Ethical Participatory Design of Social Robots Through Co-Construction of Participatory Design Protocols

Isha Datey[1,a], Hunter Soper [1], Khadeejah Hossain [1], Wing-Yue Geoffrey Louie[1], Douglas Zytko [1,b]

***Abstract*— Ethics have become a core consideration in human-robot interaction (HRI) due to ample opportunity for both positive and negative impact on humans. HRI literature has expounded on ways to produce ethical social robots, especially participatory design (PD) that integrates anticipated users and other stakeholders as designers themselves to ensure their values are integrated into robot design. We draw attention to the ethics of participation in robot design, distinct from the ethics of the robot ultimately designed. We propose an approach to foregrounding ethics in PD processes through co-construction of robot PD protocols with stakeholders. We call this "pre-PD" because it entails expanding the boundaries of PD beyond the product of design (the robot) to also include the participatory activities that enable design. Contributions of the paper include: (1) a case study of pre-PD for sexual violence mitigation robots to demonstrate feasibility of stakeholders co-constructing robot PD protocols, and (2) an actionable framework for HRI researchers to use when constructing their own PD protocols with stakeholders, informed by reflection on the case study.**

I. INTRODUCTION

Ethics have become a core consideration in human-robot interaction (HRI) [1]-[5] and are particularly germane to social robots given their purpose centers on social interaction. This poses ample opportunity for both positive and negative impact on humans. The HRI literature has elucidated ways to deliberately incorporate ethics into a social robot's design, including value sensitive design [6], [7] and participatory design (PD) [8]. The latter entails directly integrating anticipated users and other stakeholders into robot design processes as designers and key decision-makers themselves so that their values and ethical sensitivities can be directly incorporated into a robot's design.

In this paper we foreground ethics *during* PD, which is distinct from the ethics of the technologies being co-designed, in order to consider and prevent adverse impacts on stakeholders incurred through participation in design [9]-[13]. In short, we focus attention on the ethics of a PD protocol, distinct from the ethics of the product of PD. Discussions and actionable approaches to foregrounding ethics in PD protocols are relatively absent in HRI, despite unique ethical concerns that can be exacerbated or introduced by co-design of social robots, particularly along the dimensions of harm, exploitation, and agency. As just a few examples:

1) Harm: Emergent behaviors of prototypical social robots could lead to physical *harm* of participatory designers in ways not possible with non-embodied technologies.
2) Exploitation: The often-lengthy design and development processes for social robots can inadvertently lead to *exploitation* of stakeholders if the immense time and resources they commit to the project are not commensurate with realized benefits.
3) Agency: There is risk to stakeholder *agency* due to disparities in expertise between researcher and stakeholder. Robot PD protocols often involve intricately structured design activities that, while beneficial to remedying stakeholder confusion, may raise questions as to what influence researchers have on the ideas and decisions of stakeholders through PD protocols.

We propose an approach to foregrounding ethics in PD processes through co-construction of robot PD protocols with stakeholders. We call this *"pre-PD"* because it entails expanding the boundaries of PD beyond the product of design (the robot) to also include the participatory structures, activities, and processes that enable design. This allows stakeholders to identify and remedy ethical concerns that may arise during the *process* of robot design. Contributions of the paper include:

- A case study to demonstrate feasibility of stakeholders co-constructing robot PD protocols. Using sexual violence mitigation robots as a context, we unpack our methodological choices, challenges, and insights from involving diverse stakeholders with minimal (or no) familiarity with social robots in PD protocol construction.
- Through personal reflection on our case study we present an actionable framework to help HRI researchers prepare for conducting their own co-construction of robot PD protocols with stakeholders.

II. RELATED WORK

We first review prior work in PD of social robots to elucidate the method's potential for producing ethical robot designs. We then motivate the importance of considering ethics of participation in robot design and contextualize our approach to foregrounding ethics during PD among others from the broader HCI literature.

[1]Department of Computer Science and Engineering, Oakland University, Rochester, MI [a]ishadatey@oakland.edu, [b]dzytko@umich.edu

I. Datey, H. Soper, K. Hossain, W. G. Louie, and D. Zytko, "Ethical Participatory Design of Social Robots Through Co-Construction of Participatory Design Protocols," 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Busan, South Korea, 2023, pp. 1-7, doi: 10.1109/RO-MAN57019.2023.10309539.

The final version of this record is available at: https://ieeexplore.ieee.org/document/10309539.

A. Participatory Design of Ethical Robots

The ethics of HRI are a common point of discussion and debate [1]-[5], [14], often motivated by the unintended consequences of robots [15]-[17] and the ethically gray areas of how a robot could shape human behavior [18]-[20]. The field has produced and reflected on codes of ethics–or rules–for determining if a designed robot is ethical [20]-[22], although the notion of generalized ethical guidelines has received critique on the basis that ethical challenges are context-specific and cannot be satisfactorily addressed with a one-size-fits-all list of ethical principles [3], [23], [24].

In response to this, other work has proposed ethical frameworks that can be applied to HRI research and practice on a case-by-case basis, such as Ostrowski and colleagues' interpretation of the Design Justice Framework for HRI [1] that elucidates seven areas for ethical inquiry. Two of these areas, Equity and Beneficiaries, frame participatory design (PD) as a key approach towards context-sensitive robot ethics: *"The Equity area adapted for an HRI context focuses on who is included in robot design, seeking to promote and encourage researchers to leverage participatory methodologies [and the] beneficiaries area has a particular focus on the intended users for robots and calls for mapping out desires, needs, and preferences for robot design among various intended users"* (p. 5).

PD has gained considerable traction in HRI [25]-[31],[36], [48], [49] with a diverse range of stakeholders involved in robot design such as children [49], the elderly [25], [48], [36], and people with disabilities [27], [28], [30], [31]. Robot PD is typically conducted through synchronous and repeating design workshops [30], [31], supported by physical tools, worksheets, and software to scaffold ideation. Some design activities are explicitly intended to incorporate ethics in a robot's design, such as Axelsson and colleagues' "canvases" or worksheets to guide stakeholders through making key decisions for a robot's design [8].

B. Foregrounding Ethics in the Act of Robot Design

While PD has been lauded as an approach to producing ethical robots, there has been little consideration in HRI of the ethics of the PD protocol or process used to produce an ethically sound robot.

The field can agree that robots should not harm humans; we extend this thinking to robot design processes—the act of participation in robot design should not result in harm. While this form of meta-ethics, so to speak, is a new quandary within HRI, the broader human-computer interaction (HCI) literature has begun to contend with it, in some cases by offering first-hand examples of participatory designers feeling "manipulated" and disrespected by otherwise well-intentioned PD protocols [34].

It can be tempting to pursue a generalized set of guidelines for conducting ethical PD protocols given the obvious benefit of scalability, yet this idea has been cautioned against. Per Spiel et al. [35]: "The general approach of ethics guidelines systematically overlooks a multitude of situated judgments" that PD facilitators make "on the spot." Read and colleagues [33] call on PD facilitators to reflect on "whose values are being considered" in a PD procedure - generalized PD guidelines may subvert this reflection due to the assumption that values derived from one (or more) stakeholder groups can be transposed to others.

Instead, the literature advocates for situated, context-specific approaches to foregrounding ethics in PD processes [32]-[35], amongst which pre-PD can be positioned. One approach involves checklists of questions to prompt researchers to self-reflect on ethics of their PD protocols. CHECK 1 and 2 (founded on value-centered/sensitive design) are value checklists consisting of questions that researchers should ask themselves before a PD study [33]. Another example is the reflectivity reminder card [34], with questions derived from various forms of ethics (ethics-of-the-other, pragmatist ethics, and virtue ethics) and exemplified with anecdotes from Steen's PD work regarding in-situ adjustments and post-PD reflection.

A limitation to self-reflection approaches is a researcher's bias and blindspots: they may not identify ethical concerns that would be more apparent to their target stakeholders. Supplementing these self-reflection exercises with pre-PD can confirm and expand one's ethical reflections.

Another approach, based on micro-ethics, is found in Spiel et al. [35]: "We suggest actively identifying ethically charged situations after each encounter with participants, determining the choices and the judgments made and then reflecting on them with others." A similar strategy of affording stakeholders the opportunity to "reflect [on] the co-design process throughout the year" is recommended in Ostrowski and colleagues' long-term co-design guidelines for robots [36].

A key difference between these mid-process reflections and pre-PD is when the involvement of others occurs: in pre-PD it is before PD begins, which enables researchers to preemptively identify "ethically charged situations" and mitigate ethical concerns with stakeholder input. Steen [34] recounts instances when they "missed several opportunities to learn from [stakeholders mid-project] and to let their ideas affect the project" that were noticed only after post-PD ethical reflection, and Spiel et al. [32], [35] write "researchers are required to make judgments on the spot, which either may have been unforeseeable or may create a contradiction to over-arching ethical principles."

We do not intend to disparage mid-project reflection, but rather to demonstrate the benefit of combining it with pre-PD to continually reflect on ethical decisions before, during, and after PD. Pre-PD could help researchers preempt some of these formerly unforeseeable situations and mid-project challenges, instead of relying largely on reactive measures imposed only after harm or adverse impact has occurred.

III. CASE STUDY: SEXUAL VIOLENCE MITIGATION ROBOTS

We personally facilitated pre-PD in the context of sexual violence mitigation robots, which we unpack here to demonstrate the method's utility. Sexual violence (SV) involves any "sexual act that is committed or attempted by another person

without freely given consent of the victim" [37]. SV is a challenging problem because it often occurs without conscious intent to cause harm due to problematic consent practices [38] that can obfuscate whether a partner actually agreed to sex (e.g., relying on nonverbal cues [39] or assuming consent through time and place [40]). Various technology-facilitated tools have been studied and proposed for sexual violence (e.g., [41]-[43]), although seldom with robotics. Social robots are an opportune emerging technology to consider for prevention of SV because their physical embodiment can mediate sexual interactions when such mediation is most needed: when nonconsensual sex is about to occur. We opted to pursue robot-assisted SV mitigation with PD because the success of such robots is contingent on willingness of users to incorporate them into their most intimate interactions.

A. Pre-PD Method

We conducted an IRB-approved study with 19 stakeholders to co-construct a protocol for PD of sexual violence mitigation robots. Demographics recruited were women [44] and LGBTQ+ individuals [45] due to their disproportionate risk of SV as well as practitioners/providers of SV victim services and researchers of SV and adjacent topics (e.g., gender, sexuality) to leverage their professional experience. We opted for a combination of purposive and snowball sampling. This started with the lead researcher emailing SV practitioners and researchers in the geographic area while student researchers tailored recruitment messages to woman- and LGBTQ-identifying university students who had demonstrated interest in SV or intersectional issues through research publications, course projects, and activism campaigns. See Table I for demographic details.

The PD protocol was co-constructed over two rounds of sessions with a total of 19 stakeholders (15 stakeholders in round 1 and 12 in round 2; 8 participated in both rounds). Sessions across both rounds ranged from 57-127 minutes, totaling 1012 minutes of session time. Fifteen of the 19 stakeholders participated in private sessions through Zoom video calls. Four participants preferred to participate in-person as a group at a secluded table in a restaurant for mutual social support given all had previously been victims of SV. Each session was moderated by 2-3 members of the research team who alternated between taking notes on the conduct of the pre-PD activities and verbally interacting with the stakeholders.

In round 1, each stakeholder constructed a specific element of the PD protocol that reflected on ethical considerations most important to them. This ideation was facilitated through a loosely structured and highly visual presentation by the research team about PD (e.g., common design activities) and social robots to stimulate open conversation. For example, P12-15 gravitated to activities for evaluating prototypical robot designs in scenarios where it would impact their sexual activity due to perceived risks of physical harm if the robot functions incorrectly. P7 focused on emotional support and counseling structures to be integrated into design activities due to concerns of victims reliving past SV trauma when explaining and justifying their SV mitigation robot designs.

Round 1 transcripts were then analyzed using a constructive open coding approach [46] to identify similarities and differences in PD protocol ideas. These were organized in a virtual worksheet along the following dimensions: recruitment and incentives, stakeholder comfort, robot design ideation activities, robot prototyping, and evaluation/testing of prototypes. Round 2 sessions were intended for stakeholders to converge on a singular, final PD protocol

TABLE I
DEMOGRAPHIC DETAILS OF STAKEHOLDERS [1]

P	Age	Ethnicity	Gender, sexuality	Self-described traits	Rounds participated
1	20	White	Queer	Equal rights activist	Round 1 (ind.[2])
2	21	Black	Woman	Published on SV research	Round 1 (ind.), Round 2 (grp.[3])
3	24	Black	Woman	Degree in Family Studies	Round 1 (ind.)
4	26	White	Non-Binary; pansexual	Published on computer-mediated SV	Round 1 (ind.), Round 2 (grp.)
5	38	White	Man; gay	Gender and sexuality researcher; familiar with PD	Round 1 (ind.), Round 2 (grp.)
6	n/a	White	Woman	Certified Sexual Assault Nurse Examiner	Round 1 (ind.), Round 2 (grp.)
7	42	White	Woman	Published extensively on SV research	Round 1 (ind.), Round 2 (grp.)
8	25	Asian	Woman	Cyber security background	Round 1 (ind.)
9	27	White	Man; bisexual	Professional voice actor	Round 1 (ind.)
10	n/a	White	Woman (she/they)	College student (writing major)	Round 1 (ind.)
11	20	Indian	Woman	Immigrant to US; AI PD researcher	Round 1 (ind.)
12	27	Asian	Woman; heterosexual	Self-identified victim of sexual violence	Round 1 (grp.)
13	27	Asian	Woman; heterosexual	Self-identified victim of sexual violence	Round 1 (grp.), Round 2 (grp.)
14	30	Asian	Woman; heterosexual	Self-identified victim of sexual violence	Round 1 (grp.), Round 2 (grp.)
15	31	Asian	Woman; heterosexual	Immigrant to US; Self-identified victim of sexual violence	Round 1 (grp.), Round 2 (grp.)
16	26	White	Woman; heterosexual	Psychology, HCI researcher	Round 2 (grp.)
17	27	Chinese	Woman; heterosexual	Quantitative HCI Researcher	Round 2 (grp.)
18	25	Black	Woman; bisexual	HRI researcher	Round 2 (grp.)
19	n/a	White	Woman	Published on SV research	Round 2 (grp.)

[1] Some opted not to disclose their age, gender, and/or sexuality.
[2] Individual sessions
[3] Group sessions

selecting from the Round 1-proposed options in each of the aforementioned categories. This decision-making was performed sequentially, with initially-booked stakeholders in round 2 (P12-15) making their decisions, after which the next stakeholder confirmed or revised those previous decisions, and so on until the last-booked stakeholder who confirmed/finalized the decisions about the now-completed PD protocol. Stakeholders were allowed to "pass" a decision onto the next stakeholder if they were undecided.

B. Co-Constructed Participatory Design Protocol

The two-round pre-PD process culminated in stakeholder-made decisions about the following aspects of a protocol for PD of an SV mitigation robot: recruitment and incentives, stakeholder comfort, robot design ideation activities, robot prototyping, and evaluation/testing of prototypes. The decisions for these categories were visually documented and updated in a Miro board during each round-2 session [Figure 1]. Audio transcripts were also analyzed with a constructive coding process [46] for justification and elaboration on the decisions.

The protocol is not completely linear and sessions, particularly for design and prototyping, should repeat as necessary.

1) Recruitment: The PD process should primarily consist of those who are most at risk of or have experience with SV (women, LGBTQ+ individuals, college students, online daters, and young adults aged 18-30 were mentioned by stakeholders). Other groups, particularly heterosexual men, were also encouraged for inclusion, but in separate design sessions to avoid discomfort that their presence may pose to the aforementioned groups.

2) Compensation: There should be four types of compensation offered for participation. The most commonly advocated choices were food (take home and in-session) and money ($20-$40 USD per hour). Participatory designers should also be afforded long-term access to therapy (described as the longitudinal duration of the robot development process) due to risk of re-traumatization from memories of SV experience and disclosure of such experiences to other designers. Stakeholders emphasized that limited therapy access (e.g., one free session) would *not* be acceptable because restricted access to continued healthcare could worsen one's emotional state. The fourth was professional recognition for participation in design such as participation certificates, thank you notes, and opportunities for co-authorship on future publications if a participatory designer is an academic.

3) Comfort structures: Given the sensitivity of the subject matter, participatory designers are to be given choices regarding modality of participation (in-person or online) and presence of others (solo or group design sessions - with transparency over the demographics present in a given group session). New participants that join midway in the PD process should not join an ongoing group composition due to camaraderie-based comfort that may have already developed.

Comfort structures during design sessions were also requested. These include a "safety button" that allows a participant to skip an uncomfortable question. A dedicated button was recommended to normalize the option of foregoing participation; requiring participants to actively verbalize discomfort could itself be uncomfortable. Stakeholders also recommended a clear and comprehensive informed consent process throughout participation that goes beyond minimum standards set by an IRB. Suggestions for the consent form included best/worst-case scenarios for participation, trigger warnings, specific expectations for participation, and how participation can be discontinued.

4) Robot design ideation activities and prototyping: Group design sessions should be a maximum of 90-120 minutes with a maximum of 4 to 5 people, and individual sessions a maximum of 60 minutes. Sessions should leverage scenario-based design activities in which researchers first provide several abstract scenarios of an SV mitigation robot (e.g., a robot that reminds a person to ask for permission before touching the other's body) to familiarize participants of the possible scope of design. Participants would then produce their own verbal or written scenarios that articulate where their envisioned SV mitigation robot would be used and how it would function at a conceptual level. Worksheets/canvases [8] are to be provided to participants to aid in brainstorming (selection or construction of specific worksheets was not decided in our pre-PD method however; this limitation is discussed in the next section). Participants would then produce low fidelity sketches to add specificity to their scenarios. Group discussion (if a group session) would be interspersed after both of these activities.

Sketches would then be translated into virtual reality (VR) prototypes by researchers and/or participants if they are technically versed. For those involved in groups, further discussions/sessions would inform iterations to the VR-based robot designs. The purpose of these recurrent group discussions is to reach consensus on a singular (or limited set) of SV mitigation robot designs for physical prototyping. This may include a formal voting process among all groups and individually-participating designers.

5) Evaluation/testing of prototypes: Stakeholders recommended giving participants three options for evaluating physical prototypes of their SV mitigation robot designs. One pertained to actors simulating scenarios of use which participants observe and then discuss in terms of whether the robot's actions match their expectations. Another option is to allow participants to interact with the robot themselves if they feel comfortable. Although this option may have limited practicality depending on the sensitivity of contexts in which the robot is expected to be used (e.g., in one's bedroom during a sexual encounter). The third option involves letting participants take the prototypical robot home to interact with in a private setting, after which they would provide written or verbal self-reports to the research team. Biometrics could also be collected if consented to by participants.

IV. A FRAMEWORK FOR CONDUCTING PRE-PD

Another data source produced in our method involved internal team notes about the conduct of pre-PD to inform changes to subsequent sessions with stakeholders. These

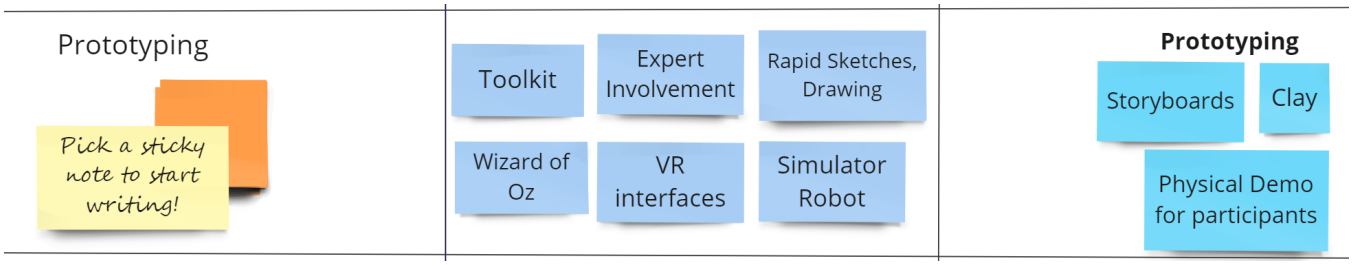

Fig. 1. A part of the Miro board worksheet showing the 'prototyping' idea section. Leftmost column: space for the stakeholder to make a decision on the given protocol element; center column: ideas from the literature (originally from the round-1 Powerpoint presentation); right most column: ideas from previous stakeholders.

notes were analyzed through an iterative card sorting process, culminating in a series of questions that we as researchers did or should have asked ourselves when preparing our pre-PD method. We report on those questions here because we consider them to constitute a decision-making framework for HRI researchers to follow when preparing their own pre-PD method.

A. Preparing for Pre-PD

1) What is the scope of what can be designed and decided in PD?: Often times robot PD initiatives begin with practical constraints (or pre-made decisions) that limit what stakeholders can design or change. Examples include the context in which the robot will be used, the timeline for development, and technical capabilities of the robot. We had purposely avoided stating any constraints in our round 1 sessions, yet some of our stakeholders found this "openness" of scope to be overwhelming and they actually requested some constraints to hone their ideation. In response to this we arbitrarily clarified the anticipated use of the SV mitigation robot to be in a private bedroom when a sexual act between two partners is about to occur. We recognize that future HRI researchers may have legitimate (rather than manufactured) scope constraints, and we encourage transparency of these constraints as early as possible to lend clarity to stakeholders about what can and cannot be modified in a PD protocol.

2) What ethical concerns are motivating pre-PD?: Researchers may have an initial set of ethical concerns motivating co-construction of a robot PD protocol. We opted to be transparent about these with stakeholders early in round 1 sessions, which they appreciated and often elaborated on or added to. Almost all ideation of individual elements of the PD protocol stemmed from identification of an ethical concern of most interest to each stakeholder. We thus recommend that HRI researchers acknowledge their own ethical concerns with stakeholders early because it can be an effective way to stimulate ideation and decision-making for the PD protocol. For reference, the ethical concerns that we broached to stakeholders were emotional harm (re-traumatization of SV victims), exploitation (potential for researchers to be insensitive to the invisible costs of participation in robot design by marginalized groups), and agency (inadvertently imposing a sexual consent practice onto the robot's design that stakeholders may inwardly disagree with).

3) Which stakeholders should be involved in co-constructing the PD protocol?: There may be a range of backgrounds and experiences that would be informative to co-constructing a PD protocol for a given robot; these demographics may go beyond those that would be expected to participate in the eventual robot design sessions. Our case study involved a diverse group of stakeholders including individuals at disproportionate risk of SV, practitioners of SV victim services, and researchers with expertise in adjacent topics. In retrospect this diversity was quite beneficial to the PD protocol. For example, stakeholders aligning with at-risk groups (women and LGBTQ+) appeared to have an easier time imagining themselves as eventual designers and users of an SV mitigation robot. Their attention gravitated to how the PD protocol could sustain their participation, resulting in ideas around variable forms of participation (individual vs. group sessions) and non-financial compensation structures such as therapy. Practitioners and researchers focused on other aspects of the PD protocol reflective of their professional experience. For example, P5 centered on inclusive procedures for data analysis and decision-making while referencing literature about participatory studies that were negatively received by LGBTQ+ stakeholders involved. We encourage HRI researchers to seek out diverse perspectives in their own pre-PD methods to address "blind spots" that stakeholders from only one demographic may have.

B. The Act of Participation

1) How should stakeholders be prepared for participation?: Stakeholders have varying knowledge gaps that may prevent them from immediately understanding or making decisions about PD protocols for social robots. HRI researchers may be inclined to incorporate preparatory materials or steps for stakeholders, however we urge consideration of how these may accidentally discourage participation on the grounds of under-qualification. In round 1 sessions we prepared a presentation with introductory content about PD, social robots, and SV as reference/backup material for the stakeholder. Initially these slides were quite intricate, amounting to literature reviews in visual form. Presenting these slides took significant time and inadvertently reminded some stakeholders of their own gaps in knowledge, leading a few to openly doubt if they were qualified to inform the robot PD protocol. The complexity of the slides was significantly reduced in

later sessions to normalize the stakeholder talking earlier and more frequently. We also engaged in two strategies to establish each stakeholder as an expert uniquely qualified to inform the robot PD protocol. One was to "dismantle" our own expertise through self-deprecating comments about our abilities to construct a "good" PD protocol on our own, often with specific requests for assistance to make the stakeholder feel needed (e.g., the researcher exclaiming "I'm not really confident in any of our own ideas about who should be involved in making this robot. What do you think?"). Another strategy involved starting the session by having the stakeholder give a lengthy introduction about themselves during which we remarked on traits and experiences that made the stakeholder uniquely qualified for informing the PD protocol ("building up" the stakeholder's confidence).

2) How should stakeholders be involved in co-construction of the PD protocol?: There are a range of ways in which stakeholders may contribute to the construction of a PD protocol. We opted for a two-round process, with the first round for divergent thinking (ideation of multiple, potentially divergent, PD protocol ideas) and the second round for convergent decision making (distilling the multiple PD protocol ideas into a singular PD protocol). Motivated by the SV literature highlighting systemic issues with loss of agency of sexual experiences when dealing with authority figures [51] we left round 1 sessions deliberately open-ended so that our stakeholders could direct us to elements of the PD protocol they deemed most important. We had initially intended for each stakeholder to create a complete PD protocol in these divergent, round 1 sessions but quickly found that untenable due to time constraints and the importance of having stakeholders discuss and reflect on specific PD protocol ideas with us. Round 2 sessions were much more structured, during which stakeholders engaged with a Miro board to make decisions about clearly identified components of a robot PD protocol.

A limitation of our approach is that stakeholders did not individually produce ideas for every component of the PD protocol (although they did have an opportunity to weigh in on final decisions about each component). In retrospect we should have made round 1 sessions more structured, such as by identifying specific aspects of a PD protocol that "must" be deliberately designed. This may have produced more individual ideas for the PD protocol. We did provide such a structure in round 2 by identifying aspects of a PD protocol for collective decision-making: recruitment and incentives, stakeholder comfort, robot design ideation activities, robot prototyping, and evaluation/testing of prototypes.

3) What is produced by stakeholders during co-construction of the PD protocol?: What form does a co-constructed PD protocol take, and what may individual stakeholders be expected to produce that represents "their" protocol? Pre-identifying such artifacts may ease the need for "translation" of stakeholders' ideas and ensure the protocol accurately reflects their intentions. Our round 1 sessions produced purely verbal ideas for a robot PD protocol, which we transcribed for incorporation into a visual Miro board. A limitation of these artifacts is that specificity of some protocol decisions was inevitably under-developed, especially structures for scaffolding PD activities. For instance, many design activity options were suggested by stakeholders - drawing, quick writing, scenario-based ideas, prompt-based discussions, rapid on-the-spot design visualization and iteration using a 3D/2D paint function or a custom robot designer during focus groups. It was left to us researchers to decide if and when to implement any or all of these myriad of ideas contingent on what felt most appropriate to eventual participants and the context (e.g., whether to use rapid sketching or not).

One could argue that specificity of PD protocol decisions could be improved through further rounds of pre-PD activities. However, most of our stakeholders indicated comfort with allowing us researchers to settle the specifics of some protocol elements like design worksheets because they were confident their underlying ideas were taken into account for the protocol already. We recommend future HRI researchers pre-identify the PD protocol "artifacts" in their pre-PD method, and allow stakeholders to clarify the extent with which they want to be involved in the specificities of PD protocol materials.

4) How are decisions about the PD protocol made?: When involving multiple stakeholders in brainstorming for a robot PD protocol there is the possibility, if not probability, of conflicting ideas. How convergent decision-making is conducted can have ramifications on stakeholders' perceptions of whether their ideas have been taken seriously. We opted for a sequential decision-making approach in our round 2 sessions in which stakeholders individually validated the decisions made by stakeholders from previous sessions (e.g., stakeholder 1 selects from x different options for robot prototype evaluation, stakeholder 2 validates or changes that design, stakeholder 3 validates or changes stakeholder 2's decision, and so forth until the last stakeholder makes the final decision). We acknowledge that this decision-making process gives disproportionate power to stakeholders booked last, however we opted for this process at the recommendation of stakeholders in the earlier sessions whose decisions were most susceptible to being overridden. Their reasoning was to enable future stakeholders to make more informed protocol decisions through awareness of the preferences of stakeholders that came before them.

One notable limitation of our decision-making process - and our pre-PD method more generally - is the relative abstraction of later parts of the PD protocol chronologically speaking. For example, decisions around recruitment and initial design sessions were much more specific than decisions around how SV-mitigation robot prototypes should be evaluated. This was due to dependencies of later portions of the protocol on previous design stages (e.g., stakeholders could not offer specific methods for testing a robot prototype without knowing what that specific robot prototype is intended to do and what form it might take). We would not consider this a fault of our pre-PD method so much as an indicator that PD protocol co-construction should be revisited

throughout the conduct of the PD process.

V. CONCLUSION

The ethics of social robots have been widely discussed in literature and participatory design has been proposed and used to produce ethical robots. This paper distinguishes the ethics of a robot's design from the ethics of the design process leading to the robot. The potential adverse effects of stakeholder participation in robot design are highlighted, and to remedy potential ethical concerns of such participation the authors advocate for expanding stakeholder involvement into the co-construction of the PD protocol - called pre-PD. The paper presents a case study of pre-PD for sexual violence mitigation robots to demonstrate the feasibility of involving a diverse range of stakeholders in co-constructing a PD protocol. Through reflection on the case study the paper concludes with a framework for key decisions researchers must make in scaffolding stakeholder participation in constructing robot PD protocols.

ACKNOWLEDGMENTS

This material is based upon work partially supported by the U.S. National Science Foundation under Grant No. IIS-2211896.

REFERENCES

[1] A. K. Ostrowski et al., "Ethics, Equity, & Justice in Human-Robot Interaction: A Review and Future Directions," 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Napoli, Italy, 2022, pp. 969-976, doi: 10.1109/RO-MAN53752.2022.9900805.

[2] Carros, Felix, Störzinger, Tobias, Wierling, Anne, Preussner, Adrian and Tolmie, Peter. "Ethical, Legal & Participatory Concerns in the Development of Human-Robot Interaction: Lessons from Eight Research Projects with Social Robots in Real-World Scenarios" i-com, vol. 21, no. 2, 2022, pp. 299-309. https://doi.org/10.1515/icom-2022-0025

[3] Colombino T, Gallo D, Shreepriya S, Im Y and Cha S (2021) Ethical Design of a Robot Platform for Disabled Employees: Some Practical Methodological Considerations. Front. Robot. AI 8:643160. doi: 10.3389/frobt.2021.643160

[4] Riek, Laurel, and Don Howard. "A code of ethics for the human-robot interaction profession." Proceedings of we robot (2014).

[5] M. Brandão, "Normative roboticists: the visions and values of technical robotics papers," 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Vancouver, BC, Canada, 2021, pp. 671-677, doi: 10.1109/RO-MAN50785.2021.9515504.

[6] EunJeong Cheon and Norman Makoto Su. 2016. Integrating roboticist values into a Value Sensitive Design framework for humanoid robots. In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 375–382. DOI:https://doi.org/10.1109/HRI.2016.7451775

[7] Haiyi Zhu, Bowen Yu, Aaron Halfaker, and Loren Terveen. 2018. ValueSensitive Algorithm Design. Proc. ACM Human-Computer Interact. 2, CSCW (November 2018), 1–23. DOI:https://doi.org/10.1145/3274463

[8] Minja Axelsson, Raquel Oliveira, Mattia Racca, and Ville Kyrki. 2022. Social Robot Co-Design Canvases: A Participatory Design Framework. ACM Trans. Human-Robot Interact. 11, 1 (March 2022), 1–39. DOI:https://doi.org/10.1145/3472225

[9] L. Dombrowski, E. Harmon and S. Fox, "Social justice-oriented interaction design: Outlining key design strategies and commitments", Proceedings of the 2016 ACM Conference on Designing Interactive Systems, pp. 656-671, 2016.

[10] C. Harrington, S. Erete and A. M. Piper, "Deconstructing Community-Based collaborative design: Towards more equitable participatory design engagements", Proc. ACM Hum.-Comput. Interact, vol. 3, no. CSCW, pp. 1-25, Nov. 2019.

[11] K. Turner, A. Verma and D. Wood, "Intersectional antiracism technology design: Building frameworks to advance justice and equity in complex sociotechnical systems", AGU Fall Meeting Abstracts, vol. 2021, 2021.

[12] S. Costanza-Chock, "Design justice: Towards an intersectional feminist framework for design theory and practice", June 2018.

[13] S. Costanza-Chock, Design justice: Community-led practices to build the worlds we need., The MIT Press, 2020.

[14] P. Lin, K. Abney, and G. A. Bekey, Robot Ethics: The Ethical and Social Implications of Robotics. MIT Press, 2014.

[15] M. Mori, K. F. MacDorman and N. Kageki, "The Uncanny Valley [From the Field]," in IEEE Robotics & Automation Magazine, vol. 19, no. 2, pp. 98-100, June 2012, doi: 10.1109/MRA.2012.2192811.

[16] Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K. (2015). Towards Safe and Trustworthy Social Robots: Ethical Challenges and Practical Issues. In: Tapus, A., André, E., Martin, JC., Ferland, F., Ammi, M. (eds) Social Robotics. ICSR 2015. Lecture Notes in Computer Science(), vol 9388. Springer, Cham. https://doi.org/10.1007/978-3-319-25554-5_58

[17] Kerstin S. Haring, Michael Misha Novitzky, Paul Robinette, Ewart J. de Visser, Alan Wagner, and Tom Williams. 2019. The Dark Side of Human-Robot Interaction: Ethical Considerations and Community Guidelines for the Field of HRI. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 689–690. DOI:https://doi.org/10.1109/HRI.2019.8673184

[18] C. Lacey and C. Caudwell, "Cuteness as a 'Dark Pattern' in Home Robots," 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 2019, pp. 374-381, doi: 10.1109/HRI.2019.8673274.

[19] J. P. Sullins, "Robots, Love, and Sex: The Ethics of Building a Love Machine," in IEEE Transactions on Affective Computing, vol. 3, no. 4, pp. 398-409, Fourth Quarter 2012, doi: 10.1109/T-AFFC.2012.31.

[20] Borenstein, J., Arkin, R. Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being. Sci Eng Ethics 22, 31–46 (2016). https://doi.org/10.1007/s11948-015-9636-2

[21] K. Abney, "Robotics, ethical theory, and metaethics: A guide for the perplexed," Robot ethics: The ethical and social implications of robotics, pp. 35–52, 2012.

[22] P. Lin, K. Abney, and G. A. Bekey, "Introduction to robot ethics," Robot ethics: The ethical and social implications of robotics, 2012.

[23] Rachel Charlotte Smith, Heike Winschiers-Theophilus, Daria Loi, Asnath Paula Kambunga, Marly Muudeni Samuel, and Rogerio de Paula. 2020. Decolonising Participatory Design Practices: Towards Participations Otherwise. In Proceedings of the 16th Participatory Design Conference 2020 - Participation(s) Otherwise - Volume 2 (PDC '20). Association for Computing Machinery, New York, NY, USA, 206–208. https://doi.org/10.1145/3384772.3385172

[24] John Vines, Rachel Clarke, Peter Wright, John McCarthy, and Patrick Olivier. 2013. Configuring participation: on how we involve people in design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). Association for Computing Machinery, New York, NY, USA, 429–438. https://doi.org/10.1145/2470654.2470716

[25] Norina Gasteiger, Ho Seok Ahn, Christopher Lee, Jongyoon Lim, Bruce A. MacDonald, Geon Ha Kim, and Elizabeth Broadbent. 2022. Participatory Design, Development, and Testing of Assistive Health Robots with Older Adults: An International Four-year Project. J. Hum.-Robot Interact. 11, 4, Article 45 (December 2022), 19 pages. https://doi.org/10.1145/3533726

[26] I. Zubrycki, I. Szafarczyk and G. Granosik, "Project Fantom: Co-Designing a Robot for Demonstrating an Epileptic Seizure," 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Nanjing, China, 2018, pp. 1045-1050, doi: 10.1109/ROMAN.2018.8525609.

[27] E. Antonioni et al., "Nothing About Us Without Us: a participatory design for an Inclusive Signing Tiago Robot," 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Napoli, Italy, 2022, pp. 1614-1619, doi: 10.1109/RO-MAN53752.2022.9900538.

[28] Shiri Azenkot, Catherine Feng, and Maya Cakmak. 2016. Enabling building service robots to guide blind people a PD approach. In 2016 11th ACM/IEEE International Conference on HumanRobot Interaction (HRI), IEEE, 3–10.

[29] Simran Bhatia, Elin A. Björling, and Tanya Budhiraja. 2021. Exploring Web-Based VR for Participatory Robot Design. In

Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, 109–112. DOI:https://doi.org/10.1145/3434074.3447139

[30] Theodoros Georgiou, Lynne Baillie, Martin K. Ross, and Frank Broz. 2020. Applying the Participatory Design Workshop Method to Explore how Socially Assistive Robots Could Assist Stroke Survivors. In Companion of the 2020 ACM/IEEE International Conference on HumanRobot Interaction, ACM, New York, NY, USA, 203–205. DOI:https://doi.org/10.1145/3371382.3378232

[31] Catherine Feng, Shiri Azenkot, and Maya Cakmak. 2015. Designing a Robot Guide for Blind People in Indoor Environments. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, ACM, New York, NY, USA, 107–108. DOI:https://doi.org/10.1145/2701973.2702060

[32] K. Spiel, E. Brulé, C. Frauenberger, G. Bailley, and G. Fitzpatrick, 'In the details: the micro-ethics of negotiations and in-situ judgements in participatory design with marginalised children', CoDesign, vol. 16, no. 1, pp. 45–65, 2020. Archive.nordes.orgarchive.nordes.org

[33] J. C. Read, M. Horton, G. Sim, P. Gregory, D. Fitton, and B. Cassidy, 'CHECk: a tool to inform and encourage ethical practice in participatory design with children', in CHI'13 Extended Abstracts on Human Factors in Computing Systems, 2013, pp. 187–192.

[34] M. Steen, 'Upon Opening the Black Box of Participatory Design and Finding It Filled with Ethics', Nordes, no. 4, Art. no. 4, Mar. 2011, Accessed: Nov. 08, 2022. [Online]. Available: https://archive.nordes.org/index.php/n13/article/view/95

[35] K. Spiel, E. Brulé, C. Frauenberger, G. Bailly, and G. Fitzpatrick, 'Micro-ethics for participatory design with marginalised children', in Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1, 2018, pp. 1–12.

[36] A. K. Ostrowski, C. Breazeal and H. W. Park, "Long-Term Co-Design Guidelines: Empowering Older Adults as Co-Designers of Social Robots," 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Vancouver, BC, Canada, 2021, pp. 1165-1172, doi: 10.1109/RO-MAN50785.2021.9515559

[37] Kathleen C Basile, Sharon G Smith, Matthew Breiding, Michele C Black, and Reshma R Mahendra. 2014. Sexual violence surveillance: Uniform definitions and recommended data elements. Version 2.0. (2014).

[38] Douglas Zytko, Nicholas Furlo, Bailey Carlin, and Matthew Archer 2021. Computer-Mediated Consent to Sex: The Context of Tinder. Proc. ACM Human-Computer Interact. 5, CSCW1 (2021), 27. DOI:https://doi.org/10.1145/3449288

[39] William Simon and John H. Gagnon. 1986. Sexual scripts: Permanence and change. Arch. Sex. Behav. 15, 2 (April 1986), 97–120. DOI:https://doi.org/10.1007/BF01542219

[40] Jennifer S Hirsch, Shamus R Khan, Alexander Wamboldt, and Claude A Mellins. 2019. Social dimensions of sexual consent among cisgender heterosexual college students: insights from ethnographic research. J. Adolesc. Heal. 64, 1 (2019), 26–35

[41] Syed Ishtiaque Ahmed, Steven J Jackson, Nova Ahmed, Hasan Shahid Ferdous, Md Rashidujjaman Rifat, A S M Rizvi, Shamir Ahmed, and Rifat Sabbir Mansur. 2014. Protibadi: A platform for fighting sexual harassment in urban Bangladesh. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2695–2704

[42] Nicholas Furlo, Jacob Gleason, Karen Feun, and Douglas Zytko. 2021. Rethinking Dating Apps as Sexual Consent Apps: A New Use Case for AI-Mediated Communication. In Conference Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21), ACM New York, NY, USA, 1–4. DOI:https://doi.org/10.1145/3462204.3481770

[43] Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Gianluca Stringhini, Thamar Solorio, Munmun De Choudhury, and Pamela J. Wisniewski. 2021. A Human-Centered Systematic Literature Review of the Computational Approaches for Online Sexual Risk Detection. Proc. ACM Human-Computer Interact. 5, CSCW2 (October 2021), 1–38. DOI:https://doi.org/10.1145/3479609

[44] World Health Organization. 2013. Responding to intimate partner violence and sexual violence against women: WHO clinical and policy guidelines. World Health Organization

[45] Human Rights Commission. Sexual Assault and the LGBTQ Community - HRC

[46] Kathy Charmaz. 2006. Constructing grounded theory: A practical guide through qualitative analysis. sage

[47] S. R. Khan, J. S. Hirsch, A. Wamboldt, and C. A. Mellins, '"I Didn't Want To Be 'That Girl'": The Social Risks of Labeling, Telling, and Reporting Sexual Assault', Sociological Science, vol. 5, no. 19, pp. 432–460, 2018.

[48] Hee Rin Lee, Selma Šabanović, Wan-Ling Chang, Shinichi Nagata, Jennifer Piatt, Casey Bennett, and David Hakken. 2017. Steps Toward Participatory Design of Social Robots: Mutual Learning with Older Adults with Depression. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, 244–253. DOI:https://doi.org/10.1145/2909824.3020237

[49] Emma J. Rose and Elin A. Björling. 2017. Designing for engagement: Using PD to develop a social robot to measure teen stress. In Proceedings of the 35th ACM International Conference on the Design of Communication, ACM, New York, NY, USA, 1–10.