Letters from the Future: Exploring Ethical Dilemmas in the Design of Social Agents

Michal Luria
mluria@cs.cmu.edu
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA, USA

Stuart Candy
scandy@usc.edu
Berggruen Institute
University of Southern California
Los Angeles, CA, USA

ABSTRACT

We present a case-study of using Ethnographic Experiential Futures (EXF) to surface underlying divergences, tensions and dilemmas implicit in views of the futures of "social agents" among professional researchers familiar with the state of the art. Based on expert interviews, we designed three "letters from the future," research probes that were mailed to 15 participants working in the field, to encounter and respond to. We lay out the elements and design choices that shaped these probes, present our remote and asynchronous study design, and discuss lessons learned about the use of EXF. We find that this form of hybrid design/futures intervention has the potential to provide professional communities with opportunities to grapple with potential ethical dilemmas early on. However, the knowledge and tools for doing so are still in the making. Our contribution is a step towards advancing the potential benefits of experiential futures for technology designers and researchers.

CCS CONCEPTS

• **Human-centered computing** → **User centered design**; *Interaction design process and methods*; *Scenario-based design.*

KEYWORDS

ethnographic experiential futures, social agents, robots, artificial intelligence, speculative design, design fiction, probes, foresight, ethics

ACM Reference Format:
Michal Luria and Stuart Candy. 2022. Letters from the Future: Exploring Ethical Dilemmas in the Design of Social Agents. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3491102.3517536

1 INTRODUCTION

Emerging technologies frequently raise social and ethical challenges, many of which are not well anticipated. Examples include Microsoft's Tay chatbot, which had to be removed from Twitter after less than 24 hours due to its racist and misogynistic comments [9]; AI algorithms that have reinforced societal biases and discrimination [41]; smartphones that have caused long-term harm to children and teenagers [43] and conversational agents that have exposed people's private information [44]. While some tech companies have made significant commitments and efforts to tackle challenges brought about by their new offerings (e.g., by defining ethical principles and creating designated ethics research teams [1, 2]), this may not be enough.

Some researchers have suggested that the engagement with ethical implications of technologies' possible consequences should occur very early in the design and creation process, for example, using policymaking [28] or games as part of the ideation process [4]. Yet it remains a challenge to anticipate the impact of technological developments beyond their immediate goals and outcomes, especially with emerging technologies that do not yet exist [27]—design and development teams tend to center around their own positive visions and intentions for the future, leaving underlying ethical concerns and risk contingencies invisible and undiscussed [30].

In this work we test out a design methodology that directly addresses the need to surface ethical concerns and potential second- and third-degree impacts of emerging technologies early on, when the design vision is still being formulated within a research or design community. To do so, we present a case study of Ethnographic Experiential Futures (EXF) [14]. This approach sets out to create tangible experiences of possible futures, to enable constituencies of would-be creators or users to engage more fully with an array of social and cultural implications of those futures, and as a result reflect more deeply upon them. We also extend EXF practice by suggesting a remote and asynchronous approach that can present a diverse palette of futures to audiences who are not co-located.

In the tradition of hybrid design/futures methods that have been taken up by designers to explore prospective outcomes more multidimensionally, including Design Fiction [7, 38], Speculative Design [18] and Experiential Futures [11, 39], we attempt to look into selected social and cultural aspects of some futures currently being incubated among certain professional communities—particularly Human-Agent Interaction (HAI) and Human-Robot Interaction (HRI) researchers—with the hope of encouraging and enabling deeper reflection around current (as well as possible alternative) visions, assumptions, and concerns. We interviewed three experts in the field and created a trio of "letter probes," each designed to concretize and bring to life a different "what if?" for how social agents might develop and be used in years to come. We then recruited 15 researchers to encounter these embodied narratives of hypothetical social agents. We report on the qualitative responses to these narratives, and on the themes of ethical concern surfaced in the process.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3517536

Lastly, we discuss our findings in this deployment of EXF, considering both successes and areas where it could be improved on and used more effectively in future work. This case study is a contribution towards better understanding the value that the design and deployment of "artifacts from the future," particularly via the Ethnographic Experiential Futures process, can bring to professional communities of technology researchers and designers.

2 RELATED WORK

In the field of HCI, design methods such as storytelling, prototyping and probe creation have been established as powerful tools to explore, examine and critique current and future technologies. While some of these approaches are intended to be part of design processes as a form of discovery (e.g., Cultural Probes [20]), we will focus on approaches that, like EXF, emphasize storytelling, provocation and immersion to actively encourage discussion and reflection through a finalized artifact.

In this community, such methods have sometimes been referred to collectively as *Discursive Design* [40] or *Design Futuring* [24]; moves that leverage design to explore and reflect on possible future technologies. The work presented in this paper is also situated within this space—being a case study of EXF as a probing and storytelling effort to understanding more about the design space and potential implications of future social agents. We set out to examine some of the blindspots and ethical considerations at play within this research domain, in contrast to more traditional forms of ethical discussion conducted in the fields of Human-Agent and Human-Robot Interaction [21, 33].

Below, we lay out some prominent methods under the umbrella term of "Discursive Design," and compare and contrast the approach taken here.

2.1 Fiction and Speculation

Among the popular Discursive Design approaches is Speculative Design, which often seeks to enable new conversations about potential future developments through provocative probes [17, 18]. It has been used in the past to explore, for example, future social robots [3]. Speculative Design uses hypothetical design artifacts' to critique existing technology and enable discussion on opportunities for change; via propositions that may or may not strictly lie within the "possible", let alone the "probable" zone of the landscape of future scenarios (Candy in [18]). The present study, in contrast, deals with scenarios deemed to be significant and worthy of closer attention by individuals steeped in the subject matter.

One way of doing EXF is to start and ground the research process in the imaginaries of specific informants or participants, incorporated here through in-depth interview. This differs from most Speculative Design, which frequently focuses on empowering the designer or researcher herself to envision and propose futures for discussion. Some critiques of Speculative Design have emphasized that much of the work conducted within that tradition is showcased in gallery settings, which may not necessarily include opportunities to stage or shape the discussion that is intended or claimed to occur [19, 35, 42].

Design Fiction (although not necessarily definitionally distinct from Speculative Design) is another banner under which designers in recent years have often explored future technology. It makes use of diegetic (in-world) prototypes to "tell worlds rather than stories" [7, 38], which is similar to the goal here. Design Fiction approaches may emphasize the coherence of an underlying narrative about a future world, and those seeking to practice it sometimes use written fiction to do so. Similarly, Pastiche Scenarios make use of fictional storytelling for future discovery and discussion purposes, and what distinguishes this approach is the effort to tell richer stories by leveraging well-known fictional characters (such as Bridget Jones or Ebenezer Scrooge) instead of inventing new ones [8]. Schulte et al. create a more open-ended approach to storytelling as a form of inquiry, a hybrid between Design Fiction and Cultural Probes [34], and Pierce developed "Design-led Inquiry," a form of Research through Design [45] that imagines alternative uses of existing technology while highlighting blindspots [32].

The stories told in the approaches above, in most cases, are the creation of designers, researchers or artists, although sometimes, as in "Futuristic Autobiographies", the stories are created directly by participants [15]. This particular deployment of EXF begins with expert interviews to gain an understanding of some of the currently-imagined trajectories of the technology of interest (social agents), according to relatively "well-informed" imaginer-informants, which are then transformed into high-resolution design probes purportedly "from" the futures in question.

In many Discursive Design projects the focus is on the future probe or prototype itself—a technological proposition imagined without a fuller social context that would presumably exist around it in the future posited. Such artifacts are frequently placed by their designers and met by their audiences in a gallery or museum space—or, they might be handed to people to consider and discuss in a relatively scripted setting, such as a workshop. This study differs in that it tried to engage people in considering probable (according to our interviewees) social agent futures through probes that also embedded details of context and environment where the social agent might live within the hypothesized future, and for participants in the study, the encounters with these futures were staged in a perhaps more organic and intimate domestic setting, through letters sent to their home addresses. On our part, this required considering and including diegetic/in-world reasons why receiving a physical paper letter would make sense in each case.

2.2 Embodiment and Immersion

Some prior work has focused on the degree of immersion of a design probe or intervention as a key attribute. User Enactments, for instance, use prototyping and Wizard-of-Oz techniques to place participants in a future context, and thereby seek to have them reflect on future technologies within their intended site of interaction [16, 25, 31]. Other researchers have used immersive theater to involve people in considering the risks of a given technology, and to encourage ethical discussions about their possible implications moving forward [37]. In this work we attempt to balance immersion and provocative storytelling by using some elements inspired by immersion (albeit practically hemmed in by Covid-19 pandemic restrictions), and other elements of provocative storytelling to bring alternative future worlds to life, with the goal of

promoting reflection on not only the technologies themselves, but also on their systemic contexts and consequences.

2.3 Ethnographic Experiential Futures

Ethnographic Experiential Futures (EXF) is a direct descendant of two existing traditions of research and scholarship: Experiential Futures (XF), a wide-ranging, transmedia design space that incorporates not only future artifacts (objects), but also immersive scenarios (situations) that physically place participants in experiences that aim to portray possible future worlds [11], and Ethnographic Futures Research (EFR), a method that maps out desired, feared and expected views of the future held by individuals or by a community [39]. EXF directly combines the two. It invites researchers to consider *both* how they might surface images of the future within a field (as in EFR) *and* how these possibilities might be manifested as tangible experiences for others to engage in (as in XF) [14]. EXF thus not only elicits statements about particular futures, but seeks in a way to translate them into embodied and rich *experiences*, in hopes of promoting reflection and change [11, 13]. For example, Kornet interviewed environmental activists about the impacts of nearby petrochemical industry facilities on health in their communities [23], and Jenkins and colleagues explored possible futures of supermarkets and the core values that accompany them [22]. In this work, we put the EXF framework to use in creating a novel set of future encounters for participants, under remote and asynchronous conditions, and extending the practice by implementing a more structured form of reflection and qualitative data analysis than prior work has done.

3 EXF AS RESEARCH APPROACH

This study uses EXF to elicit reflections about a set of posited future scenarios and underlying ethical tensions within a particular professional community. The research combines design and futures methods, both of which require making design choices and judgements in the process. While not seeking to produce replicable results in a social science sense, this research provides an instructive example of using EXF to design discursive probes and evaluate their effects as a form of targeted insight-prompter or conversation-starter. We have attempted to fully report the design and evaluation process here for other designers and researchers to learn from, whether that be by applying similar approaches in future research efforts or advancing the method by learning from our mistakes (See Section 9).

EXF invites engagement with diverse participants through what are intended to feel like tangible experiences of possible futures. It builds on participatory foresight and design, in hopes of making a more diverse array of scenarios available for consideration.

As noted above, EXF builds in part on Ethnographic Futures Research (EFR), the purpose of which is to map images of the future held by a particular community or individuals [39]. According to anthropologist Robert Textor, EFR's principal developer, *"Just as the cultural anthropologist conventionally uses ethnography to study an extant culture, so the cultural futures researcher uses EFR to elicit from members of an extant social group their images and preferences with respect to possible or probable future cultures for their social group."* (Textor, 1980, p. 10) [39]. EXF, in contrast to EFR, does not stop at mapping a community's vision of the future, but also processes and translates it into experiential probes for the community to interact with, and potentially offer responses and reflection.

The EXF method includes four essential parts in its workflow, and one optional part (see Figure 1):

- **Map:** Inquire into people's images of the future in relation to a particular topic—for instance, future scenarios that they expect, desire, or fear.
- **Multiply (optional):** Generate alternative images of the future that challenge or extend existing thinking.
- **Mediate:** Translate discussions and ideas about the future into tangible probes or immersive experiences.
- **Mount:** Stage these experiences to engage the interviewees, or others, in deeper discussion about the futures in question.
- **Map:** Investigate and record responses.

The rest of the paper follows this structure. For each stage, we describe the process and decisions made towards our final findings and discussion.

4 MAP: SURFACING POSSIBLE FUTURES

OR: FORMATIVE EXPERT INTERVIEWS AS A BASIS FOR EXF

In the initial *Map* stage, the goal was to identify a series of specific images of the future of social agents through semi-structured interviews with experts in the HRI/HAI field. The aim was to inquire more deeply into particular future design directions that people in the community are thinking about and striving towards, and that the experts consulted therefore regarded as significant.

We worked with three interviewee-informants, each bringing a different perspective on how robots and agents may develop in years to come. One was an academic researcher in a leading robotics institution; another a designer who showcased their robots in exhibitions around the world; and the third an artist known for their work on immersive AI and social agents. For these purposes quality was prioritized over quantity—and the three were carefully selected for their expertise as well as their divergent perspectives on the topic of social agents and robots.

Interviews were conducted remotely via video link, each conversation lasting 30 to 45 minutes. The questions explored interviewees' images of the future using a classic trio of lenses from EFR: desired, undesired, and expected futures. For example, interviewees were asked to imagine a future that they feared, or one that is 90% desired [39]. The goal of these interviews was to seek out "key uncertainties" or questions in the domain of social agent design, and on that basis to identify promising scenaric ideas or premises'—"what ifs" that could be turned into tangible probes as the next step in the EXF process [14].

5 MEDIATE: FROM CONVERSATION TO DESIGN FORM

OR: TRANSLATING EXPERT INTERVIEWS INTO PROBE IDEAS

The goal of the next stage of the process was to render particular ideas and concerns about the future, surfaced during the expert interviews, into a tangible form that others could experience. As each interview spanned many questions and topics, they did not necessarily point in each case to a single, specific vision for future agents. To help address this challenge in proceeding towards a concrete provocative probe, Candy and Kornet suggested identification

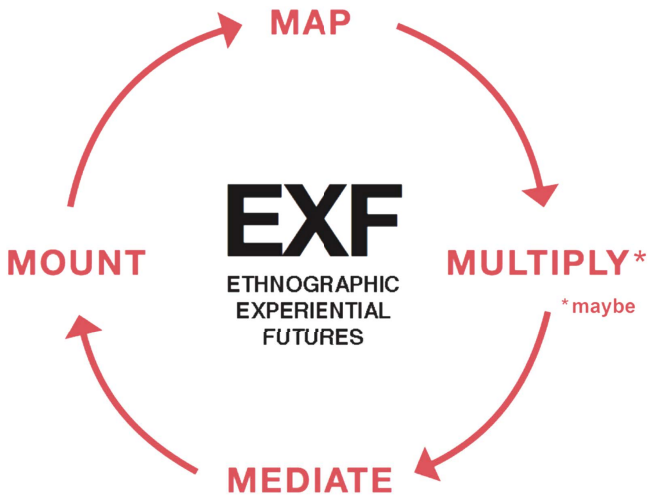

Figure 1: The Ethnographic Experiential Futures method workflow: Map, Multiply (as needed), Mediate, Mount and return [14]. ***Figure courtesy of Candy and Kornet (2019).***

of "scenaric ideas" within interviews [14]—cues from the conversation content that might serve as a starting point for more tangible expressions or instantiations of the ideas in question.

Setting out to highlight ethical considerations worthy of closer scrutiny within the research community, we focused on *desired* visions of the future as described by our interviewees. We transcribed and analyzed the interviews in search of scenaric ideas—characteristics of a future social agents, or something specific that an agent could do. Then, equipped with a list of such characteristics, we brainstormed several ways in which each idea might be translated into a tangible probe. For each possible implementation, some scenaric ideas were included and others left out; not all characteristics would necessarily fit within a single, coherent narrative concept, although we attempted to include as many of the given expert's ideas as possible. For instance, the scenaric ideas selected from one of the interviews included: (1) an agent that helps with learning and education, (2) an agent that is tailored and personalized to each individual, and (3) an agent that allows tracking of personal achievements.

While the "mediation" process—turning a perhaps vague idea about a future into something concrete enough to be experienced—inevitably requires a measure of creative interpretation and input from the researcher, we sought to focus on those ideas that best characterized each interviewee's thinking, and that were already offered or expressed by them in relatively concrete terms (as in the example above).

The goal was to mediate each of the three conversations into a corresponding future artifact or probe. Thus, the process resulted in three social agents, one per interview. We ideated within the team to expand on the initial scenaric ideas, concluding with the following descriptions for each probe: (1) the agent's main goal or function (e.g., an at-home tutoring agent for middle school children); (2) its behavioral qualities (e.g., empathetic, encouraging, etc); (3) other clues about the world in which it may exist (e.g., algorithms are integrated in the learning process, teachers and principals can make use of data to improve teaching). This outcome allowed us to advance to the next stage of the process—to translate these ideas about future agents into physical, experiential probes.

6 MOUNT: AN EXPERIENCE OF FUTURES FOR OTHERS TO INTERACT WITH
OR: DESIGNING THE PHYSICAL PROBE

Translating a list of characteristics into a final probe is a critical step, as each design decision necessarily shapes the story of the future that is being told. (We discuss some of these aspects further in Section 6.1). We aimed to create an engagingly tangible, and even mildly immersive, encounter with each future agent, which is why here the Mediate and Mount stages of EXF are interleaved—alongside ideating *what* was the thing that participants would experience, it was just as important to consider *how* they would experience it. As the intended outcome was a reflection of ethical concerns of future social agent designs, our target audience were emerging or trainee researchers and designers in the HRI and HAI fields. This was taken into consideration when designing the probes described below.

We considered four key aspects of designing an experience for participants: technology, storytelling, location and timing. Decisions around each aspect highlighted potential pros and cons, and helped determine the physical probes to be created.

(1) Technology—What is the desired balance between the effort invested in creating high-fidelity, technologically complex probes, and the provided value? The final probes focused on evoking fragments of *the world and use context in which future agents may exist*, rather than developing agent prototypes themselves. In earlier iterations of this stage we first attempted to develop working agent prototypes that would simulate an interaction "from the future," as if the user were to meet the hypothesized social agent today. However, the complex interfaces introduce many technological and logistical challenges, which in turn, we came to see, would limit their storytelling potential. That is, attempts to plausibly prototype the interactions themselves would tend to pull focus on the nuances of interaction and prototype design (e.g., the agents' appearance, voice, language, and so on), which was not the intention of this study. Rather, we wanted to encourage participants to think about the future context in which these artifacts might sit, including the very *goals* of their existence, and their potential social and cultural meanings and consequences.

In addition, we wanted to take advantage of the affordances of engagement with respondents through physical production, rather than something more performative (perhaps labour intensive) or screen-reliant (perhaps abstract). For that reason we elected to use perhaps the simplest tangible medium available; paper. This choice enabled design output to a very high fidelity, creating a believable, engaging, high-resolution and even somewhat immersive experience of a thing just as it might exist in the posited future (not to be confused with quick, low-fidelity "paper prototypes" [36]), while expending minimal resources on technological development, and retaining the ability to tell stories about this series of future social agents in their broader context.

(2) Storytelling—How well does the probe tell the story of an alternative world, with its social, cultural or political

complexities? Having determined that we would use probes using paper and print, we next needed to generate an array of options for instantiating the expert interviewees' future agent stories in that medium. For this purpose, we used the Futures Wheel, a method for ideating beyond the obvious or intended results of a particular change (say, a technological development) and prompting the researcher/designer to examine broader second- and third-order societal impacts [5]. Alongside this, we generated a generic list of paper objects that people use in their daily life, any of which might in principle become an encounter with a future artifact—reports, notes, letters, notebooks, books, notices, and so on. The two brainstorms resulted in a two lists; one, a set of possible societal consequences for each type of social agent (aligned with each of the three expert perspectives), and the other, a set of common paper artifacts that could be the "tip of the iceberg" for a story of social change we wanted to demonstrate [11]. (The two lists combined or multiplied created a rich set of candidate artifacts for the project. A similarly structured approach can be seen in the design/futures card game *The Thing From The Future*, from which this stage of the process drew some inspiration [12].

(3) Location—Where should the study take place? Should it be a remote or a co-located experience? Due to Covid-19 safety restrictions there were clear limits on how research participants could encounter physical artifacts. The target audience of HRI/HAI researchers and designers was quite specific, so a remote setup was more appropriate (in contrast to, say, staging "posters from the future" in a downtown or campus locale where a larger and more diverse audience might see them). One option considered for creating experiential scenarios that could be experienced remotely was to create a digital experience, such as an online document, e-mail, or calendar reminder. At one stage we also began prototyping a remote "live" interaction with an online chatbot from the future. But given that most of our intended participants' interactions during Covid-19 were already remote, there was a concern that they might be desensitized to additional digital interaction, which could undermine the goal of creating evocative, engaging probes. We wanted to have them interact with the physical artifact wherever they happened to be, and so the solution that naturally emerged was to send artifact probes by mail. This would also make the material relatively easy to distribute anywhere.

(4) Timing—Should the experience be designed as synchronous or asynchronous? When synchronized, researchers can take an active part by controlling the technology (in our case, agent) behind the scenes while making it seem autonomous (also known as Wizard-of-Oz control) [26, 33]. The downside of this option is that it would have required researchers to be "always on" to play the role of the agent (as in the study conducted by McCarthy [29]). A possible alternative execution would have been setting a specific "interaction time" with participants, but this would have sacrificed the opportunity to let people "discover" and organically interact with the designed probes in their own time.

The design decision to send mail-in probes meant that the timing of participants' future encounters was unpredictable; so the probes and overall study had to work for asynchronous conditions. The advantage of this was that we could rely on the fact that participants would view the letter when they had availability to engage (as opposed to say, push notifications that might arrive when they were otherwise occupied).

6.1 Final Probes

Mailed letters are relatively straightforward to design while still able to embed a compelling story, not only about a technological development, but also about the world in which it exists; the second- and third-order aspects of particular scenario trajectories. This led us to focus more on storytelling than on the aspects of technical development that working prototypes would have required, and enabled inexpensive but high fidelity probes with the look and feel of real letters.

In each of the letters we attempted to build a bridge between familiar and recognizable, on one hand [13], and potentially novel aspects of social agents in years to come, on the other. While the expert interviews had primarily focused on positive visions, we also sought to embed some elements of ethical, social and cultural complexity or dilemma for each context (a sort of twist on the "Multiply" step of EXF), which resulted in somewhat debatable or polyvalent futures instanced in the three letter probes. This was part of the effort to evoke deeper consideration, both of desired futures and of possible consequences attendant on them.

Sending letters also allowed some consistency across the three experiences. Each of the three futures, inspired respectively by the interviewees distinct visions, was instanced or mediated through a letter that someone might receive in that future. We tried to imagine the action of sending and receiving artifacts as being part of the story, so that the whole experience would ideally unfold or "behave" as it might if this future actually occurred. For example, to make the letters more believable, they were personalized by inserting the family name of the recipient (e.g., statistics on the education of the participants' child, Briar Joni *Last Name of Participant*).

6.1.1 Educational Agent: A Letter from Gates Middle School. The first letter expresses a vision of future agents in education, where personal agents might collaborate with children when learning (Figure 2). In this vision, personalized social agents are designed to assist individuals in learning, both at home and in schools when teachers are unavailable. These social agents would adjust to special needs, learning styles, and individual progress and interests.

The letter is addressed to the parent of a child who studies with the assistance of a social agent system called Tutori. It offers an automated, quarterly snapshot of the child's progress, essentially a next-generation school "report card," but it is accompanied by a letter from the school principal to the child's parent/guardian, recounting the concerning story of the child's attempt to "cheat" the agent's algorithm to improve their performance.

Thus, the probe not only portrays a general vision of an educational social agent, but also points towards additional questions that may be implied by this possibility: What might the agent track in education? How would it do so (what behaviors does it notice and report)? What is the company that provides these social agents, and what is the nature of their collaboration with the educational institution? What are the boundaries and limitations of AI in this context? Can a human outsmart or game the system? How does the system respond when a child attempts to do this? Designing the probe involved thinking through and making decisions about

Figure 2: A letter that summarizes the achievements and overall evaluation of the participants' child at school. The letter combines a personal note from the school principal regarding a disciplinary incident that occurred in school, along with the future alternative to a "report card," generated automatically by an educational agent.

many such questions, and was intended to encourage participants in turn to consider them and their ethical implications.

6.1.2 Values Agent: A Letter from Veritas. This letter expresses a vision in which social agents are personalized to support users in living according to their values. They track "alignment" between behavior and professed values, evaluate and suggest possible life changes that a user might make, and are designed to motivate the user to be the best version of themselves—as defined by them (Figure 3). The letter received by participants depicts an annual achievement overview from Veritas, the company that runs these "value-oriented" social agents. The letter also comes with an "update choices" questionnaire for the recipient to fill out and return; a kind of interaction-by-post that may be familiar to many of us in the present day.

Some of the questions this letter might raise include: How much should a social agent know about your personal values, and how

Figure 3: A letter from a personalized "values agent." The letter summarizes successes and failures in aligning with users' personal values. The letter also includes a form for participants to fill out in order to update their agent on current ideologies, self-concept and priorities.

much should it ask? Could or should it leverage information from other sources to assist an individual with their mission? Who runs this kind of company and what is their agenda? How proactive should an agent be in encouraging a user to pursue their ideological goals? What happens when there is misalignment between one's values and behaviors? What are the ideological attitudes or practical limits, if any, of the company's willingness to support a user in these ways? Similarly to the previous probe, the design choices made imply, for the sake of argument, some possible answers to these questions, and in doing so aim to invite participants to consider their own answers.

6.1.3 Privacy Agent: A Letter from an Underground Organization. This letter expresses a future in which agents behaves as largely autonomous "creatures" that help users resist data collection and surveillance by big tech and governments. These agents are not controllable—not by companies, governments or by individuals. Rather, they are designed to develop in their own way. However, their fundamental goal is to help and serve people (Figure 4).

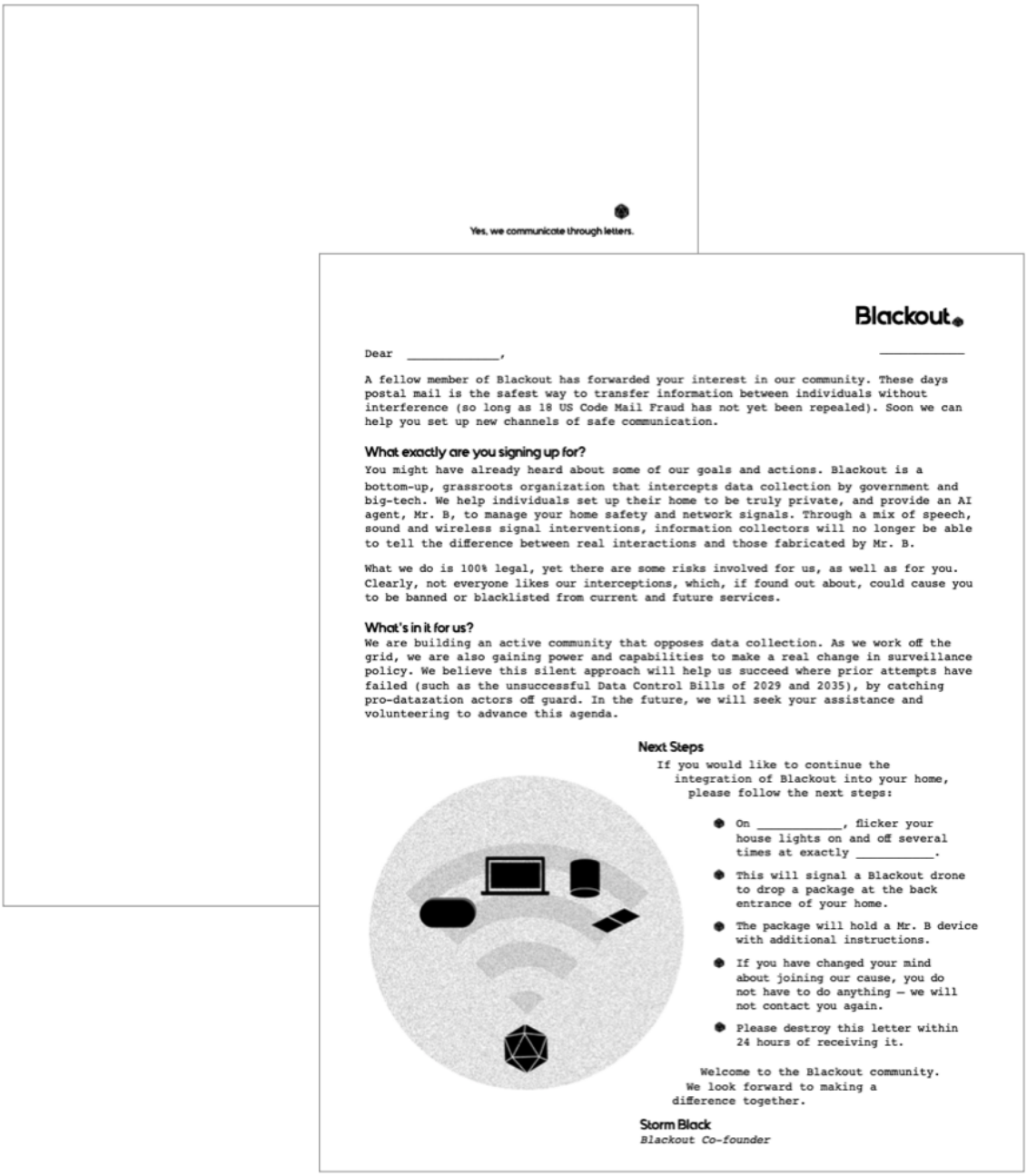

Figure 4: A letter from an "underground" social organization sent to participants. The letter invites the receiver to join the initiative, and guides them about how they would implement a privacy agent in their home. The letter is sent through mail to avoid surveillance and data tracing in this future world.

The letter comes from a nonprofit organization that sets out to distribute these particular social agent "creatures" to counter the status quo in this future; what are (in its view) invasive data collection practices. This agent therefore functions like a type of pet; just as a guard dog might protect your home physically, this one would protect your home on the digital front. It does so using signals and data generation to interfere with surveillance. Some of the questions this probe may raise include: What kind of social agents might be designed "for society"? How would they develop independently? How might they compare to agents created by

companies? How much control could people have, and what kinds of backlash or politicized technology movement might be produced by the continuation and expansion of ubiquitous surveillance? Can one really count on a social agent to solely serve one's needs with no ulterior motive?

7 MAP: RECORDING EXPERIENCES AND REFLECTIONS

OR: STUDY DESIGN AND EXECUTION

This study used the EXF method to collect reflections and investigate ethical concerns that surfaced in the process in a relatively structured way, going beyond the presentation of hypothetical design artifacts to audiences in an exhibition or gallery context. Instead, we set out to capture participant responses more fully by using an open-ended survey (sent electronically) soon after each future letter. Since each letter conveyed one future story, each was followed by a separate link to collect participant responses.

7.1 Participants

Fifteen participants took part in the study (M = 30; 11 identified as female and 4 as male). All were Human-Robot Interaction or Human-Agent Interaction researchers across the US; eight Ph.D. students, two postdoctoral researchers, two researchers and three professors. All declared that they considered interactions with agents and robots as part of their professional work; this was why they were recruited as respondents.

7.2 Questionnaire Design

An online questionnaire captured each participant's reflections on the ethics of future agent design after they had received and interacted with each probe. It included general open-ended questions to get at sentiment (e.g., "How did you feel after reading the letter? Why did you feel this way?"), several 5-point Likert scale statements ("This letter manifests a future that is likely"), and questions about the ethics of designing the portrayed agent ("Seen as a thought experiment, what ethical issues does this experiential scenario raise for the field?") Some questions were adapted from Candy [11] and Kornet [23] and others were formulated for the purpose of the study. The questionnaire is attached as supplementary material.

7.3 Procedure

Participants were recruited via word of mouth through mailing lists and communication channels (e.g., Slack) used by the HAI/HRI research community. Interested individuals were asked to fill out a signup form that included informed consent, basic information, several demographic questions, and three sentiment-towards-privacy questions adapted from [10]. The letter probes were created in advance of recruitment, except for minor personalized elements in each case. A week after sign-up, the first batch of letters was dated, personalized, and mailed to participants.

The three letter probes were sent to participants in random sequence, with an email follow-up two or three days later, indicating that they should have received or would soon receive a mailed letter. They were asked to read it and reflect on their response through the online questionnaire as soon as possible afterwards. Once a participant's response was collected, the next letter was personalized, dated, printed and sent out.

Participants answered 13 questions via the questionnaire for each of the three letters (39 questions in total). The questionnaires were filled out using a consistent participant number provided in an email, to ensure anonymity in the questionnaire analysis. A separate sheet was used to cross-reference participant numbers answered each questionnaire, and their names for the next personalized letter. This allowed the researchers to analyze the questionnaires while maintaining responders' anonymity. The procedure and data collection protocol were approved by the university's institutional review board (IRB).

7.4 Analysis

Likert-scale questions were analyzed using basic inferential statistics such as median, mean and mode. Open-ended questions were analyzed using affinity diagramming, a common design research practice to reveal patterns in a qualitative dataset [6]—each qualitative response was grouped with similar responses, revealing recurrent themes across multiple participants and interactions.

7.5 Study Results

We report on three underlying ethical questions on the design of future social agents that we have identified. Overall, participants agreed on the *questions* that need to be considered in the community, although they did not agree on the answers to them. The findings are HRI and HAI specific, but they may serve as a case study for more widely applicable considerations emerging from the methodology and research approach.

7.5.1 Are Data Privacy and Algorithm Ethics Relevant to HRI Research? It was not always clear to participants whether each probe referred to a social agent, a robot, or simply an artificial intelligence algorithm. Some were *"not sure [they] see the engagement with robotics"* [P08], or *"the connection to HRI"* [P09], while others thought these topics were *"highly related to HRI"* [P10]. P12 said that *"when reading the letter, HRI aspects didn't occur to [them at first], but now [they] think it is highly relevant."* Part of this confusion might have been due to the probes' paper medium, which apparently led some participants to assume that the "agent" was just software or an algorithm. That said, we believe the polarized responses on whether the probes are or are not relevant to the field of HRI may point to a larger concern: the topics of ethics, data privacy and algorithm fairness are commonly seen as falling outside the sphere of responsibility of HRI researchers, and this could be why some participants considered their discussion surprising or irrelevant for the design of robots and agents.

We conclude that HRI and HAI researchers should see themselves as active participants in resolving the challenges that algorithms raise, as these will likely be implemented alongside physical aspects of social robots. EXF and other efforts can contribute to this goal. For example, P11 said that: *"This [study] encourages [them] to consider how when we talk about ethics of human-agent interaction. We also need to be thinking about what is happening during periods of non-interaction [e.g., data storage and privacy], and to actually focus on those instances in our research"*.

7.5.2 Should Agents Make Social Assessments of People? There were differing views among participants as to whether agents should make social assessments of people, and how the collection of data by agents and robots compares to other widespread technology (e.g., social media). P12 argued that *"collecting data on social interactions at home [for example] seems a bit alarming at first, but from a perspective of personalization to improve the quality of human-robot interaction it seems essential"* and that *"this type of AI-supported, personalized systems can be hugely beneficial."* Other participants presented contradicting arguments, for instance, that their research community should always strive to *"develop HRI systems using minimal data"* [P03].

Here we see an underlying tension around the issue of whether agents should make social assessments of their users or not, and how much data is acceptable to collect for the purpose of interaction with social agents. More discussion among the community is needed to define where the boundaries that matter ought to be drawn.

7.5.3 Should Policymakers Be Involved? Participants also took two contrasting stances on how topics of ethics should be evaluated within the community: internal assessment and external audit.

The *internal assessment* approach argued that data can be collected and used for AI, but with *"more transparency in the algorithms"* [P14], *"communication of data sharing policy and procedures"* [P11] and *"constant discussions of ethics"* within the field [P10].

The second, *external audit* approach argued that the solution should be external—to *"increase regulations limiting what people can build and what data they are allowed to collect and analyze"* [P13]. Participants who took this approach advocated for the involvement of policymakers in all stages of robot, agent and AI development. As P11 put it, *"I see [an ideal] future as equally tech-driven and policy-driven"*. Neither approach is fully implemented today. This suggests that there may be space for HRI and HAI researchers to take initiative to make both internal and external examinations of ethics a significant part of their research process.

8 FINDINGS ABOUT THE EXF METHOD

In this section we consider participant responses that inform **our use of EXF**: We discuss how each element may have contributed to or detracted from the overarching experience of the alternative futures, as well as associated ethical considerations.

8.1 Prototype Immersiveness

The probes were deliberately low-cost, 2D, and did not require technical development, yet they still successfully engaged participants in a range of futures they aimed to portray. The seamlessness with which the letters entered people's everyday lives contributed to bringing participants into the alternative worlds that they represented. Participants were *"surprised by how realistic it felt"* [P08], *"felt intrigued to receive the letter, as it was fun to envision a different future, like a game!"* or even felt it was like *"being part of a science fiction novel"* [P07].

As fully immersive face-to-face experiences were not possible due to Covid-19 restrictions (to say nothing of cost as a factor), we turned to other means of engagement. It appears that the use of high-fidelity letter designs, the US Postal Service as a platform, and the implicit logics for why the content was presented as a mailed letter in each case, all contributed to a sense of realism or verisimilitude in the stories told. One participant even said that *"At least two of the three [letters] had [them] completely fooled for several minutes before [they] realized what was going on"* [P09].

8.2 Narrative Distance from Person

Each of the letters told a story with a different protagonist. The Veritas letter told the story of a "future you" and your use of a value-oriented agent. The Gates letter told the story of your future child's education with the assistance of an agent. The Blackout letter told the story of an organization that promotes privacy within a community. In this sense the letters each embed different relationships, and incorporate different degrees of "distancing" between the hypothetical technology and the individual recipient. In the first letter, participants were asked to imagine *themselves* in the future. While this design choice was helpful in telling the story, participant responses suggested that it may have been too difficult for them to imagine themselves in a future scenario, especially when they had to reason with a choice that they did not think they would make today. P02 commented: *"I realized this was a service I had signed up for. I was a little confused about why I had signed up"*. Similarly, P10 highlighted the unknowns about a future self: *"I'm not sure whether future me did give any consents to what was tracked by the AI agent"*.

In contrast, the Gates school prototype presented a story about the participant's (hypothetical) child and the educational agent implemented in their school. Given that implementing the agent was not the participant's choice in this scenario, but rather, was given as part of the education system portrayed, it seemed to be easier for participants to accept. In a way, they appeared to accept the putative reality of the future situation more readily. For example, P15 said that they *"feel a bit sad for [their] imaginary child because it seems like they are only known to the school through quantifiable interactions with software"*, and P01 *"felt proud of Briar Joni [their child], who in the letter was being blamed for 'cheating' and manipulating the intelligent system."*

In other words, participants seemed to be more accepting of the presented narrative when the prototype discussed a protagonist that was not them, but instead someone close to them (e.g., their future child)—even if the participant did not actually have a child today. P01 said that *"Imagining myself as a parent getting this complaint from my child for not complying with a biased system made me care more about addressing this root issue [technosolutionism] now and not later [when I have kids]."*

In contrast, it seemed more difficult for participants to accept a future framed as being due to their own supposed choices within the presented future (as the Veritas letter suggests). Such a case seemed to divert some focus away from the issues themselves, and towards discussion of why they would have made the choices that put them in this future situation in the first place, or not. As a result they were less likely to reflect to the situation and its implications.

8.3 Narrative Distance from the Present

Another element that impacted perceptions and reflections on each narrative related to the time horizon and perceived temporal distance of events depicted. The letters were dated to three different years: 2031 (Gates—10 years out), 2034 (Veritas—13 years out) and

2039 (Blackout—18 years out). These dates were selected based on the story told. While the experts had been asked to describe their visions for 10 years out, the variable dates were selected or inferred by the authors based on the extent of specific changes described.

It consistently seemed easier for participants to consider the ethical questions and considerations of a future agent design that was closer to the present, both in terms of the actual date, and the likelihood of the scenario. For example, participants said that the topic raised in the closest, Gates letter *"needs urgent attention"* [P13], as it provides *"exposure of the possible and plausible negative effects that can come with these systems"* [P04].

The farthest narrative on the timeline was portrayed in the Blackout letter, which was also seemingly more "distant" in terms of likelihood—the story diverged from what today may be perceived as the likely trajectory of future technology, and rather presented a complex story of data collection by authorities and an opposing social movement. Participants found this narrative *"too unrealistic"* [P13], and had difficulty imagining ethical concerns of a future with an alternative socio-political reality. As P14 noted, the letter encouraged one to *"think more about the group of people behind the agent than the AI agent itself. [They were] more concerned with the political situation than with the technology used to address it."*

8.4 Response Collection Method

This study collected responses from participants using online surveys (Qualtrics), which participants received soon after each distribution of a future letter in the mail. The survey offered several advantages as an instrument of data collection here: it allowed the study to be both remote and asynchronous—participants received the letters at their homes, and when they were ready, logged reflections through the online survey available via email; the step-by-step process of receiving and responding to letters allowed participants to reflect on each individually; and it was quite easy to include additional respondents, since the probes and surveys were already designed whether few or many took part (as opposed to qualitative interviews, for example, where each inevitably adds time to the collection effort).

The process in this instance contrasts with many prior design futures projects that have used exhibitions to showcase their probes. Here a more structured approach allowed *every* observer to share their reflections, especially valuable since participants were carefully recruited for their career orientation to professional HRI/HAI research. A survey is also more controlled in the order of exposure and response, and in scaffolding reflections around different elements of each scenario.

However, the approach also had some key disadvantages. It did not allow back-and-forth conversation between participants and researchers in the way that a semi-structured interview or group interaction would have—and there wasn't a mechanism for following up on a particular response or asking clarifying questions. Further, answering many open-ended questions in a single survey can be more fatiguing than an interview. As a result, this process likely gathered more basic or self-evident responses than a live conversation might have done. In addition, the structural choice to capture recipients' responses one letter at a time also meant forgoing the opportunity to draw out a comparative dimension among these alternative futures and prospective applications of social agents. Such a comparative element, especially if pursued in conversation, and with different respondents in dialogue, seems a missed opportunity afforded by the research design up to that point, but not able to be taken up in this case. We suspect that these two features, although partly brought about by practical pandemic considerations, contributed to the lack of significant novelty and surprise in the results.

9 DISCUSSION

This project made use of and extended the EXF method in several ways in the effort to create opportunity for ethical reflection on tomorrow's possible social agents. We believe the approach could be valuable to those interested in revealing and exploring underlying tensions and dilemmas within the technology design space and community, and even in promoting change, on all sorts of technical fronts. That said, the study findings were somewhat limited in depth and novelty. We discuss the strengths and weaknesses of the approach, and suggest ways of building on this work to further develop EXF methodology for speculation and reflection.

9.1 Probes: Simple but Engaging

The paper letters used in the study, created in high fidelity, were successful in drawing people into the story (even to the extent that, as we learned, on occasion participants were convinced for a moment they were real). Making these convincing did not require technical development per se, but rather a focus on storytelling, including diegetic justification for why the narrative was being told through the seemingly old-fashioned mechanism of a physically mailed letter (e.g., to avoid data-tracing). The choice of medium also deliberately kept the probes low-cost, while still inquiring into the social, cultural and political implications of possible futures. Moreover, the risk of malfunction for paper probes was minimal—if the letter arrives, the participant will likely experience it as planned. If the letter returns (which did occur several times), the researcher could easily address the problem by re-sending it.

9.2 Experiences: Limited but Widespread

In contrast to previous experiential futures (and EXF) projects in which participants were physically present to encounter elements of future world [11, 23], here we were able to include a broad range of participants in our study remotely, while keeping the experience engaging. Participants were distributed across the US, and could participate thanks to the reliability of mailing systems; an especially valuable affordance at a time when in-person gathering was not possible. Further, the study was conducted asynchronously, so participants could attend to the letters they received at their convenience. Arriving over the span of several weeks, they provided a slower pace for reflection and consideration, while also letting anticipation build between letters. A downside is that these remote, asynchronous experiences, and certainly the fragmented mechanism used to elicit responses, may have compromised the depth and novelty of the "conversation" that followed. To fully benefit from experiential research at this relatively modest scale, more social modes of engagement would be recommended, such as semi-structured follow-up interviews, or better yet, focus group

discussions allowing the surfacing of richer compare-and-contrast type responses.

9.3 Narratives: Distanced but Relatable

The stories told through the letters maintained some distance from today and from participants—we believe this distance contributed to their ability to reflect and observe ethical concerns in these possible futures, and in the field more broadly.

First, the scenarios were situated in the distant—but not too distant—future. This allowed participants to reflect on possible scenarios, while perhaps avoiding the sense of it being too late to change the collective trajectory of the technology. Second, the scenarios varied between telling a story about the participant themselves, about their family member, and about a social movement. It seems the middle distance of someone socially close was the most engaging, giving participants a clear "role" in the future, but not casting them as the "protagonist," which may have required more effortful investment in the suspension of disbelief.

The "distance" of the narratives from each other is another design factor in play. Each letter told a story within a completely different domain (educational, personal and communal), and arguably, in mutually exclusive future worlds. Had the artifacts differed less, and so been more directly comparable, that could make the topic amenable to investigation in a different way (e.g. three social agents differently inflected but all in the education realm). The breadth of the topical net that was cast here made it difficult to draw conclusions across all three scenarios, not just because their contents did not invite direct comparison, but critically, because the debrief process did not either. The general and somewhat predictable responses received might have been quite different, though, had the conversation been more deliberately constructed around identifying points of ethical concern, dilemmas, or possible technology- and policy-based resolutions. Future work might use a similar approach, but limit the probes to a more specific design space and enabling a more deliberate form of compare-and-contrast conversation.

9.4 Response Collection: Structured but Flexible

Lastly, we attempted an alternative form of reflection and response collection from what has been previously done. Prior studies have used, for example, on-the-go surveys [11] and informal discussions [22, 23] to evaluate and reflect on possible future experiences. The distributed, asynchronous approach in this case gave participants more time to think.

A disadvantage of using the survey mode was that it did not make the fullest use of people's experiences and reflection processes. Certainly then, future efforts might consider alternative modes of response collection; for instance, asking participants to respond to letter probes with letters of their own. A more straightforward approach would be to redesign a similar study in synchronous form, and elicit responses through interviews or group discussions. Even amid the pandemic, a live online conversation might have more successfully tapped the "potential energy" of insights elicited by the artifacts.

10 CONCLUSION

This case study used the hybrid design/futures method Ethnographic Experiential Futures (EXF) in an effort to materialize and prompt discussion of a selection of currently under-examined ethical dilemmas within the community of Human-Agent and Human-Robot Interaction researchers. We interviewed three subject matter experts about futures worth paying attention to in the emerging design space of social agents, and engaged 15 research respondents in a series of modestly-scoped but tangible experiences of these three futures. We mailed them three "letters from the future", asking them to reflect on each one and its implications.

The contribution here is primarily methodological—presenting the design decisions made in this process and the lessons learned from each step; which choices worked well, which worked less well, and how future research efforts might elicit richer conversations and insights into the design spaces and dilemmas of emerging technologies. A key limitation was that responses were elicited to probes that, while similar in format, varied across multiple other substantive dimensions simultaneously—domains, characters, stories, and images of the future. Work to follow might learn from this by zooming in on a narrower portion of the design space; for example, a subtopic like social agents applied within education. Alternatively, the wide-ranging diversity seen here might be more richly leveraged to seed a commensurately wide-ranging, comparative ethical conversation, via a live, facilitated debrief process such as interviews or focus group formats.

Equally, as this study intentionally canvassed a wide swath of future possibility and design space with just three artifacts, there is also exciting potential for future research that could involve enlisting much larger numbers of people in exploring an even wider array of issues, instantiated through a greater number of exemplary future artifacts; a more systematic distributed mapping of the promises and perils of the technological futures that we are collectively creating.

ACKNOWLEDGMENTS

This work was supported by Situation Lab (@sitlab), as well as the National Science Foundation under Grant Number SES-1734456.

REFERENCES

[1] AI Google 2018. Artificial Intelligence at Google: our principles. https://www.blog.google/technology/ai/ai-principles/.
[2] AI Microsoft 2018. The Future Computed: Artificial Intelligence and its role in society. https://news.microsoft.com/uploads/2018/01/The-Future-Computed.pdf/.
[3] James Auger. 2014. Living with robots: A speculative design approach. *Journal of Human-Robot Interaction* 3, 1 (2014), 20–42.
[4] 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*. 421–433.
[5] David N Bengston. 2016. The Futures Wheel: a method for exploring the implications of social–ecological change. *Society & Natural Resources* 29, 3 (2016), 374–379.
[6] Hugh Beyer and Karen Holtzblatt. 1997. Contextual design: A customer-centered approach to systems designs.
[7] Julian Bleecker. 2009. Design Fiction: A short essay on design, science, fact and fiction. *Near future laboratory* 29 (2009).
[8] Mark A Blythe and Peter C Wright. 2006. Pastiche scenarios: Fiction as a resource for user centred design. *Interacting with computers* 18, 5 (2006), 1139–1164.
[9] Petter Bae Brandtzaeg and Asbjørn Følstad. 2018. Chatbots: changing user needs and motivations. *Interactions* 25, 5 (2018), 38–43.
[10] Laura Burbach, Chantal Lidynia, Philipp Brauner, and Martina Ziefle. 2019. Data protectors, benefit maximizers, or facts enthusiasts: Identifying user profiles for life-logging technologies. *Computers in Human Behavior* 99 (2019), 9–21.

[11] Stuart Candy. 2010. The futures of everyday life: Politics and the design of experiential scenarios. *University of* (2010).
[12] Stuart Candy. 2018. Gaming futures literacy: The thing from the future. In *Transforming the Future*. Routledge, 233–246.
[13] Stuart Candy and Jake Dunagan. 2017. Designing an experiential scenario: The people who vanished. *Futures* 86 (2017), 136–153.
[14] Stuart Candy and Kelly Kornet. 2019. Turning foresight inside out: An introduction to ethnographic experiential futures. *Journal of Futures Studies* 23, 3 (2019), 3–22.
[15] EunJeong Cheon and Norman Makoto Su. 2018. Futuristic autobiographies: Weaving participant narratives to elicit values around robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 388–397.
[16] Scott Davidoff, Min Kyung Lee, Anind K Dey, and John Zimmerman. 2007. Rapidly exploring application design through speed dating. In *International Conference on Ubiquitous Computing*. Springer, 429–446.
[17] Anthony Dunne. 2008. *Hertzian tales: Electronic products, aesthetic experience, and critical design*. MIT press.
[18] Anthony Dunne and Fiona Raby. 2013. *Speculative everything: design, fiction, and social dreaming*. MIT press.
[19] Tony Fry. 2007. Redirective practice: an elaboration. *Design philosophy papers* 5, 1 (2007), 5–20.
[20] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *interactions* 6, 1 (1999), 21–29.
[21] 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.
[22] Tom Jenkins, Laurens Boer, Juliane Brigitta Busboom, and Ivar Østby Simonsen. 2020. The Future Supermarket: A Case Study of Ethnographic Experiential Futures. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–13.
[23] Kelly Kornet. 2015. Causing An Effect: Activists, Uncertainty & Images of the Future. (2015).
[24] Sandjar Kozubaev, Chris Elsden, Noura Howell, Marie Louise Juul Søndergaard, Nick Merrill, Britta Schulte, and Richmond Y Wong. 2020. Expanding Modes of Reflection in Design Futuring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.
[25] Michal Luria, Judeth Oden Choi, Rachel Gita Karp, John Zimmerman, and Jodi Forlizzi. 2020. Robotic Futures: Learning about Personally-Owned Agents through Performance. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 165–177.
[26] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 633–644.
[27] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. Social Boundaries for Personal Agents in the Interpersonal Space of the Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[28] Jennifer Manuel and Clara Crivellaro. 2020. Place-Based Policymaking and HCI: Opportunities and Challenges for Technology Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.
[29] mccarthy2018 2018. Feeling at Home Between Human and AI. immerse.news/feeling-at-home-between-human-and-ai-6047561e7f04/.
[30] Evgeny Morozov. 2013. *To save everything, click here: The folly of technological solutionism*. Public Affairs.
[31] William Odom, John Zimmerman, Scott Davidoff, Jodi Forlizzi, Anind K Dey, and Min Kyung Lee. 2012. A fieldwork of the future with user enactments. In *Proceedings of the Designing Interactive Systems Conference*. 338–347.
[32] James Pierce. 2019. Smart home security cameras and shifting lines of creepiness: A design-led inquiry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
[33] Laurel D Riek. 2012. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012), 119–136.
[34] Britta F Schulte, Paul Marshall, and Anna L Cox. 2016. Homes for life: a design fiction probe. In *Proceedings of the 9th nordic conference on human-computer interaction*. 1–10.
[35] Tristan Schultz, Danah Abdulla, Ahmed Ansari, Ece Canlı, Mahmoud Keshavarz, Matthew Kiem, Luiza Prado de O Martins, and Pedro JS Vieira de Oliveira. 2018. What is at stake with decolonizing design? A Roundtable. *Design and Culture* 10, 1 (2018), 81–101.
[36] Reinhard Sefelin, Manfred Tscheligi, and Verena Giller. 2003. Paper prototyping-what is it good for? A comparison of paper-and computer-based low-fidelity prototyping. In *CHI'03 extended abstracts on Human factors in computing systems*. 778–779.
[37] Michael Warren Skirpan, Jacqueline Cameron, and Tom Yeh. 2018. More than a show: Using personalized immersive theater to educate and engage the public in technology ethics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[38] Bruce Sterling. 2009. Cover Story Design fiction. *interactions* 16, 3 (2009), 20–24.
[39] Robert B Textor. 1995. The ethnographic futures research method: An application to Thailand. *futures* 27, 4 (1995), 461–471.
[40] Bruce M Tharp and Stephanie M Tharp. 2019. *Discursive design: critical, speculative, and alternative things*. MIT Press.
[41] Times 2019. Artificial Intelligence Has a Problem With Gender and Racial Bias. https://time.com/5520558/artificial-intelligence-racial-gender-bias/.
[42] Cameron Tonkinwise. 2015. Just design. Being dogmatic about defining speculative critical design future fiction. Medium.
[43] Sherry Turkle. 2017. *Alone together: Why we expect more from technology and less from each other*. Hachette UK.
[44] Sam Wolfson. 2018. Amazon's Alexa recorded private conversation and sent it to random contact. https://www.theguardian.com/technology/2018/may/24/amazon-alexa-recorded-conversatio.
[45] John Zimmerman and Jodi Forlizzi. 2014. Research through design in HCI. In *Ways of Knowing in HCI*. Springer, 167–189.