From Episteme to Techne: Crafting Responsible Innovation in Trustworthy Autonomous Systems Research Practice

Pauline Leonard, University of Southampton

Chira Tochia, University of Southampton

1. Introduction

A key challenge for Responsible Research and Innovation (RRI) is to integrate its principles fully into every aspect of research project design (Forsberg et al 2021). In many science and technology projects, RRI is routinely conceived as an add-on, discrete category or activity which does not fundamentally change the ways in which the research is designed and conducted (Åm 2019; https://www.rri-practice.eu/). This can in part be explained by the epistemological foundations of 'science': corresponding to Aristotle's notion of *episteme,* and growing out of Enlightenment thinking, the gold standard of scientific knowledge is traditionally conceived as that which is objective, value-free and independent (Sovacool et al 2020). In contrast, sociological, feminist, post-positivist approaches centralise acknowledgement that contextual and normative factors inevitably impact upon research processes and knowledge production. By working to develop methods that render these transparent, researchers can act to minimise some of their undesirable and potentially harmful side effects and strive towards social good. To this end, these approaches adopt principles of *techne*, which conceptualise research and the production of knowledge as forms of proficient craftwork (Prasad 2005). Techne involves openly interrogating scientific protocols and strategically adjusting research practice to ensure that RRI principles are thoroughly embedded in methods and design.

This paper brings together the insights of these social science methodologies and the United Kingdom's Research and Innovation (UKRI)'s Engineering and Physical Sciences Research Council (EPSRC)'s Framework for Responsible Innovation[i] to demonstrate how a combined, interdisciplinary approach can offer benefits to RRI. To illustrate how this can be achieved in practice, we draw on reflections and decision-making processes which arose during our qualitative case study research on industrial cleaners' attitudes towards robotics, conducted as part of a larger research project within the UKRI's Trustworthy Autonomous Systems (TAS) Hub, *Trustworthy Human-Robot Teams* (https://www.tas.ac.uk/research-programmes/agile-programme-2/trustworthy-human-robot-teams/). As robotics are increasingly being introduced into a broad range of industrial contexts and employment fields, the project aims to investigate issues of trust and trustworthiness for people working with and alongside robots. The use of robots not only raises issues of jobs and employment, but also retraining and reskilling, economic productivity, safety and reliability, and changes in social relations and even social identities (Elliott 2019). These predicted transformations are as yet unfolding, and the use of robotics is still relatively new in many contexts. As such, 'the jury is still out' in terms of whether we trust them to do the job as well as a human. For the 'fourth industrial revolution' so much cited by politicians and industrialists to turn to reality, robots must be deemed to be fully trustworthy by those on whom they impact: the people whose jobs and/or skills they are replacing as well as the people for whom they are providing a service.

However, researching complex and sensitive issues such as trust raises important RRI questions and challenges. It is these which are the focus of this paper: how may we, as researchers, conduct our studies according to the best principles of RRI? In what follows, we argue that techne offers a way forward. We start by expanding our discussion of the contributions of techne to epistemologies grounded in episteme before describing a range of sociological and feminist conceptual developments which may provide resources by which researchers can become more aware and critical of conceptual frameworks, positions, biases, political affiliations, expectations and justifications (O'Reilly 2012). All these impact on the research craft: the framing of research questions, the decisions taken on research methods and participants, the collection of empirical data and the construction of findings, and the presentation of knowledge. We then turn to our own research to discuss how we crafted these insights into our methods. We concentrate on the qualitative ethnographic interviews we conducted to argue that, by moving from episteme to techne, our research practices can be made both more responsible and innovative.

2. From Episteme to Techne

The idea that 'gold standard' scientific claims, methods and results can -and should be -independent of particular perspectives, value judgements or normative interests has achieved a position of dominance across scientific communities, leading to misrecognition or even denial of the impact of contextual factors on research design and findings. With the concept of objectivity dominating scientific legitimacy in society, how can the normative values of RRI be fully integrated and reconciled within research? The simple answer for the design of many research projects is that they are not: they remain separated, solved by forming an additional work-package wherein the implications of, for example, the UKRI EPSRC's RRI 'AREA' (Anticipate, Reflect, Engage and Act) Framework (https://epsrc.ukri.org/research/framework/area) are considered, rather than being embedded and interrogated within each and every stage of the research process (Åm 2019). However, as Flyvbjerg (2001) has pointed out, within the Aristotelian approach, *episteme* is only one means by which to approach science. Indeed, the broader term *epistemology,* which combines *episteme* (knowledge) with *logos* (account, argument or reason) (Steup and Neta 2020), refers to the study of the different varieties of knowledges, and how these come to be established. Taking this focus exposes how social and personal factors are intertwined with the construction and production of knowledge, with feminist epistemology in particular highlighting the ways in which subjective interests and personal values affect methods of collecting evidence, undertaking analysis and the presentation of knowledge (Letherby 2003). Epistemologists argue that while these factors are often hidden or unacknowledged in the presentation of science, they are always there, nonetheless. In a similar vein, 'interpretivists' such as Max Weber, posit that there can be no knowledge independent of thought and language (Weber 1968). As a result, it has become regarded as good practice in *post*-positivist research practice to identify one's own subjectivity, bias and politics and make these transparent. For example, for researchers aiming to take a feminist approach to their research, a statement of commitment to making gender visible is foundational.

In consequence, within post-positivist social science, and particularly influential within qualitative research methodologies, it is *techne* rather than episteme which dominates. While this Special Issue on Trustworthy Automated Systems (TAS) will illustrate that many research projects within the TAS Hub do not draw on qualitative research methods, we argue that *techne* offers broad and valuable insights for RRI within research on trustworthy automated systems whatever the research methods. Influenced by

notions of artisanship and craftwork (Prasad 2005), *techne* is an activity which is 'concrete, variable and context-dependent. The object of *techne* is the application of technical knowledge and skills according to a pragmatic instrumental rationality' (Flyvbjerg 2001: 56). Here, each stage of the research process is carefully crafted to embed the ontological and epistemological assumptions of the paradigm (Kuhn 1970) within which the project is set. Thus, if the paradigm in TAS is RRI, then the assumptions and values embraced within this paradigm, such as creating automated systems which are socially desirable, serve the public interest and minimise harm, need to be crafted into the whole research process. Rather than aiming for universality as in episteme, techne openly acknowledges the uniqueness and distinctiveness of context, searching for appropriate tools by which to gain rich and detailed space- and time-specific understanding.

All of the projects within the TAS Hub are designed with the aim of improving the quality of people's lives. In our own project, Trustworthy Human-Robot Teams, we explore the challenges to trustworthiness for industrial cleaning robots, which offer health and safety benefits. What are the concerns of the people involved in their use: cleaners, facilities managers as well as the space users? It is clear that these groups will encompass people from a wide variety of backgrounds. To craft our research responsibly, to make sure we minimise potential social harms, we need to take a detailed and granular approach to understanding the diversity of voices within these categories, and the differential impact of the project design upon different individuals.

Further, researching human-robot relations present other distinctive features for the craftwork involved in RRI. As science and technology studies (STS) emphasise, technology is far from a neutral or inert presence (Mackenzie and Wajcman 1999), and technological devices can play an important role in the production of social outcomes (Halford et al 2015). Actor Network Theory (ANT) exposes how the relations between humans and non-human actors are contingent on particular times and spaces, such that the effects of technological affordances are never universal, fully knowable or determined (Latour 2005). ANT thus argues that technology cannot be placed in a separate category to humans within our ontological and epistemological frameworks. The interactions between humans and machines means that boundaries are blurred, and outcomes co-constituted through the human/non-human nexus. That robots are often subject to anthropomorphism illustrates this ambivalence further, with gender often playing a key role in this humanisation. For example, that the majority of digital assistants have feminine voices may not only impact on the form of human/machine relations but perpetuate gender/technology stereotypes in the process (Costa 2018).

Frameworks for achieving RRI thus need to embrace these multiple complexities. We now turn to consider how sociological and feminist approaches can contribute not only to our conceptual understandings but also to our methodological practices.

3. Crafting the Framework

To achieve the principles of Responsible Innovation, the UKRI EPSRC recommend taking an approach that encompasses the following steps: Anticipate, reflect, engage and act (AREA). The key principles involved are summarized in Table 1 below:

Anticipate	Describing and analysing the impacts, intended or otherwise, (for example economic, social, environmental) that might arise. This does not seek to predict

	but rather to support an exploration of possible impacts and implications that may otherwise remain uncovered and little discussed.
Reflect	Reflecting on the purposes of, motivations for and potential implications of the research, and the associated uncertainties, areas of ignorance, assumptions, framings, questions, dilemmas and social transformations these may bring.
Engage	Opening up such visions, impacts and questioning to broader deliberation, dialogue, engagement and debate in an inclusive way.
Act	Using these processes to influence the direction and trajectory of the research and innovation process itself.

Table 1 1: Key Principles of RRI. *Source: (https://epsrc.ukri.org/research/framework)*

Strong synergies exist between the EPSRC's AREA principles and postpositivist sociological approaches to empirical research, particularly those informed by feminist debates on the relationship between the processes and the products of research, where issues of responsibility and politics are central. However, a key contribution of a postpositivist approach is to challenge chronological processes, as are implied in Table 1, instead understanding that research involves iteration and circularity, and constant revisiting of assumptions, reflections and decisions. Thus, while in our ensuing discussion we take as a heuristic device each of the AREA principles in turn, to consider how social scientific theoretical and methodological insights can add flesh to their bones, we prefer to see these principles as mutual, intertwined and helical. As such, we focus in most detail on the 'Anticipate' category, conceptualizing the aspects involved here *not* as preliminary but as threads which weave through all the other principles.

3.1 Anticipate

A fundamental principle of feminist research is to take responsibility for the knowledge that we produce and acknowledge our accountability in terms of the impact that our research may have (Letherby 2015). This starts with anticipating the effect of the research process on the people involved: research participants, members of the research team itself as well as stakeholders who may be affected by the outputs. To attempt to understand the diversity of positions, the sociological practice of 'making the familiar strange' (Mills 1959), whereby alternative subjectivities, perspectives and experiences are de-familarised, is an important conceptual tool. Central to this is sensitivity to issues of difference such as gender, race and ethnicity, social class, age, religion, regionality/nationality, sexuality and disability etc., and an awareness of the interplay of politics, power and emotion within the research process. Implicit here also is an appreciation that context matters: that knowledge is situated and produced in specific configurations of time and space and founded upon meanings and understandings that are constructed or reconstructed within locales (Edwards and Holland 2020; Kvale and Brinkmann 2015). Important questions are therefore raised as to who is to be involved in the research, where it is to be conducted, and when, and over what time period. To what extent, for example, can the experiences of one social group, for example, young white male students involved in a laboratory experiment in the morning represent the experiences of another social group with an entirely different set of social characteristics in the multiple, and highly diverse, times and spaces of the 'real world'? Further, to what extent do the

social characteristics of the researchers themselves impact on the conduct and outcomes of the research (Bourke 2014)?

While such questions have predominantly been regarded as most apposite to qualitative research methods such as interviews, RRI means that they must also be considered within positivist and quantitative approaches where implicit biases such as 'default male thinking' can be pervasive (Criado-Perez 2019). This is particularly evident within technology related research. The fact that women are chronically under-represented in science and engineering feeds through into who is most likely to be involved in researching robotics and human-automated systems interaction. The fact that men are much more likely to be technology researchers and designers, and women are more often technology users (Wang and Young 2014), needs to be anticipated in the forming of teams, framing of research questions and research methods decisions. Of course, diversity does not stop here: gender intersects with a vast range of other social factors, including skills levels, social positions, prior experience and societal norms and expectations (Kendall et al 2020: 585).

How can we, as researchers, hope to anticipate 'the possible impacts and implications that remain otherwise remain uncovered and little discussed' as recommended within the AREA framework? Towards this end, we outline some of the conceptual tools adopted by feminist/qualitative researchers for responsible research which may have general value for anticipating the multiplicity of social issues. These include positionality, power, insider/outsider status, emotion and sensitivity.

3.1.1.Positionality

Anticipating the impact of our research on others involves turning the spotlight on ourselves: how are we, as researchers, positioned in relation to the research process and our research participants? Acknowledging our own 'positionality': our own social status and identity, biases, motivations for doing the research and expectations of participants is a complex process, but 'better to understand the complexities within research rather than to pretend that they can be controlled' (Letherby 2003:71). How we refer to those involved in the research can be revealing: are they objects, subjects, informants, respondents, participants, advisors, or beneficiaries, for example (Edwards and Holland 2013; Letherby 2015)? The choice of term will reflect the perceived positionality of the researcher in relation to those on whom the research impacts, as well as the relations between the researcher, the researched, and any other stakeholders, and reveal to what extent others are understood to be involved in the co-production of knowledge.

3.1.2. Power

Researchers will inevitably exercise power through their positionalities, as well as their research design, and the effects of these need to be anticipated. To avoid research as something which is 'done *on* the relatively powerless *for* the relatively powerful' (1978 : 25, original emphases, cited in Edwards and Holland 2013: 79), consideration needs to be given to how participants may respond and be made to feel, and, if necessary, steps taken to alleviate senses of powerlessness and enable participants to feel comfortable. Power imbalances may also result from broader social-structural relations and the entrenched differences in power and resources which exist within society around race, gender, social class, age, educational background etc. (Edwards and Holland 2013). As well as any benefits and/ or challenges which may be derived from these social positions, academic researchers come to the research context already powerful, through holding significant degrees of power and status in society as

"experts". This may feel threatening to some participants, particularly members of marginalised groups who may be very unused to being involved in academia/research. At the same time, following Foucauldian notions of power (Baxter 2003), researchers may be simultaneously power*less*: their need to collect data positions them as 'vulnerable knowledge seekers' (Edwards and Holland 2013: 78), reliant on participants for their research to generate the outputs desired. Further, conducting research with "elites": powerful members of society, can subject researchers to different sorts of social relations. Elites may seek to set the agenda, challenge the researcher about the project and its intentions, and even exert control over the research process (Cousin et al 2018). Anticipating these behaviours, and being prepared for them, will necessitate different strategies for different researchers, depending on their social structural identities and the different resources these deliver (Mikecz 2012). As Edwards and Holland note (2013) both age and gender play important parts in 'studying up' and researching as a young woman, for example, can result in older male participants wanting to exercise control and exert power.

3.1.3. Insider/Outsider

Conducting research with people from different socio-economic backgrounds and in different contexts will generate issues of insider/outsider status. In some instances, researchers may be 'insiders' where, due to their social identities and personal experiences, they may share backgrounds and knowledge with research participants: for example, both researcher and researched may be of a similar age, race and gender, and both may have knowledge of automated systems. In this situation, there may be assumptions made as to what does or does not need to be articulated. Being an insider is seen as both an advantage and a disadvantage (O'Reilly 2012). Some argue that having some pre-existing knowledge, familiarity and/or sharing social background with research participants is advantageous: it may be quicker to build up trust and investigate sensitive issues such as trustworthiness if commonalities in identities exist (Becker 2010). For example, for feminist researchers, it is assumed that the richness of data collection will be improved if women researchers conduct research with/on other women with whom they can empathise and share stories. Conversely, others argue that insider status can dull the sense of enquiry. It is easier to 'make the familiar strange' as an outsider, when the researcher has little pre-existing knowledge and goes into the empirical stage bursting with questions. Outsiders can more easily note the 'unconscious grammar' of a situation and make assessments with a fresh eye. However, these binary categories are arguably in themselves misleading, as all of us to some extent can position ourselves simultaneously as both insiders and outsiders at different points in the research.

3.1.4. Emotion

The power relations produced through the research process, as well as the content of the research itself, will inevitably trigger emotions in both researchers and the research participants. Acknowledgement and anticipation of the emotional dynamics which may occur, and how these may impact on the production of knowledge, are essential components of RRI, as is giving thought to emotion management. From a psychosocial perspective, power and emotion come together in both conscious and unconscious ways (Edwards and Holland 2013). The notion of the 'defended subject' encapsulates how anxiety is inherent in the human condition and consequently unconscious defences can come into play in research interactions, just as they do in all types of social interactions (Hollway and Jefferson 2012). There are thus emotional costs on both sides: for researchers who may have to deal with displays of anxiety, mistrust, resentment, sadness and even anger, and for the participants

experiencing such emotions. Recognition of, and attention to, the dynamics of power and emotions within research can only enhance the quality of the knowledge production process (Edwards and Holland 2013).

3.1.5. Sensitive Issues: Trust

Issues of power and emotion are especially pertinent when researching complex conceptual and sensitive issues, such as trust and trustworthiness, the key conceptual foci of the TAS Hub. At the same time, trust/worthiness are difficult concepts to research: complex and 'slippery' (Guillemin et al 2016); 'non-material, abstract and lack[ing] a generally agreed-upon definition' (Ayrton 2020: 1233). Trust can be understood as fundamentally relational, occurring between interdependent agents, both human and, in the case of the TAS research, technological/material. Trust is particularly relevant in contexts of uncertainty and even vulnerability, such as, in terms of our own research study, when we put our work tasks in the hands of robotics, and when we enter a physical space that we hope is clean and safe. Without trust, action will not proceed or may proceed in a context of high emotion.

How the members of the research team understand and conceptualise trust will impact on the research design and the presentation of findings. However, philosophical understandings of complex concepts such as trust often remain underdiscussed within research teams (Guillemin 2016). Ayrton (2020) notes that two competing understandings of trust dominate: trust conceived as an attitude and trust as a form of cooperative behaviour (Li 2012). Both approaches assume a rational form of strategic thought and/or action, whereby the trustor calculates the extent to which 'the trusted encapsulates the interest of the trustor and therefore has the incentive to be trustworthy in fulfilling the trustor's trust' (Hardin 2002: 24). Much of the research into trust is based on the assumption that this rational calculation can be measured, for example through attitudinal surveys, games and experiments (Rousseau et al 1998; Calhoun et al 2019). However, critics challenge the assumptions underpinning the trust as attitude/behaviour dichotomy: that trust is uni-directional (from the trustor to the trusted), stable and intentional. Instead, a third approach to conceptualising trust is preferred, which understands trust and the gaining of trustworthiness as a *process:* context-specific and contingent, always ongoing, being made and remade, or otherwise (Möllering 2013). This approach to understanding and researching trust is, however, far from straightforward in methodological terms and will demand careful yet creative craftwork (Ayrton 2020). In all of these approaches however, trust requires the goodwill of the trustor towards the trusted, reinforcing how the trustor is in a structurally vulnerable position, having to rely on the benevolence of the person or the reliability of the technical affordance, such as a robot, without any certainty that this exists (Guillemin et al 2016).

In terms of the 'Anticipate' phase of RRI, the key point is that the research team acknowledge the range of approaches to conceptualising trust, the relationship these have with research design and selection of methods, and the impact these understandings and choices will have on research participants. From the RRI perspective, trust is thus both a noun and a verb: something we aim to understand *and* actively do (Guillemin et al 2016).

3.2 Reflect

The above discussion underscores how taking a critical approach to the research process and its implications, both before we start and during, involves thinking methodologically about our methods. Feminists argue that methodology is concerned with the 'getting of knowledge' (Stanley and Wise 2008:222) and is key to understanding the relationship between knowledge and power. Challenging episteme, this approach argues that 'the view from nowhere was always in fact a view from somewhere' (Spencer 2001:444). *Reflexivity* involves reflecting critically on the 'somewhere': the contexts of knowledge production, which are often highly specific scientific and disciplinary environments, such as universities.

In addition, reflexivity involves reflecting on the social structural relations within the research team and with research participants. Whatever the time period of the project, social relationships will be built, and friendships made, which subsequently may either be retained or ended. While hierarchical relationships pre-exist the research, these may be challenged during the research process, and huge potential exists for misunderstandings and mistakes (O'Reilly 2012). Coffey (1999) advises careful reflection about the possibility of exploitation of social relationships, such as when we ask friends to act as gatekeepers or key informants. At the same time, as researchers, we are also vulnerable to feeling let down or even hurt. RRI can thus also involve careful reflection on, and craftwork of, ourselves.

Reflexivity also includes analysing how we interpret and represent our research findings, for example, in the texts where they are written up. Feminist scholars have shown how our choices are both personal and political, bound up with our own biographies, motivations for undertaking the research project, relationships with our research participants and desired-for outcomes. Reflexivity acknowledges this complexity and gives time and space to their consideration and is thus an integral part of RRI.

3.3 Engage

Being open about our reflections raises important questions about inclusivity: who should be involved in our methodological decisions? While it is deemed responsible to include all research team members, feminist and collaborative researchers believe that research should also be a two-way process with participants, involving them in the research process from initial design, through data collection and analysis, to the practical application of findings (O'Reilly 2012). RRI thus demands that we give due consideration to where and how we can engage with both our team members *and* our participants. At the same time, there may be good reasons for not disclosing certain aspects of the research to participants, if this involves breaking confidences, risking anonymity or causing harm (O'Reilly 2012).

3.4. Act:

Our discussion highlights how 'decisions made implicitly or explicitly at the pre-empirical stage reverberate and continue throughout, defining what is discoverable' (Ayrton 2020: 1245). We have drawn on sociological, feminist and postpositivist insights to deconstruct the range of issues: social, political, ethical, epistemological, and methodological, which are embraced within the AREA framework and on which we need to act. We now turn to introduce our Case Study, proceeding then to discuss our techne: the ways in which we drew in practice on the conceptual tools offered by these perspectives to produce a recrafted AREA framework by which to embed our RRI approach into our research methods.

4. The Case Study

The Case Study was set within a university setting, focusing on cleaners whose work involves industrial levels of cleaning and disinfecting of laboratories and lecture theatres. A key innovation of the project was the prototyping of a UV disinfecting robot, or 'cobot', a robot designed to collaborate with humans through sensors for verification. The appearance of our robot, 'Zappy', is in part humanised, taking the shape of a large square white box with a digital screen showing an animated face and with wheels to move around. Cleaners were introduced to Zappy and watched them perform before our interviews.

To explore trust issues in depth, semi-structured interviews were selected as a key method. These allow us to generate a list of questions and topics we want to cover, but also afforded flexibility in how and when the questions are put, enable additional probes and follow-ups as relevant and interesting topics arise in our discussions with our participants. The aim is to allow much more space and time for interviewees to answer on their own terms than structured interviews or surveys. As such, we felt as a team that they offered and important extension to our RRI within our research design, to some degree enabling the sharing of power within the interview context. At the same time, the presence of common set of topics provides some structure for comparison across interviewees (Edwards and Holland 2013).

In addition, to further enhance RRI, we wanted to conduct these ethnographically, *in context*. Ethnography involves spending time 'in the field', or site of study, and enables 'interpretation of local and situated cultures based on paying attention to the singular and concrete' (Edwards and Holland 2013: 30). This experience of 'being there', and observing what happens on the ground, helps to deepens researchers' understanding of the aspect of participants' lives under investigation: in this context, this was the spaces where the cleaners worked daily. Pink (2009) notes how ethnography is a *sensory* experience, involving not only observation but engaging the senses of sight, smell, hearing, touch and even taste, as well as the emotions, both in practice and in purview (O'Reilly 2012). In contradistinction to any idea of 'the detached, independent researcher' therefore, this approach understands that enhanced understanding is gained by sharing experience with research participants. The benefits are not only to the researcher, who is able to experience the locale of interviewees' lives, but also to the researched, who may feel more empowered and relaxed in a setting which is familiar to them, rather than a space identified and provided by the research team.

A key aspect of ethnography is the production of fieldnotes, which enable and even prompt reflection on the research process. These vary hugely in style according to the author and are inevitably quite personal. Especially at first, they provide an important record of early, and often emotion-filled, impressions: which O'Reilly (2012: 102) terms *'insider sensitivities*.' While these notes are rarely shared or included in published outputs, in this article we flout this convention to draw on Chira's fieldnotes as a means by which to illustrate the RRI considerations that evolved in situ and in practice. We use these to focus on the reflections on our research tools and research processes and our interactions with our participants, once again structuring the discussion heuristically through the principles of the AREA framework.

5. **Techne: Re-Crafting the AREA Framework**

5.1 Anticipate

The anticipate stage involved the research team discussing their pre-existing assumptions and imaginations of participants and how they might respond to the project topic and our questions. To

encourage ongoing reflection, Chira was invited to diarise her thoughts and responses. From the beginning, the topic of our research: Trustworthy Cleaning Robots, was perceived to raise a range of issues which needed to be anticipated in terms of their potential for sensitivity. The question 'will robots steal our jobs?' is a noisy one in the media and public debate, with pundits predicting that many jobs will disappear due to automation (Ford, 2016; Elliott 2019). These concerns formed the background for Chira's anticipations of the interviews themselves:

I believe there is an overarching view that those in the cleaning industry are not paid well and therefore it also might be uncomfortable to discuss themes such as job risk and security, or make claims that their roles could be taken over with robot automation. I think if I was in their position and someone was asking me about the idea of an autonomous robot conducting my job and I really thought about how this interview might be used I might even say certain things so that the outcome resulted in suggesting it is a bad idea, if I was worried about my own job security.

Bearing this in mind, a key skill in research craftwork therefore is to build trust, reassure participants on the provenance, aims and context of the research, as well as ethical principles such as confidentiality. However, when thinking about our participants, our feminist approach reminded us how, as we start to design the research tools we will be using to investigate our research questions with or research participants, such as interview guides, we often rely on stereotypes to build a preliminary and general picture of our respondents (Pickering 2001). Acknowledging the predilection to use stereotypes is important, as it helps us to anticipate and question our own preconceived positions and possible biases, not least in terms of marginalised characteristics such as gender and race (Connell 1987). In this honest extract, Chira is transparent about her initial thoughts as she anticipates interviewing the cleaners:

Writing up the interview guide, I noticed I paused over using certain words. I always want to make sure I'm using the most appropriate language for my participants; however, I caught myself thinking thoughts such as, "I have no idea on the average education level of cleaners" and "a lot of cleaners that I have previously interacted with before tend not to be fluent in English or native English speakers". So I felt that I should be even more cautious with the vocabulary I plan to use. I've also noticed I'm putting in more examples, ready to explain in case the question is not clear for them too. I'm annoyed at myself, why is it that I think because I'm going to be speaking to cleaners that they won't understand the work we're doing as easily as other workers, or that they'll need more clarity on the questions I'm going to ask? I assume they use technology I've never even thought of in their roles and have a much higher understanding of solvents than I do too. [field notes 30/11/21]

Chira is highly reflexive of the ways in which her own biography and personal experiences shape and frame her preliminary picture of our research participants at the anticipation stage. On the one hand, she uses her reflections to craft her use of language carefully, and plan ahead with illustrative examples. On the other, she is also aware that she may need to recraft her approach in situ, in response to her participants as she finds them.

In response to these initial thoughts on cleaners, Chira was encouraged to reflect further on her ideas about cleaners:

5.2 Reflect

I can't think off the top of my head if I have any friends or family in cleaning roles and therefore feel very ignorant about what this role can entail. I remember temping at the hospital my mum works at. She knew all the porters and cleaning staff, so we would always stop for a chat if I was walking around with her. At my own institution, I exchange niceties with the cleaning staff when I work late or if I pass them in the corridor but sometimes it wouldn't be reciprocated, I don't know if that was a language issue or I wasn't loud enough or just they didn't want to engage in conversation. The ones I did have were very shallow chats, such as 'how are you?', 'have a nice evening', 'see you tomorrow'. I always felt a bit awkward as they came into the office to empty the bins and hoover around me, or I'd roll back in my chair so they can do that spot, I'm not really a fan of being waited on and it felt odd that they were cleaning up after me, even if my area was pretty much spotless.

Chira is honest about her positionality and her emotional feelings of discomfort around this. Many women feel guilty about other people cleaning for us: deeply embedded assumptions about cleaning being 'women's work' mean that women routinely feel that we should be doing our own cleaning (Poortman and Lippe 2009). The knowledge that cleaners are disproportionately poor women, women of colour and migrants (British Cleaning Council 2017) who have long histories of being underpaid, overworked and hidden from view, can adds to white middle class women's 'shame' (Bothelho 2021). In addition, Chira's reflections also identify that many cleaners remain invisible; performing their work at times when other workers are not present, or in the process of arriving and leaving (Orgad and Higgins 2021). However, importantly, as individual researchers, we bring our own identities, politics and insider/outsider subject positions to mediate these social relations, as Chira reflects:

Being brought up in a working class family, I could assume we might have the same political views. I myself am left wing, with the view that we should have more policies that are focused on helping those who need more support, or benefit working class groups more.

From past experience, my assumption is that the majority of cleaners are not white or from England - being mixed race myself, and a first-generation immigrant, might make me appear more relatable - do I share this information with them to appear as such?

On the other hand, I have never really had a deep conversation with someone in this position.
I really have no knowledge of the depth of their role, I only understand domestic cleaning tasks. I don't understand the nuances of cleaning such large rooms, types of products they need knowledge of, how often they work, do they work alone or are they sometimes in teams, do they have set spaces they always clean or are they shuffled about?

Chira is aware that all of these issues will impact on the power relations between interviewer and interviewee. Anticipation and reflection become an iterative process as the implications of these challenges are addressed in terms of how they might impact on the interview process:

There might be some other interesting power dynamics, naturally more power is held in an interview with the interviewer as they know what line of questioning is coming and are the ones in control of the situation. The fact I may be younger than the majority that I interview, yet I may appear to be their senior in the interviewer-participant relationship, could also be intimidating. I will also most likely be more educated than them and I wonder if that will make them more self-conscious on how they answer

questions, I've had it before in interviews where people have this notion that because I was studying towards my PhD that I was super smart and therefore they were embarrassed if they don't answer a question eloquently, apologising, as though I would judge them for it.

The ways in which answers to interview questions might reflect broader emotions and anxieties is further acknowledged in Chira's recognition that interviewing employees about their work at their place of work is complex. Indeed, a criticism of the interview method is that they generate statements which are liable to be determined by the situation-the interview context- rather than any empirical 'reality' beyond that context. While for qualitative researchers this does not make them any less interesting or valuable, RRI demands that we craft the research process to try and minimise feelings of concern. Crafting includes an awareness of the influence of place on perceptions, emotions and power hierarchies, and reiterating ethical considerations such as confidentiality:

Hopefully having the interviews take place either in their place of work or wherever they are comfortable when conducted virtually will make them feel more relaxed and at ease initially. Then again, they may feel uneasy talking about work whilst at their place of work if they wish to share any negative views…. there might be some resistance to discussing worries and challenges about their role and the organisation because of a fear it might be reported back, so I am making a note to self to really emphasise that everything shared is confidential and not going to have an effect on their job at all.

Within the Anticipation/Reflect stages, RRI is thus crafted by being as open and transparent as possible, drawing on feminist insights to be rigorous and systematic about addressing sometimes uncomfortable issues such as power, positionality, insider/outsider, emotion and sensitivity, and practising reflexivity in taking account of their potential impact. At the same time, despite this preparation, we need to be prepared that our preconceptions may well still be challenged when we enter the fieldwork site and engage with 'real' participants, where new contingencies may present themselves. RRI is thus a continuing and ongoing process of craftwork, forged through each and every dynamic as the research unfolds. This is demonstrated in the next extract which also reveals how a further, unanticipated challenge was presented by the COVID-19 context within which we found we were operating:

The first few things that surprised me were that they were all British, and I also forgot I would be interviewing people in the North of England, so most had a different dialect to me. This made some of the interviews easier due to no language barrier, but also difficult as we conducted the interviews with masks on (due to COVID-19) and sometimes speech was slightly muffled which made certain words harder for me to understand in their accent.

One participant was [much younger than I expected] - a year younger than me at 31. Even though the first two participants I interviewed were in their 50s and had said they'd been in the industry for a couple of decades I still think I thought cleaners were always going to be older than participant 4. He technically worked in estates with an element of cleaning responsibilities, but it surprised me still.

I also had assumed that we would be conducting these interviews in their workplace and had not taken into consideration that the university has many campuses and that they had never had to clean the

building we were in, so I couldn't tell if they felt more, less, or indifferent about having the interview take place in their place of work.

These unanticipated factors impacted on Chira's positionality. The fieldnotes reveal how reflection is an immediate, in-the-moment activity, as well as being post-hoc. Continual self-monitoring and reflections on the 'next best step' are necessary to craft the gaining of trust with respondents, but they are also essential to the co-constructed process of the Engage principle.

5.3 Engage

Once we meet our participants, new issues may arise in terms of diversity and inclusion. As discussed earlier, feminist scholars have demonstrated how researchers must be highly sensitive to perceptions of inequalities in power, and the emotion and potential anxiety that this might cause. In this context, crafting responsible research skills involve flexibility and quick-thinking, to maintain trust and engage with participants to keep the process positive and productive:

I collected the demographic data whilst I was setting up my devices to record and naturally filled in some data points from observation such as gender and race, but I felt a bit uncomfortable asking about education level because I didn't want them to feel that I was trying to point out mine was higher or anything similar, so I just tried to ask in a neutral tone. One participant tried to recite off the specific grades for certain subjects that they had got for their AS/A-levels which eased any tension as we both laughed about needing such detail.

A few times I could sense that they viewed me as someone in a position of more power: when I asked what kind of person would help them trust a robot via a demonstration, participant 2 said someone like me who knows a lot about them (which I do not and had never said I do!). Also, when I had to try defining or using a different word to "challenges" because the participant didn't understand what I meant by that, it felt like I was using the wrong vocabulary, which I had noted I was concerned with before the interviews.

Participant 1 and I shared a behaviour in replying to our 'Alexas' politely and telling off our partners when they don't speak so nicely to them. I joke about a robot uprising in my house, which is why I do it, but I think for her she had humanised Alexa and therefore it felt more natural to speak to 'her' that way.

The ways in which interviews are a dynamic craft, with power shifting between interviewer and interviewee in the 'interview dance' (Hoffman 2007), is further illustrated in the following fieldnote. In addition, the emotional benefits and costs are revealed in the interviewee's reflections on the experience:

Participant 4 mentioned a couple of times that it felt like therapy, and then even an interrogation, when I thought I had probed naturally and calmly for more of a 'what do you mean' by that, to which I apologised. I think they were trying to be humorous, but I feel it made me second guess my questioning style. Now when I reflect on it, I recall that this participant digressed a lot so I think they took my push to return back to the question as 'digging'.

Chira is also made aware of how her own 'habitus': 'a structured system of dispositions formed through absorbing the history of the social field and the agent's particular trajectory within it' (Ayrton 2020:

1244; Bourdieu 1990). Our habitus can set the social limits of our understandings, unless these social limits are actively brought to light:

This participant also said about how much she loved her job and again I felt bad for being surprised as I view cleaning as a monotonous task that is a necessity. However, she did share that some cleaning tasks were more monotonous than others, such as cleaning windows which she would delegate to a robot. I think I forgot momentarily that there is more to a job than just the responsibilities itself, she may get more from it from a social aspect or managerial, as she managed a team of 29.

After the first set of interviews had been completed, the research team engaged in further reflection in terms of the 'Act' principle going forward. This provides an important opportunity for further recrafting to ensure that the principles of RRI are attained.

5.4 Act

On reflection, we realised that holding the interviews straight after participants had already engaged in other aspects of the research, meeting 'Zappy' and then completing a user survey, might have influenced their conceptualisations of robots, as well as leading to fatigue. Acting to schedule the remaining interviews for another time, and conduct these remotely proved a productive step, somewhat challenging our previous assumption that these should be done in the workplace. This enabled recrafting the framing of certain questions, for example:

Many found it hard to separate their idea of robots and their capabilities with that of Zappy who they had met. Were they skewed by seeing Zappy who is big and a bit slow? To address the latter, I knew I needed to make sure I emphasised that even though they had met Zappy to think of robots in a more general sense.

Further, and inspired by feminist interview methods of conversation and power reversal, Chira also recrafted her interview style:

There was a notable change in length, all hitting around the hour mark. I myself became more comfortable and confident with the scheduled line of questioning, not only knowing what questions were next so I could adapt the flow easily if an interviewee jumped ahead in their answers, but also learning what questions might provide more fruitful answers and therefore probing more at the right times. I think my confidence in the question guide may have created a more relaxed style of interviewing- it felt more conversational rather than a set of questions that needed to be answered. I took this feeling into all the rest of the interviews. I added more questions that I felt would add value, such as understanding the participants' thoughts on trust as a concept, trust with automation, humans and robots, as well as (before they met Zappy) what they envisioned when they thought about robots, the benefits and limitations robots had today, or in the future. I also made a point of emphasising that the participant is the expert to avoid them feeling interrogated or powerless. I wanted them to understand that what they were telling me was interesting and this encouraged them to provide richer answers.

The Act 'stage' provides the confidence to build on and perhaps recraft our methods to better meet the principles of RRI. However, importantly, the fieldnotes reveal how all the AREA principles work together in tandem, weaving through our research practice to underpin the skills of research craftwork.

6. Conclusion

This paper has shown the productivity of disrupting conventional and standardised approaches to scientific research based on episteme and, drawing on sociological and feminist insights, developing a techne approach which is more contingent, transparent, embracive, and pluralistic. We argue that the capabilities this offers is of particular relevance not only to research on Trustworthy Autonomous Systems but has general relevance, both epistemologically and methodologically, for all researchers working with, on, and for human participants, in the following key ways:

First, by drawing on feminist scholarship, we have highlighted the ways in which who we are as researchers inflect what and how we research. Continual attention to this encourages the embedding of RRI into every stage of the research process. Anticipating and reflecting upon our own biases, personal and political motivations, and limits of experience promotes consideration of the multiple ways in which we are implicated in our research. It provides the essential space and time which is needed to fully anticipate and observe the emotional, social and political responses to the phenomenon and reflect on their origins. Chira's careful recordings of her moments of 'outsider'-based unease within the empirical stage are important as instances of 'making the familiar strange'. These present valuable building blocks by which to develop better understanding of our respondents' experiences and, thereby, more responsible researcher-participant relations. They also, importantly, inform the engage and act stages where evaluation takes place on 'what works' to recraft the research design iteratively.

Second, Chira's observations of her experiences and emotions in the field remind us of the ways in which feminist concepts of positionality and power are inevitable facets of the research process. Foucault's (1980) notion of power as being shifting and circulatory also provides a useful tool by which to craft the research process as it unfolds responsively and responsibly. However, the ways in which interview talk reflects these shifting patterns of powerfulness and powerlessness is also valuable and important data. Being highly attuned to the complexities and ambiguities within these spoken interactions provides a valuable lens to crafting responsible and responsive research methods.

Third, our focus on the views of cleaners is an unusual one within research and a means by which to further responsibilise research by including the voices of people who often remain marginalised. Cleaners routinely face chronic invisibility due to overnight shifts, and/or spatial separation from coworkers and customers as well as job stigma, such as the 'dirty work' label, (Rabelo and Mahalingam 2019: 103). Our research responds to calls for greater research on such occupations to support better understanding and dignification.

Finally, and more specifically, the cleaners' responses to Zappy underscore the complexities and contingencies of human-robot relations. ANT highlights how these are highly specific to the situation, and that the emotional responses that such networks generate, such as trust and trustworthiness, involve social and relational processes which are ongoing and transient. Rather than pursuing these concepts as objects which need to be uncovered for the creation of universal laws, we argue that a more responsible and innovative approach is to shift from episteme to techne as a means by which to better understand the complexities of Trustworthy Autonomous Systems.

References

Åm, H. (2019) Limits of decentered governance in science-society policies *Journal of Responsible Innovation* 6, 2 163-178

Ayrton, R. (2020) The case for creative, visual and multimodal methods in operationalising concepts in research design: An examination of storyboarding trust issues *The Sociological Review* 69, 6: 1229-1249

Becker, S. (2010) 'Badder than "just a bunch of SPEDs": alternative schooling and student resistance top special education rhetoric" *Journal of Contemporary Ethnography* 39, 1: 60-86

Bell, C. (1978) 'Studying the locally powerful: personal reflections on a research career' in C.Bell and S. Cencel (eds) Inside the Whale Sydney: Pergamon

Benton, T. and Craib, I. (2001) *Philosophy of Social Science: the Philosophical Foundations of Social Thought* Basingstoke, Palgrave

Bothelo,F. (2021) Making others (un)equal: The social ethics of Scandinavian enclaving in Maputo, Mozambique *Critique of Anthropology* https://doi.org/10.1177/0308275X211059663

Bourdieu, B. (1990) *The logic of practice* Cambridge, Polity Press

Bourke, B. (2014) Positionality: Reflecting on the Research Process *The Qualitative Report* 19, 33: 1-9

British Cleaning Council (2017) The cleaning and support service industry: Research report 2017. http:// britishcleaningcouncil.org/wp-content/uploads/2018/04/BCC-Industry-Research-Report-2017- small.pdf

Calhoun, C., Bobko, P., Gallimore, J.J. and Lyons, J. (2019) Linking precursors of interpersonal trust to human-automation trust: An expanded typology and exploratory experiment *Journal of Trust Research* 9,1: 28-46

Coffey, A. (1999) *The Ethnographic Self: Fieldwork and the Representation of Identity* London, Sage

Connell, R.W. (1987) *Gender and Power* Cambridge, Polity Press

Costa, P. (2018) Conversing with personal digital assistants: on gender and artificial intelligence https://revistas.ucp.pt/index.php/jsta/article/view/7318

Cousin, B., Khan, S. and Mears, A. (2018) Theoretical and methodological pathways for research on elites *Socio-Economic Review* 16, 2:225-249

Criado-Perez, C. (2019) *Invisible women: Exposing data bias in a world designed for men* London, Penguin Random House

Dalton, R. and Welzel, C. eds. (2014). *The Civic Culture Transformed: From Allegiant to Assertive Citizens*. Cambridge: Cambridge University Press.

Devine, D., Gaskell, J., Jennings, W. and Stoker, G. (2020) *Exploring Trust, Mistrust and Distrust* Discussion paper for the 1st Digital Workshop of the ESRC 'TrustGov' project https://trustgov.net/working-papers accessed 261121

Edwards, R. and Holland, J. (2013) *What is qualitative interviewing?* London, Bloomsbury

Flyvbjerg, B. (2001) *Making social science matter: Why social inquiry fails and how it can succeed* Cambridge: Cambridge University Press

Forsberg, E-M, Thorstensen, E.,Casagrande, F.D., Holthe,T.,Halvorsrud, L.,Lund, A., Zouganeli, E. Is RRI a new R&I logic? A reflection from an integrated RRI project *Journal of Responsible Technology*

Foucault, M. (1980) *Power/Knowledge* Brighton, Harvester Press

Edwards, R. and Holland, J. (2020) Reviewing challenges and the future for qualitative interviewing *International Journal of Social Research Methodology* 23,5:581-592

Elliott, A. (2019) *The Culture of AI: everyday life and the digital revolution* London, Routledge

Ford, M. (2016) *The Rise of Robotics: Technology and the Threat of a Jobless Future* New York: basic Books

Guillemin, M. et al (2016) "Doing Trust": How Researchers Conceptualize and Enact Trust in Their Research Practice *Journal of Empirical Research on Human Research Ethics* 11, 4: 370-381

Halford, S. Kukarenko, N., Lotherington, A-T., and Obstfelder, A. (2015) Technical Change and the Un/Troubling of Gendered Ageing in Healthcare Work *Gender, Work and Organization* 22,5 495-509

Hardin, R. (2002) *Trust and trustworthiness* Russell Sage Foundation

Hoffman, E. (2007) 'Open-ended interviews: power and emotional labour' *Journal of Contemporary Ethnography* 36: 318: 246

Hollway, W. and Jefferson, T. (2012) Doing Qualitative Research Differently: Free Association, Narrative and the Interview Method (2nd Ed) London, Sage

Kendall, L., Chaudhuri, B. and Bhall, A. (2020) Understanding Technology as Situated Practice: Everyday use of Voice User Interfaces Among Diverse Groups in Urban India *Information Systems Frontiers* 22:585-605

Kvale, S. and Brinkmann, S. (2015) *InterViews: learning the craft of qualitative interviewing* London, Sage

Latour, B. (2005) *Reassembling the Social: An Introduction to Actor Network Theory* Oxford, Oxford University Press

Letherby, G. (2003) *Feminist Research in Theory and Practice* Buckingham: Open University Press

Letherby, G. (2015) 'Gender -Sensitive Method/ologies' in Robinson, V. & Richardson, D. (Eds) *Introducing Gender and Women's Studies* London, Palgrave

Li, P. (2012) When trust matters most: The imperatives for contextualizing trust research Journal of Trust Research 2, 2:101-106.

Kok, B.C. and Soh, H. (2020). Trust in Robots: Challenges and Opportunities *Current Robotics Reports* 1 : 297–309

Mackenzie, D. and Wajcman, J. (eds) (1999) *The Social Shaping of Technology* Basingstoke, OUP

Mikecz, R. (2012) 'Interviewing elites: addressing methodological issues' *Qualitative Inquiry* 18: 482-493

Mills, C.W. (1959) *The Sociological Imagination*

Möllering, G. (2001) The nature of trust: From Georg Simmel to a theory of expectation, interpretation and suspension *Sociology,* 35, 2: 403-420

O'Reilly, K. (2012) *Ethnographic Methods* (2nd Ed) London, Routledge

Orgad, S. and Higgins, K. (2021) Sensing the (in)visible: domestic cleaning and cleaners on Mumsnet Talk *Feminist Media Studies* DOI: 10.1080/14680777.2021.1922486

Pickering, M. (2001) *Stereotyping: The Politics of Representation* Basingstoke, Palgrave

Pink, S. (2009) *Doing Sensory Ethnography* London, Sage

Prasad, P. (2005) *Crafting Qualitative Research: Working in the Postpositivist Traditions* New York, M.E.Sharpe

Rabelo, V. and Mahalingam (2019) "They really don't want to see us": How cleaners experience invisible 'dirty' work *Journal of Vocational Behaviour* 113, 103-114

Rousseau, D., Sitkin, S., Burt, R. and Camerer, C. (1998) Not so different after all: A cross-discipline view of trust. *Academy of Management Review* 23, 3: 393-404.

Poortman A-R and Van der Lippe, T. (2009) Attitudes Toward Housework and Child Care and the Gendered Division of Labor *Journal of Marriage and Family* 71, 3: 526-541

Sovacool B.K. et al (2020) Sociotechnical agendas: Reviewing future directions for energy and climate change research *Energy Research and Social Science* 70

Spencer, J. (2001) 'Ethnography after postmodernism' in P. Atkinson, A. Coffey, S. Delamont, J. Lofland and L. Lofland (eds) *Handbook of Ethnography* London, Sage.

Stanley, L. and Wise, S. (2008) 'Feminist Methodology Matters!' in D. Richardson and V. Robinson (Eds) *Introducing Gender and Women's Studies* 3rd ed. Basingstoke, Palgrave.

Steup, M. and Neta, R. (2020) 'Epistemology' *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/epistemology/#VariCognSucc

Wang, Y. and Young, J. (2014) Beyond "Pink" and "Blue" : Gendered Attitudes towards Robots in Society *genderIT* May 7-9

Watson, N., Luff, P. Fischer, J., Leonard, P., Dasgupta, P., Evers, C., Hyde, R., Kucukyilmaz, A., Price, D., Trigo, M. (2020) *Trustworthy Human Robot Teams* TAS Hub Agile Project Proposal: PART B V1, 17 Nov

Weber, M. (1968) *Economy and Society* New York, Bedminster Press

[i] (https://epsrc.ukri.org/research/framework)