridging Contextual and Methodological Gaps on the "Misinformation Beat": Insights from Journalist-Researcher Collaborations at Speed. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI '22), Association for Computing Machinery, New York, NY, USA, 1–15. DOI:https://doi.org/10.1145/3491102.3517503

[59] Julie Hemment. 2007. Public Anthropology and the Paradoxes of Participation: Participatory Action Research and Critical Ethnography in Provincial Russia. *Human Organization* 66, 3 (September 2007), 301–314. DOI:https://doi.org/10.17730/humo.66.3.p153144353wx7008

[60] Jeff Horwitz. 2021. Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt. *Wall Street Journal*. Retrieved December 7, 2022 from https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353

[61] Noura Howell, Audrey Desjardins, and Sarah Fox. 2021. Cracks in the Success Narrative: Rethinking Failure in Design Research through a Retrospective Tri-oethnography. *ACM Trans. Comput.-Hum. Interact.* 28, 6 (December 2021), 1–31. DOI:https://doi.org/10.1145/3462447

[62] Tiffany Hsu. 2022. Misinformation Swirls in Non-English Languages Ahead of Midterms. *The New York Times*. Retrieved February 9, 2023 from https://www.nytimes.com/2022/10/12/business/media/midterms-foreign-language-misinformation.html

[63] Janise Hurtig. 2008. Community Writing, Participatory Research, and an Anthropological Sensibility. *Anthropology & Education Quarterly* 39, 1 (2008), 92–106. DOI:https://doi.org/10.1111/j.1548-1492.2008.00007.x

[64] Information Integrity Research And Development Interagency Working Group and Networking and Information Technology Research and Development Subcommittee. 2022. *Roadmap For Researchers On Priorities Related To Information Integrity Research And Development*. National Science and Technology Council, United States of America. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2022/12/Roadmap-Information-Integrity-RD-2022.pdf

[65] Caroline Jack. 2017. Lexicon of Lies: Terms for Problematic Information. *Data & Society*. Retrieved from https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf

[66] Teo Keipi, Matti Näsi, Atte Oksanen, and Pekka Räsänen. 2016. The rise of online hate. In *Online Hate and Harmful Content*. Taylor & Francis Group, 22. Retrieved January 26, 2023 from https://www.taylorfrancis.com/chapters/oa-mono/10.4324/9781315628370-4/rise-online-hate-teo-keipi-matti-n%C3%A4si-atte-oksanen-pekka-r%C3%A4s%C3%A4nen

[67] Brit Kelley and Stephanie Weaver. 2020. Researching People who (Probably) Hate You: When Practicing "Good" Ethics Means Protecting Yourself. *Computers and Composition* 56, (June 2020), 102567. DOI:https://doi.org/10.1016/j.compcom.2020.102567

[68] Ian Kennedy, Morgan Wack, Andrew Beers, Joseph S. Schafer, Isabella Garcia-Camargo, Emma S. Spiro, and Kate Starbird. 2022. Repeat Spreaders and Election Delegitimization: A Comprehensive Dataset of Misinformation Tweets from the 2020 U.S. Election. *Journal of Quantitative Description: Digital Media* 2, (June 2022). DOI:https://doi.org/10.51685/jqd.2022.013

[69] Os Keyes, Jevan Hutson, and Meredith Durbin. 2019. A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–11. DOI:https://doi.org/10.1145/3290607.3310433

[70] David Klepper. 2022. Climate disinformation leaves lasting mark as world heats. *AP NEWS*. Retrieved December 1, 2022 from https://apnews.com/article/wildfires-science-fires-american-petroleum-institute-014d4825f21084a80eb71414dbe63b9e

[71] David Klepper. 2022. Twitter ends enforcement of COVID misinformation policy. *AP NEWS*. Retrieved February 6, 2023 from https://apnews.com/article/twitter-ends-covid-misinformation-policy-cc232c9ce0f193c505bbc63bf57ecad6

[72] Natascha Klocker. 2015. Participatory action research: The distress of (not) making a difference. *Emotion, Space and Society* 17, (November 2015), 37–44. DOI:https://doi.org/10.1016/j.emospa.2015.06.006

[73] André Luiz Martins Lemos, Elias Cunha Bitencourt, and João Guilherme Bastos dos Santos. 2021. Fake news as fake politics: the digital materialities of YouTube misinformation videos about Brazilian oil spill catastrophe. *Media, Culture & Society* 43, 5 (July 2021), 886–905. DOI:https://doi.org/10.1177/0163443720977301

[74] Théophile Lenoir and Chris Anderson. 2023. Introduction Essay: What Comes After Disinformation Studies. *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill* (January 2023). Retrieved February 8, 2023 from https://citap.pubpub.org/pub/oijfl3sv/release/1

[75] Darren Linvill and Patrick Warren. 2023. Interpreting the 'Twitter Files': Lessons About External Influence on Content Moderation. *Lawfare*. Retrieved February 9, 2023 from https://www.lawfareblog.com/interpreting-twitter-files-lessons-about-external-influence-content-moderation

[76] Ioana Literat, Yoo Kyung Chang, and Shu-Yi Hsu. 2020. Gamifying fake news: Engaging youth in the participatory design of news literacy games. *Convergence* 26, 3 (June 2020), 503–516. DOI:https://doi.org/10.1177/1354856520925732

[77] Tim K. Mackey, Vidya Purushothaman, Michael Haupt, Matthew C. Nali, and Jiawei Li. 2021. Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter. *The Lancet Digital Health* 3, 2 (February 2021), e72–e75. DOI:https://doi.org/10.1016/S2589-7500(20)30318-6

[78] Randi Markussen. 1994. Dilemmas in Cooperative Design. *Participatory Design Conference* (October 1994). Retrieved December 1, 2022 from https://ojs.ruc.dk/index.php/pdc/article/view/108

[79] Alice E. Marwick, Shannon McGregor, and Elaine Schnabel (Eds.). 2022. What Comes After "Disinformation"? *Bulletin of Technology and Public Life* (October 2022). Retrieved February 11, 2023 from https://citap.pubpub.org/what-comes-after-disinformation

[80] Alice Marwick, Rachel Kuo, Shanice Jones Cameron, and Moira Weigel. 2021. Critical Disinformation Studies: A Syllabus. Retrieved from https://citap.unc.edu/research/critical-disinfo/

[81] Adrienne L. Massanari. 2018. Rethinking Research Ethics, Power, and the Risk of Visibility in the Era of the "Alt-Right" Gaze. *Social Media + Society* 4, 2 (April 2018), 2056305118768302. DOI:https://doi.org/10.1177/2056305118768302

[82] Meta. Facebook Community Standards. *Meta Transparency Center*. Retrieved December 7, 2022 from https://transparency.fb.com/policies/community-standards/

[83] Heidi J. Nast. 1994. Women in the field: Critical feminist methodologies and theoretical perspectives. *Professional Geographer* 46, 1 (February 1994), 54. DOI:https://doi.org/10.1111/j.0033-0124.1994.00054.x

[84] National Center for Immunization and Respiratory Diseases. 2022. COVID-19 Vaccine Confidence. *Centers for Disease Control and Prevention*. Retrieved December 7, 2022 from https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html

[85] Sarah Nguyễn, Rachel Kuo, Madhavi Reddi, Lan Li, and Rachel E. Moran. 2022. Studying mis- and disinformation in Asian diasporic communities: The need for critical transnational research beyond Anglocentrism. *Harvard Kennedy School Misinformation Review* (March 2022). DOI:https://doi.org/10.37016/mr-2020-95

[86] Shuo Niu, Keegan Veazey, Phoenix Pagan, and Abhisan Ghimire. 2022. Understanding Hate Group Videos on YouTube. In *Companion Computer Supported Cooperative Work and Social Computing*, ACM, Virtual Event Taiwan, 32–36. DOI:https://doi.org/10.1145/3500868.3559465

[87] Gianluca Nogara, Padinjaredath Suresh Vishnuprasad, Felipe Cardoso, Omran Ayoub, Silvia Giordano, and Luca Luceri. 2022. The Disinformation Dozen: An Exploratory Analysis of Covid-19 Disinformation Proliferation on Twitter. In *14th ACM Web Science Conference 2022* (WebSci '22), Association for Computing Machinery, New York, NY, USA, 348–358. DOI:https://doi.org/10.1145/3501247.3531573

[88] Ihudiya Finda Ogbonnaya-Ogburu, Angela D.R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–16. DOI:https://doi.org/10.1145/3313831.3376392

[89] Pedro Oliveira and Luiza Prado. 2015. Cheat Sheet for a Non (or Less) Colonialist Speculative Design. *A Parede*. Retrieved February 13, 2023 from https://medium.com/a-parede/cheat-sheet-for-a-non-or-less-colonialist-speculative-design-9a6b4ae3c465

[90] Jonathan Corpus Ong and Ross Tapsell. 2022. Demystifying disinformation shadow economies: fake news work models in Indonesia and the Philippines. *Asian Journal of Communication* 32, 3 (May 2022), 251–267. DOI:https://doi.org/10.1080/01292986.2021.1971270

[91] Julian E. Orr. 1996. *Talking about Machines: An Ethnography of a Modern Job* (1st edition ed.). ILR Press, Ithaca, N.Y.

[92] Christine Øye, Nelli Øvre Sørensen, and Stinne Glasdam. 2016. Qualitative research ethics on the spot: Not only on the desktop. *Nurs Ethics* 23, 4 (June 2016), 455–464. DOI:https://doi.org/10.1177/0969733014567023

[93] Reece Peck. 2023. The Power of News Style and the Limits of Technology: Thinking Beyond the "Infocentric" Orientation of Disinformation Studies. *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill* (January 2023). Retrieved February 9, 2023 from https://citap.pubpub.org/pub/vzttydqz/release/1

[94] Gordon Pennycook and David G. Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116, 7 (February 2019), 2521–2526. DOI:https://doi.org/10.1073/pnas.1806781116

[95] Whitney Phillips. 2018. The Oxygen of Amplification. Retrieved from https://datasociety.net/wp-content/uploads/2018/05/1_PART_1_Oxygen_of_Amplification_DS.pdf

[96] Anthony T. Pinter, Morgan Klaus Scheuerman, and Jed R. Brubaker. 2021. Entering Doors, Evading Traps: Benefits and Risks of Visibility During Transgender Coming Outs. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3 (January 2021), 272:1-272:27. DOI:https://doi.org/10.1145/3434181

[97] Christine Preiser. 2016. Conducting Open Participant Observations of Bouncers – Negotiating (In)visibility in Fieldwork*. *British Journal of Community Justice* (June 2016). Retrieved February 11, 2023 from https://mmuperu.co.uk/bjcj/articles/conducting-open-participant-observations-of-bouncers-negotiating-invisibility-in-fieldwork/

[98] Jan Willem van Prooijen. 2018. Empowerment as a Tool to Reduce Belief in Conspiracy Theories. In *Conspiracy Theories and the People Who Believe Them.*

[99] 432–442. Retrieved from https://www.researchgate.net/publication/329825465_Empowerment_as_a_Tool_to_Reduce_Belief_in_Conspiracy_Theories

[99] María Laura Ramírez Galleguillos, Damla Çay, Ipek Gürbüzsel, Ivon Bensason, Yekta Bakırlıoğlu, Asım Evren Yantaç, and Aykut Coşkun. 2022. What is Participatory Design in Turkey? In *Proceedings of the Participatory Design Conference 2022 - Volume 2*, ACM, Newcastle upon Tyne United Kingdom, 294–295. DOI:https://doi.org/10.1145/3537797.3537889

[100] Christine Reyna, Andrea Bellovary, and Kara Harris. 2022. The Psychology of White Nationalism: Ambivalence Towards a Changing America. *Social Issues and Policy Review* 16, 1 (2022), 79–124. DOI:https://doi.org/10.1111/sipr.12081

[101] Julie Ricard and Juliano Medeiros. 2020. Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil. *Harvard Kennedy School Misinformation Review* 1, 3 (April 2020). DOI:https://doi.org/10.37016/mr-2020-013

[102] Thomas Rid. 2020. *Active Measures: The Secret History of Disinformation and Political Warfare* (Illustrated edition ed.). Farrar, Straus and Giroux, New York.

[103] Sarah T. Roberts. 2019. *Behind The Screen: Content Moderation in the Shadows of Social Media.* Yale University Press.

[104] Róisín Ryan-Flood. 2009. Queering Representation: Ethics and Visibility in Research. *Journal of Lesbian Studies* 13, 2 (April 2009), 216–228. DOI:https://doi.org/10.1080/10894160802695387

[105] Kavous Salehzadeh Niksirat, Evanne Anthoine-Milhomme, Samuel Randin, Kévin Huguenin, and Mauro Cherubini. 2021. "I thought you were okay": Participatory Design with Young Adults to Fight Multiparty Privacy Conflicts in Online Social Networks. In *Designing Interactive Systems Conference 2021* (DIS '21), Association for Computing Machinery, New York, NY, USA, 104–124. DOI:https://doi.org/10.1145/3461778.3462040

[106] Amanda Seitz. 2021. Mob at U.S. Capitol encouraged by online conspiracy theories. *AP NEWS*. Retrieved March 16, 2021 from https://apnews.com/article/donald-trump-conspiracy-theories-michael-pence-media-social-media-daba3f5dd16a431abc627a5cfc922b87

[107] Scott Shane and Alan Blinder. 2018. Secret Experiment in Alabama Senate Race Imitated Russian Tactics. *The New York Times*. Retrieved January 24, 2023 from https://www.nytimes.com/2018/12/19/us/alabama-senate-roy-jones-russia.html

[108] Kate Starbird. Unraveling the Big Lie: Participatory Disinformation and the 2020 Election. In *The Conservative Dilemma*.

[109] Miriah Steiger, Timir J Bharucha, Sukrit Venkatagiri, Martin J. Riedl, and Matthew Lease. 2021. The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–14. DOI:https://doi.org/10.1145/3411764.3445092

[110] Lisa Stevenson. 2014. *Life Beside Itself: Imagining Care in the Canadian Arctic.* University of California Press.

[111] Angelika Strohmayer, Laura Cortés-Rico, Tania Pérez-Bustos, Afroditi Psarra, Daniela Rosner, Özge Subasi, Irene Posch, Sara Nabil, and Jihan Sherman. 2022. Justice-oriented Participatory Electronic Textile Making: Fostering shared spaces of knowledge dialogues through the process of making, un-making, and re-making justice-oriented participatory praxis. In *Proceedings of the Participatory Design Conference 2022 - Volume 2*, ACM, Newcastle upon Tyne United Kingdom, 259–262. DOI:https://doi.org/10.1145/3537797.3537880

[112] Lucy A. Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication* (2nd edition ed.). Cambridge University Press, Cambridge Cambridgeshire; New York.

[113] Matthew Sullivan. 2019. Libraries and Fake News: What's the Problem? What's the Plan? *Communications in Information Literacy* (2019). DOI:https://doi.org/10.15760/comminfolit.2019.13.1.7

[114] Stuart A. Thompson and Davey Alba. 2022. Fact and Mythmaking Blend in Ukraine's Information War. *The New York Times*. Retrieved February 11, 2023 from https://www.nytimes.com/2022/03/03/technology/ukraine-war-misinfo.html

[115] Cameron Tonkinwise. 2014. How We Intend to Future: Review of Anthony Dunne and Fiona Raby, Speculative Everything: Design, Fiction, and Social Dreaming. *Design Philosophy Papers* 12, 2 (December 2014), 169–187. DOI:https://doi.org/10.2752/144871314X14159818597676

[116] Carlos Roberto Torres-Parra, Roberto Cuervo-Pulido, Juliana Flórez-Flórez, and Omar Ramírez-Pérez. 2022. Diseño participativo como método para la creación de videojuegos críticos. Poder Violeta, estudio de caso de un videojuego sobre acoso sexual. *Revista Colombiana de Educación* 85 (May 2022). DOI:https://doi.org/10.17227/rce.num85-12568

[117] Kathie M. d'I. Treen, Hywel T. P. Williams, and Saffron J. O'Neill. 2020. Online misinformation about climate change. *WIREs Climate Change* 11, 5 (2020), e665. DOI:https://doi.org/10.1002/wcc.665

[118] Francesca Bolla Tripodi. 2022. *The propagandists' playbook: how conservative elites manipulate search and threaten democracy.* Yale University Press, New Haven.

[119] Elliot Trotter. 2020. At MisinfoNight, Ballard High School students share media literacy lessons. *Center for an Informed Public*. Retrieved January 22, 2023 from https://www.cip.uw.edu/2020/11/19/misinfonight-ballard-high-school/

[120] Katerina Tsetsura. 2023. The Brave New World (?) of Disinformation Research. *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill* (January 2023). Retrieved February 9, 2023 from https://citap.pubpub.org/pub/8u3v9p8w/release/1

[121] Eve Tuck and K. Wayne Yang. 2014. R-Words: Refusing Research. In *Humanizing Research: Decolonizing Qualitative Inquiry with Youth and Communities*, Django Paris and Maisha T. Winn (eds.). SAGE Publications, Inc., 1 Oliver's Yard, 55 City Road London EC1Y 1SP. DOI:https://doi.org/10.4135/9781544329611

[122] Eve Tuck and K. Wayne Yang. 2014. Unbecoming Claims: Pedagogies of Refusal in Qualitative Research. *Qualitative Inquiry* 20, 6 (July 2014), 811–818. DOI:https://doi.org/10.1177/1077800414530265

[123] Sukrit Venkatagiri, Anirban Mukhopadhyay, David Hicks, Aaron Brantly, and Kurt Luther. 2023. CoSINT: Designing a Collaborative Capture the Flag Competition to Investigate Misinformation. Association for Computing Machinery, New York, NY, USA, 32.

[124] Tamar Wilner, Kayo Mimizuka, Ayesha Bhimdiwala, Jason C. Young, and Ahmer Arif. 2023. It's About Time: Attending to Temporality in Misinformation Interventions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

[125] Tom Wilson and Kate Starbird. 2021. Cross-platform Information Operations: Mobilizing Narratives & Building Resilience through both "Big" & "Alt" Tech. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (October 2021), 345:1-345:32. DOI:https://doi.org/10.1145/3476086

[126] Sijia Xiao, Coye Cheshire, and Amy Bruckman. 2021. Sensemaking and the Chemtrail Conspiracy on the Internet: Insights from Believers and Ex-believers. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (October 2021), 1–28. DOI:https://doi.org/10.1145/3479598

[127] Jason C. Young. 2021. Disinformation as the weaponization of cruel optimism: A critical intervention in misinformation studies. *Emotion, Space and Society* 38, (February 2021), 100757. DOI:https://doi.org/10.1016/j.emospa.2020.100757

[128] Jason C. Young, Brandyn Boyd, Katya Yefimova, Stacey Wedlake, Chris Coward, and Rolf Hapel. 2021. The role of libraries in misinformation programming: A research agenda. *Journal of Librarianship and Information Science* 53, 4 (December 2021), 539–550. DOI:https://doi.org/10.1177/0961000620966650

[129] Alex Zahara. 2016. Ethnographic Refusal: A How to Guide. *Discard Studies*. Retrieved February 12, 2023 from https://discardstudies.com/2016/08/08/ethnographic-refusal-a-how-to-guide/

[130] Cat Zakrzewski. 2022. Election workers brace for a torrent of threats: 'I KNOW WHERE YOU SLEEP.' *Washington Post*. Retrieved January 9, 2023 from https://www.washingtonpost.com/technology/2022/11/08/election-workers-online-threats/