„IMAGINE Responsible Robotics“

A Card-Based Engagement Method

Authors (main contribution):

Lisa Sigl^{1, 2} (conceptualisation of the method & writing of card texts; editorial responsibility of card texts; writing of and editorial responsibility for manuscript)

Jesse de Pagter¹ (conceptualisation of the method & writing of card texts; design of the cards; substantial comments on earlier version of the manuscript)

Guglielmo Papagni¹ (conceptualisation of the method & writing of card texts; substantial comments on earlier version of the manuscript)

¹Technical University of Vienna

² University of Vienna

Abstract:

In this text we introduce a card-based engagement method we call “IMAGINE Responsible Robotics”. The basic purpose of this method is to facilitate reflection and exchange about societally relevant issues in research, development and design (RDD) of robotic technologies, particularly also in increasingly inter- and transdisciplinary contexts of robotics.

A basic precondition for such an exchange and collective reflection is mutual understanding. Before this background, we have developed a card-based engagement setting that in a first step(,) facilitates the exchange and mutual understanding on different value-laden perspectives on robotic technologies, in a second step the application of these perspectives to a specific context of application and, in a final step, the co-development of situated ways of practicing responsible and – from a societal perspective – trustworthy robot development. Ideally, in this way new ideas on how to consider societal issues and concerns in robot RDD are generated.

The method is particularly designed to be a resource for idea generation amongst (academic and other) researchers, developers and designers at relatively early stages of research endeavours. It builds on the sense that reflections on the societal responsibility involved in research and innovation can open up often black-boxed assumptions about the societal relevance of robotics and thus be an ideation resource in creative innovation processes.

Since the development of robotic technologies is often a multi-stakeholder-process, the last phase is particularly designed to reflect on and gather ideas on what kinds of stakeholders would ideally need to be included and what kinds of collaborations would make sense for a specific research idea to be specified and carried out.

We argue that while such tool-based methods can easily be misused as one-point interventions with little effect on actual research practices, they can be of great value when applied in the context of comprehensive approaches to robot RDD such as care-centred value-sensitive design, participatory design or integrative social robotics (ISR).

Card design and discussion map are available for download in the supplementary material!

1. INTRODUCTION

"Robotics Technology will become dominant in the coming decade. It will influence every aspect of work and home. Robotics has the potential to transform lives and work practices, raise efficiency and safety levels, provide enhanced levels of service and create jobs. Its impact will grow over time as will the interaction between robots and people..."

European society is currently facing important challenges. Robotics can be an integral part of wider solutions to these challenges, but using them will have important ELS impacts. Addressing these impacts needs to go hand in hand with the deployment of technology."*
(SPARC-The Partnership for Robotics in Europe 2017)

**ethical, legal, and social*

As in this introductory quote, robotic technologies are expected to have wide application areas and the potential to change many areas of our lives, e.g. by re-organising work life with collaborative robots and smart factories or by using robots in care contexts or as social companions. In the above quote, the Strategic Research Agenda for Robotics in Europe even states that robotic technologies *will* change "every aspect of work and home" (SPARC 2017). Because of these potentially pervasive influences on our lives, it is widely argued that we need a societal debate on the ethical, legal and social aspects (in the following referred to as societal issues) associated with robotic technologies. This for example includes reflections on controversially discussed questions such as how to deal with personal data and privacy issues, how to provide physical and emotional safety or how to deal with potentially new kinds of accountability in the area of robotic technologies and artificial intelligence (European Group on Ethics in Science and New Technologies 2018, Austrian Council on Robotics and Artificial Intelligence 2018, Tzafestas 2018, Taylor et al. 2018, European Commission 2020).

But what exactly does this mean in practice? As mentioned in the above quote, a popular argument is, that it is important to address these issues "with the deployment of technology" (SPARC 2017), i.e. when the technology is already developed and ready for use. This kind of "downstream" consideration however only has limited effect on shaping the ethical, legal and social character of the technologies themselves and near to no effect on the value-laden decisions taken within respective research, development and design (RDD). Thus there is a long-standing debate about how such issues could be also considered "upstream", for example in a public deliberation on which kinds of RDD should be commissioned and funded and "midstream", that is within in the many strategic decisions taken within RDD processes (Fisher et al. 2006).

While downstream-engagement (regulation/decision-making after technology development) is certainly also core to the development towards responsible RDD, it has amongst others been criticised because of possible path dependencies, i.e. that once technology development has embarked on a path that usually includes certain technological and development choices it is difficult to change them and thus de-legitimise investments that have already been made.

Metaphors for robot development like that of the "roadmap" (SPARC 2017) suggest that there is mainly one way to go. Giving more attention to the midstream phase however allows making visible and also put to public scrutiny that in technology development there are always different possible paths to choose in technology development. The card-based method we propose here is located in this midstream: it is designed to provoke reflections about ELS aspects, make visible the tacit and often unreflected issues and value-laden decisions in technology development, and, in doing so, enabling participants to deal with these issues in more deliberate ways. It is aimed at groups of (often interdisciplinary) researchers, designer and/or developers that plan collaborative work. Ideally, it is used in rather early and preparatory phases of research endeavours or pre-engagement activities (cp. te Kulve/Rip 2011). As we will argue below, rather than being used in isolation, we envision this method to be used as part of comprehensive approaches to robot RDD such as value-sensitive or

participatory design approaches and in recent, innovative approaches in robotics research, such as integrative social robotics (ISR) and critical robotics research.¹

2. RESPONSIBLE RESEARCH AND INNOVATION IN ROBOTICS

A substantive body of work (including on technology assessment, anticipatory governance and ethical, legal and societal aspects of research) has shown that following unreflected path dependencies results not only in delegitimised early investments, but also in moral luck, meaning that the ELS implications of so developed technologies are random or determined by other factors than ethical reasoning (Williams 1981). Thus, it has been argued that early-on reflection on these kinds of decision-making processes are necessary, e.g. to anticipate potential risks or societal concerns (e.g. based on the development of other technologies), and to develop settings in which relevant stakeholders can deliberate on which path to go and what kinds of collaborations throughout a research process can support societally desirable RDD. Such reflections are thought of as staying important throughout the whole RDD process because of unpredictabilities regarding risks and opportunities throughout RDD processes. Some ways of dealing with such issues have been developed in different approaches to ethical RDD like value-sensitive or participatory design (see also below). The policy framework “Responsible Research and Innovation (RRI)” circumscribes four dimensions of societally responsible research and innovation processes: ongoing **reflection** about the course of research and innovation processes, **anticipation** of societal implications (as far as possible), **deliberation** on ethical and value-related aspects and **responsiveness**, i.e. the flexibility to adapt to unforeseen problems, but also potentials. These four aspects are thought to enable researchers, other stakeholders, but also society at large to deal consciously with technology-related risks and challenges, but also to unfold their potential for desirable societal impact (Owen et al. 2013). In other words, such processes aim at moving beyond a “closing down” framing of conventional ethical review and approval” and aims at process-oriented ways of opening up new ideas for realising the public value of science and innovation (ibid).

The engagement method we propose here particularly builds on the “opening-up” aspect of reflection and deliberation, in the sense that individual reflection on and collective discussion of societal issues are used as an ideation resource in creative innovation processes.

One challenge in such engagement settings is to make visible the value-laden aspects in RDD processes, since studies have indicated that throughout academic socialisation processes, researchers sometimes learn to orient their research rather towards what is best publishable and fundable. This can make them “forget” about considering broader societal issues along the progression of their academic careers (Fochler et al. 2016). Since particularly broader societal values often stay unconscious and are deeply hidden within RDD practices and concepts, a strand of work within RRI-related work focuses on ways of facilitating open debates within the research community (Felt/Fochler/Sigl 2018), but also between the research community and other stakeholders (developers, broader publics) that make value-laden practices more visible and debate-able.

Within the robotics field, we see discussions along similar trajectories. Two particularly visible approaches are Participatory Design (PD) and Value-Sensitive Design (VSD) approaches. Šabanović, for example, has argued that in many narratives, technological development in robotics is “led by experts from academia, industry and government”, figuring them “as the primary driver of social progress, while society fills a passive role of accepting and adapting to the results of technological innovation.” (Šabanović, 2010, p. 439). Instead, she argues for participatory design processes that put robotics RDD within a “mutual shaping framework” and allow societal actors and values to play a more decisive role in technology development. CCVSD approaches are also process-oriented in that they propose different components and a step-by-step process of considering

¹ The method was developed in the framework of the „REsponsible Robotics“ pilot project at the Vienna University of Technology: http://trustrobots.acin.tuwien.ac.at/?page_id=2069 (14.09.2021).

sensitive issues “from the moment of idea generation” and continuing “throughout the design of various prototypes” (van Wynsberghe 2013).

Further, approaches like “Integrated Social Robotics (ISR)” and “Critical Robotics Research (CRR)” emerge that - amongst others – call for more interdisciplinary collaboration and a value-based debate on what robotic technologies should do rather than a utility-oriented RDD on what robotic technologies could do (Seibt 2016, Seibt et al. 2020). This approach is so far most articulated for social robotics but is beginning to broaden towards other robotics fields (Seibt-Special Issue in AI 2021??).

These approaches have in common the understanding that RRI cannot be just another box to tick. Rather, it requires a “wider engagement on social and ethical dimensions... [that] cannot be externally imposed... [but] must be nurtured by scientists... themselves” (Wilsdon, Wynne, Stilgoe 2005: 19, cf. Jasanoff 2007). To be effective on the epistemic level, i.e. in shaping direction and contents of research, reflection must become an embodied skill, part of the tacit knowledge of designing and carrying out complex and often unpredictable research processes. We thus aim for a method that allows researchers, designers and developers to access and engage with societal issues in robotics in a time-efficient way.

3. CARD-BASED METHODS

There is a growing body of work dedicated to developing methods and tools that can facilitate reflection of societal issues: workshop formats like the “mind-scripting workshop” by Allhutter (2012) or “IMAGINE RRI” (Felt/Fochler/Sigl 2018) and the recent EU project *RRI tools* that has developed a self-reflection tool and provides a collection of tools for facilitating responsible research, education or design purposes may serve as an example. Particularly the latter often also follow the aim of supporting inter- and transdisciplinary exchange (Groves 2017). A body of work is also emerging linked to facilitating the co-production of knowledge in inter- and transdisciplinary contexts (https://naturwissenschaften.ch/co-producing-knowledge-explained/methods/td-net_toolbox/cite_methods).

Within such methods, cards are often used as tools for addressing complex and dynamically changing challenges, such as in engineering education to teach students to deal with wicked problems in everyday work situations. Since engineers face complex challenges in their ethical choices (Emison 2006), card-based methods have also been used to teach practitioners how to consider issues of responsibility in their work. Linking to a common practice in design communities to use “ideation cards” to stimulate creative thinking, cards were used to reflect upon options of Privacy by Design (PbD). While PbD is increasingly also becoming a legal requirement, this participatory approach builds on the understanding that privacy is best realised when it is a guiding value during the whole design process, thus aiming to shift participants’ mind-sets from compliance-to-regulation to “ownership and buy-in” (Luger et al. 2015). To achieve this, the cards make the often complex and incomprehensible legal requirements more accessible to practitioners, making the card game suitable for educational purposes as well as for being used in concrete design practices. Further, there are card-based engineering “design toolkits” that “are ‘playable’ board or card games” and let students experiment with design decisions in complex, but realistic situations that derive in dynamic work processes (He/Adar 2016). They are designed to allow students to link their learning experience to concrete practice contexts and often include both, individual and collective working phases to avoid the workshops to be dominated by just one leading participant (ibid).

“Envisioning” or “elicitation” cards have also been used in value-sensitive design (VSD), user-centered (UCD) and participatory design (PD) (Beck et al. 2008; Friedman et al. 2011; Friedman/Hendry 2019; Schwaninger et al. 2021). Schwaninger et al. for example use visualisation techniques on the cards to support elicitation, reduce power imbalances between researchers and participant and “enhance participants’ ability to elaborate on their own conceptions of the world” in a playful way (2021). In this way, a “shared understanding of trust” is established, which is described as “important and should be created in early design phases to support the

proper transferal of knowledge and more sensitive design decisions” in participatory design processes for social robots (ibid).

Card-based approaches have also been used in public engagement activities to facilitate reflection about the values associated with emerging technologies in card-based public engagement (e.g. Felt et al. 2014) or as a tool for studying responsible research (Felt et al. 2018; IMAGINE RRI). “Dilemma games” (e.g. on “professionalism and research integrity”²) provide – in a game-like manner – a repertoire of values or value-laden decisions as teaser for participants’ statements and facilitator for reflection. A recent example for an ethics by design tool is the “Moral IT-Deck” used for reflecting normative issues related to technology development” (Urquhart/Craigon 2021). These are just a few examples of how cards are currently a frequently used tool for supporting reflection and to support collective ideation processes (Peters et al. 2020).

One challenge addressed by such card-based approaches that is particularly important for the “Imagine Responsible Robotics” method we propose here, is the problem of inter- and transdisciplinary communication. As has often been emphasised, robotics RDD and responsible robotics in particular deals with very complex issues and thus requires “strong, mutual, and long-lasting interdisciplinary collaboration” (Hannibal/Weiss 2020:1-2). However, even though interdisciplinarity has been seen as an important approach to finding new and innovative problem solutions, there are still challenges involved in practicing interdisciplinary exchange, integrating knowledge from different disciplines (ibid), and in developing shared epistemic commitments towards research (Granjou/Arpin 2015).

Moreover, in many inter- (and also trans-) disciplinary settings, there is a degree of incommensurability in respective disciplinary definitions, concepts and reference points. For such settings it has been argued that despite such incommensurability, exchange can be facilitated if groups agree on rules of exchange and “can hammer out a local coordination, despite vast global differences” (Galison, cit. Collins/Evans/Gorman 2007). Creating such spaces – sometimes called “trading zones” – then allows for better mutual understanding and respectful co-production of new lines of thinking. For robotics research particularly, this need for inter- and transdisciplinary collaboration has amongst others been highlighted by Selma Šabanović (see e.g. Šabanović 2010) who showed the potential of humanities in robotics, and called for interdisciplinary collaboration in robot RDD.

4. THE CARD-BASED ENGAGEMENT SETTING „IMAGINE RESPONSIBLE ROBOTICS“

The card-based engagement setting that we introduce here addresses this problem of communication. It aims at opening a debate about often subjective value-based attitudes towards responsible conduct of research and innovation and allows participants to reflect on their personal value orientations in relation to robotic technologies, while at the same time also potentially allowing to relate one’s professional expertise to these reflections. Hence, the text on the cards points to publicly discussed and often controversial societal issues and is designed to provoke reflection. While it is designed to be accessible to colleague scientists, developers or designers, it partly uses professional language to emphasise that the addressed issues are issues of public debate, but are also issues of scientific research, tackled by research fields like human-computer-interaction, philosophy of technology or science and technology studies. The overall setting is focused on reflection, i.e. the aim is not to teach researchers what responsible robotics is but to provide a repertoire of perspectives and questions to reflect upon and to provide a moderated space for exchange.

The method builds on the reasoning that new and emerging technologies like robotic technologies always have the potential to develop in different directions, depending on which aims are pursued, what values guide RDD and what issues are considered in bringing them into being. To facilitate a process in which technology development happens in ways that are societally desirable, it requires that researchers and developers are

² <https://www.eur.nl/en/media/2020-12-original-dilemma-card-game> (13.08.2021)

individually aware and self-reflective of societal issues and values related to robotic technologies, that they are able to apply different societal perspectives to the particular application contexts they work on, and to be able to envision and design RDD processes that are embedded in multi-stakeholder settings. Using the language of science and technology studies we could say that what we need are methods of transforming “matters of fact” into “matters of care”, i.e. of developing ways of acting upon societal issues and concerns in robotics research (Latour 2004, de la Bellacasa 2017).

The proposed engagement method allows participants to individually reflect those societal issues but also to exchange with others on their situated perspectives and knowledge. It is designed for discussions of app. 5-8 participants and 1-2 facilitators who moderate the discussion and take care of making notes and recording (if required). Possible participants are robotics researchers and other stakeholders in the design and development of robotic technologies. A discussion of 3 hours (incl. a 15min break) is envisaged. The engagement is card-based, meaning that cards provide a narrative infrastructure and haptic element for the debate (cf. Felt et al. 2018), and flexible enough to find their own way of addressing societal issues and concerns. The role of the moderator(s) here is that of (a) facilitator(s), meaning that they guide through the steps, make a list of speakers (if necessary), care for timing and see to it that everyone gets to voice his/her opinion.

It consists of three steps of discussion: (1) A first step, supported by 11 “Perspective Cards”, providing short introductions and core questions on value-laden, societally relevant issues with regard to robotic technologies, (2) a second step to relate a set of perspectives to specific application areas, and (3) a third step that facilitates an open and creative debate on possible practices and ideas/options for making these issues matter in robot research and development. Since the field of robotics is huge, each discussion setting focuses on one field of application. In the current version, cards related to companionship robots and manufacturing robots are included, but other fields of application would easily be added to the card set. To facilitate concentrated and productive discussions, discussion organisers choose only one of these contexts and invite respective participants who are stakeholders in this particular context of application (either Companionship or Manufacturing).

Moving along these three steps, the engagement method thus allows facilitating debates and exchange about values that guide robotics research, but also goes beyond this towards the concrete imagination of practices that allow considering societally relevant issues, in particular application contexts for robotic technologies. The cards however should mainly serve the purpose(s) of providing shared narrative references and language for exchange within a heterogeneous group of participants. The aim of the moderation is to move the discussion beyond this frame towards a meaningful exchange of perspectives between participants. This is structurally encouraged by providing empty cards to add additional perspectives or contexts to the discussion. In the following, we give a rough overview of the three steps of the discussion. A more detailed outline of the discussion process and moderation are provided in the Moderation Outline for “IMAGINE Responsible Robotics”.

1st Step of the Discussion: “Perspectives”

For the first stage every participant gets a set of 13 “Perspectives” cards and a board for ordering the cards according to what is most / least relevant to individual participants. The cards address societal issues related to research, design and development of robotic technologies. The “perspectives” are chosen to touch upon rather general challenges to a broad set of robotic technologies and include: democracy and participation, emotional safety, physical safety, transparency and security, rule of law and accountability, responsible design, distribution of autonomy, human dignity, sustainability, wealth distribution, personal data and privacy. The cards each describe the perspective very briefly with one paragraph and a few questions. The cards are meant as a means for opening perspectives and providing key aspects of societal issues that are

publicly and scientifically discussed. Participants with specific expertise in some of these issues will find the text oversimplifying, but the set includes 2 empty cards in case participants want to add a perspective or to extend perspectives. Therewith, participants are encouraged to think also beyond what is written on the cards.

The main inspiration for choosing the topics of the 11 Perspective cards were reports by international expert groups, such as the *European Group on Ethics in Science and New Technologies'* Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems (2018), a report on a "consultation with multidisciplinary experts" by *Next Generation Internet* (Taylor et al. 2018), the White Paper of the Austrian Council on Robotics and Artificial Intelligence (2018), and a White Paper on Artificial Intelligence by the European Commission (2020). Cards' text was also grounded by current discussions in the field of science and technology studies, science and engineering ethics, philosophy of science and debates within the Doctoral School "Trust Robots" at the Technical University of Vienna. Please see supplementary material for additional literature sources of the individual cards.

During a first step, participants are asked to read through the cards (15min, reading 1 card quickly takes app. 40sec) and prioritise the "Perspective" cards according to individual relevance, i.e. arranging the cards on a board according to what they find most/least relevant for themselves and their work area (see Image 1). While it can be assumed that there is a large overlap between personal relevance and relevance to work context, we ask them to choose personal relevance, as by "relevance" here we mean a subjective valuation that results from one's own personal reflection. This step is meant to allow participants to gain a sense of ownership over the issues at stake, which is considered an important condition for in-depth discussions and creativity in the later stages of the discussion. Particularly this first discussion step is designed to tap into the affective relation people have to societal issues and concerns in robot RDD.

Figure 1: Possible Orderings of Perspective Cards (Clustering or Spectrum along Relevance Priority)

While reading through the perspectives cards, participants thus lay down either a spectrum or a clustering according to relevance, but they may also find a different way of ordering the cards that allows them better to think and reflect about them. After this ordering, the moderator facilitates a round in which everyone explains their individual ordering. This will help to get to know where the other participants come from and what kinds of issues matter to people in this specific setting.

2nd Step of the Discussion: “Contexts”

The **second step of the discussion** is about relating the Perspectives to specific fields of application. In this step, a new type of **“Context” cards is introduced** that describe sub-contexts of application in which robotic technologies are either already applied or are likely to become applied in the future. In the supplementary material we provide cards for the broad area of robotic technologies for *Companionship* (with the sub-contexts Assistance, Care, Education, Friendship, Intimacy) and for *Manufacturing* (with the sub-contexts Collaborative Robots, Wearables, Smart Production, Industrial Robots). However, these sub-contexts can also be defined by the respective group of researchers, developers and designers and be written on empty context cards. The aim in this step is that participants choose collectively 1-2 sub-fields of application and reflect on which of the “perspectives” from step one are most important to consider here. In practical terms, the respective Context-Cards are placed in the middle of the table (or, for example, a large white board) to make the focus of the current discussion visible on the table, and participants describe in a round, which “perspectives” they are most concerned with in relation to it. While doing so, they lay down the respective perspective cards (ideally, max. 3 cards) to the respective context card.

After this round, the group looks at the picture that has emerged and chooses 1-2 contexts that they want to focus on in the 3rd discussion step. Participants are free to decide on the rationale of this choice; it may for example be inspired by which context(s) most people are working on – or are interested to work on in the future, by which perspectives they find most interesting to discuss with the other people in the group, or they may want to discuss contexts/perspectives that are often marginalised in discussions, or of which you find that they are discussed in a controversial manner because they touch upon particularly sensitive issues. Usually, it is advisable to take a 15min break now for informal discussions.

3rd Discussion step “Practices and Collaborations”

This last step in the discussion is dedicated to discussing how robot development can take place in a trustworthy way; i.e. in a way that societies / the people can trust that societal, and partly controversial issues are reflected on and considered in the practice of robot research, design and development. In this step, participants who already have good experiences share their best practices and, together, all participants can discuss and develop new, creative and also speculative ideas on how to consider societal issues and concerns in robotic RDD.

Guiding questions for the discussion in this third step are: What could be interesting research questions if we consider the societal issues at stake? What would it mean to organise research in this particular context in a responsible way, i.e. with the aim of increasing the likelihood that robotics applications will have a positive impact? Who would have to be included in the research process to ensure that relevant societal perspectives are considered?

This last step is rather characterised by open debates and utopian brainstorming, meaning that participants should not feel restricted to thinking about the actual “doability” of a particular research process but think about what an ideal research endeavour in this context would look like.

This step can be done collectively in the large group or in smaller groups forming around one particular sub-context. Participants are provided with lapboards (small whiteboards, pens and eraser(s) or, alternatively, just paper sheets to make notes), on which they can first do collective brainstorming on their ideas. Ideas are envisioned to include for example new kinds of research questions, ideas for applications, structures for collaboration or participatory research, etc.. Important in this step is that participants can really tap into their creativity, i.e. notes on ideas should be taken regardless of how realistic they may seem in practice, either by the participants or the moderator him/herself.

The last 15 minutes of the discussion are then reserved for talking about what kinds of ideas participants find potentially worthy of being followed-up for future research endeavours and what their key take-away reflections are. As a documentation of the workshop, pictures from the taken notes and brainstorming should be shared via e-mail amongst participants as a resource for potential follow-up work.

5. EXPERIENCES FROM USING THE METHOD IN THE CONTEXT OF AN "IDEATION WORKSHOP"

In this section we briefly summarise experiences of using this engagement method in an ideation workshop we organised for a research group. The group was in a stage in which it was just about finalising some large research projects and thinking about carving out new research trajectories. The workshop was organised along the lines described above. Since in this particular setting the workshop was about imagining new sub-contexts for application, the second step included an open brainstorming of possible sub-contexts. The following is based on an analysis of the workshop transcript (excl. small group discussions in Step 3), as well as immediate feedback by the eight participants on the workshop format, given within the workshop itself and anonymous written feedback by four of the participants a few days after the workshop.³

It was experienced as important aspect of opening up reflection and exchange that the method did not have an explicit aim or output expectation, as this participant put it.

„[It was] good and important that there was no ‚goal‘, but that we could rather just spin. ... The ... cards were a good visual tool to open up thinking about important technology implications.“ (P1)⁴

The cards were seen as a good inspiration for “thoughts that you would otherwise not have had” (P2). Even though the amount of text on the cards was at the beginning experienced as a bit intimidating, the pointed questions on the cards amongst others were seen as helpful for thinking about participants personal relation to and prioritisation of the addressed issues (RGP-1285; RP2).

In particular, phase 1 was experienced as preparing the individual participants for the collective ideation process. In a brief exchange amongst the participants about the usefulness of the workshop, it was seen as a “good start... to get your brain prepared” (RGP-1285). The cards were experienced as a tool for open individual reflection, as supporting and provoking further thinking with giving “different perspectives on the issue” (FP4) and a “multitude of considerations” (RGP-1290), but not as edifying or framing too much. One participant said it was helpful that “for a short time, everyone can have thoughts in his/her own way...the cards are there, but the thoughts are your own” (RGP-1290). Another participant then followed up on this by saying: “I think that it makes you aware of what you think yourself” (RGP-1295).

Overall, the setting was also seen as useful for exchange on controversial societal issues with colleagues. The open approach as well as seeing and comparing one’s own prioritisation of cards with those of colleagues seems to have provoked discussion. A participant for example said that this was due to a “field of tension that emerges between the terms and between the interpretation [of the terms] by participants” (FP2). At least in this group, the setting also seems to have been useful for reflecting and exchanging on issues that are usually considered less relevant. An interesting dynamic of the discussion was that one card was rather consistently seen as little relevant or even “not relevant at all” (RGP-147) by many participants. Importantly, they didn’t see it as unimportant (RGP-337), but saw it as more urgent to talk about issues that are legally or otherwise regulate-able. One participant said for example that „this is the reason why [this card] is further at the back, because I think that it would be beyond the scope and that rather... there needs to be a reduced set of rules first.“ (RGP-38; cp. -315) The issue on the card seemed to be “less graspable” (RGP-80; -312). Interestingly, the

³In the following we use pseudonyms to refer to participants: RGP[-line number] for „random group participant“; FP[number] for „feedback participant“

⁴German original quote: „Gut und wichtig: Dass es kein "Ziel" gab, sondern man auch einfach spinnen durfte. ... Die Themenkarten waren ein gutes visuelles Tool um das Nachdenken über wichtige Technologie- Implikationen zu öffnen.“

following discussion then showed that the card-based setting provoked a quite lengthy discussion about exactly this issue. This dynamic was seen as remarkable and surprising within the group, both during the discussion and (in) the feedback (e.g. P2). It led participants to tap into their experiences and societal debates about robotics and reflect upon them in relation to their work (RGP-298f). This suggests that the setting was able to facilitate reflection and exchange beyond the kinds of issues that are usually seen as relevant in a given context. One participant saying that “usually the conversations relate a lot to very concrete projects and therewith strongly to technology and specific concepts” (FP3) The “open design of the workshop” (FP4) as discussed above, was seen as important for this.

The „external moderation“ was seen as contributing to this open approach, as it prevented the discussion from implicitly narrowing down the imagination to questions and problems that seem doable in a particular work context (FP1). One participant put it in his feedback like this:

„the moderation helped going beyond old thinking bubbles. Even though many ideas were already present internally [in the group], it was nice to see the old ideas in a new light and to generate new ones. If there would have been a goal beforehand, I think there would have been less wild, but because of that also [less] exciting ideas.“ (FP2)⁵

In this sense, the workshop seems to have served the purpose of ideation quite well.

It was noted however that participants would have liked to follow up on this open ideation process with a workshop phase that facilitated breaking down the generated ideas to more specific ideas for new kinds of RDD, because “a lot of thinking work is already done at this point” (FP4) and at this point, “you are mentally within the experiment, then you get started” (RGP-1281). For following up the three steps of the discussion that are described above, participants also had quite specific ideas on how to bridge to a step that would specify the generated ideas. One participant suggested the following:

“Towards the end, when the clustering [of ideas] is on the wall, one option would be to offer the group to specify a bit. For example, each group has the aim to formulate just one sentence that describes its vision of the topic. It could also be a sketch... [this] could help specifying and remembering the idea better.“ (FP2)⁶

Other suggestions for specifying the ideas more were to “generate just ONE future scenario” (FP1). Depending on the context-specific aim for such an engagement though, it could be useful to add a further step for specifying ideas and widen the workshop to 4 hours or even a whole day workshop to further follow-up on promising ideas.

6. DISCUSSION AND OUTLOOK

The card-based setting we describe here is designed to help reflections on and explorations of how societal issues and concerns can be addressed and considered in robotics RDD, particularly within interdisciplinary groups of researchers, developers and designers. Like other card-based approaches it can be helpful in unpacking quite complex issues “through channelling attention” (Urquhart/Craigon 2021; cf. Kitzinger, 1994). Rather than being used as providers of information, the cards in this engagement method are used as a tool to make the knowledge of the participants accessible to other participants. Using the narratives on the cards as “anchors” (Urquhart/Craigon 2021) or “reference points” for a collective discussion (Felt/Fochler/Sigl 2018)

⁵ German original quote: „Ja, da Moderation hat sehr dabei geholfen die alten Gedankenblasen zu überwinden. Auch wenn schon viele Ideen intern vorhanden sind, war es schön die alten Ideen in neuem Licht zu sehen und neue zu generieren. Hätte es am Anfang eine Zielsetzung gegeben wären wohl weniger wilde aber auch deshalb spannende Ideen aufgekommen.“

⁶ German original quote: Gegen Ende - wenn die Cluster an der Wand sind - könnte man anbieten ob die Gruppe nun weiter konkretisieren wolle. Z.b. jedes Grüppchen setzt sich zum Ziel einen einzigen Satz zu formulieren der ihre Vision zum Thema beschreibt. Kann auch eine Skizze sein. ...[das] könnte helfen das Thema zu konkretisieren und sich besser an die Ideen zu erinnern.“

allows participants to better understand reference points and value structures of other participants in relation to the discussed issues. This allows participants to think beyond what each participant could think of individually.

It is important to emphasise that the here proposed method should be used in the context of comprehensive, collective and process-oriented approaches like PD, CCVSD, ISR, or CRR (see above). Ideally it is used in early stages of idea generation, to inform the trajectory and forms of collaboration and participation in subsequent RDD steps. It could for example be used for initial discussions between researchers, developers and designers to prepare participatory design settings. A power-sensitive moderation of a discussion along the three stages proposed here could help participants from different fields, institutional contexts and partly different epistemic authority “to speak up and engage in a collective discussion” (cf. Felt et al. 2018: 207).

The card-based discussion is potentially also compatible with an ISR approach because it considers the five principles: According to the “process-principle” (1), the products of RDD processes are not tools/robots but social interactions, as “they are, in effect, engineering socio-cultural interactions” (Fischer et al. 2020: 149). The “perspectives” in the here presented card-based method shift attention to the different dimensions of these socio-cultural interactions and societal aspects (such as emotional safety, physical safety, or distribution of autonomy). Due to this method being designed for exchange between different stakeholders, it brings together expertise from many perspectives, thus can be useful in following parts of the “quality principle” (2) of involving all relevant scientific disciplines. The third “ontological complexity principle” (3), states that the perspectives of I, you and socio-cultural norms need to be considered, because “the ‘truth’ of a socio-cultural interaction is always a matter of procedural, dialogical mediation between these three perspectives” (Fischer et al. 2020: 150). The proposed method invites participants at different points during the debate to shift between these perspectives: during the first stage, they read perspective cards (that represent societal concerns and different socio-cultural “perspectives” that can be applied to reflecting societal issues with regard to robotic RDD), and are then invited to self-reflect, which of these perspectives are important to themselves (“I” perspective). The rounds in which participants explain their positions (and other participants listen), as well as the collective discussions, particularly in stage 1 and 3, then invite participants to consider the perspectives of other participants. Besides that, different cultural norms can come in by including respective people to participate in the discussions. The fourth “context principle” (4) states that socio-cultural interactions are always context-specific, which is why the discussion format foresees an application of the perspectives to specific sub-contexts. And finally, the “values first principle” (5) is compatible with the method, as the text and the questions on the cards are meant to support a debate on which values (should) guide robot development. For example, the perspective card ‘Democracy and Participation’ raises the issue of who – and whose values – should be considered in decisions in robot RDD. The perspective card ‘Emotional Safety’ raises the question whether or not robots **should** perform certain tasks (e.g. such that involve emotional bonds) even if they **can** perform them, and the perspective card ‘Responsible Design’ explicitly introduces the distinction “between what the robot can do from what the robot should do”.

Within such comprehensive approaches, an engagement method like “IMAGINE Responsible robotics” can be of great value as it is time-efficient and relatively easy to accommodate within academic cultures in which time for reflection is scarce and opportunities for exchange beyond the immediate work tasks are often rare. Before this background such methods have the important function of creating a concentrated time and space for reflection and of catalysing and speeding up exchange.

Acknowledgements:

Special thanks go to Helena Frijns (Technical University of Vienna) for contributing to the conceptualisation of the method, the writing of card texts and substantial feedback on an earlier version of the manuscript. The authors would like to thank all students of the PhD school Trust Robots at the TU Wien whose reflections in the class “Responsible Robotics” brought

to the fore issues and perspectives that informed the development of this engagement method. We also thank Sabine Köszegi and Markus Vincze for making the concern about “Responsible Robotics” a part of this PhD school. Further, we thank Vice Rector Kurt Matyas for supporting the pilot project “Responsible Robotics” and Marjo Rauhala for continuous kind support and feedback throughout the project. We also thank Rumena Trendafilova (SKIAT) for the beautiful and professional design of the cards. Finally, our thanks go to our colleagues Florentine Frantz, Anna Pillinger and Wolfie Christl (Cracked Labs) for feedback on an earlier version of the cards and/or the discussion setting. Last but not least, we want to thank the researchers who took part in the engagement workshop and provided valuable and constructive feedback.

References:

- Austrian Council on Robotics and Artificial Intelligence. (2018). Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz positiv gestalten. White Paper des Österreichischen Rats für Robotik und Künstliche Intelligenz. Vienna.
- Beck E, Obrist M, Bernhaupt R, Tscheligi M (2008) Instant card technique: how and why to apply in user-centered design. In: Proceedings of the tenth anniversary conference on participatory design 2008, PDC '08. Indiana University, Indianapolis, IN, USA, pp 162–165. <http://dl.acm.org/citation.cfm?id=1795234.1795261>
- Delgado, Ana, & Åm, Heidrun. (2018). Experiments in interdisciplinarity: Responsible research and innovation and the public good. *PLoS Biol*, 16(3), e2003921. doi: 10.1371/journal.pbio.2003921
- European Group on Ethics in Science and New Technologies. (2018). Artificial Intelligence, Robotics and ‘Autonomous’ Systems. Brussels: European Commission - Directorate-General for Research and Innovation.
- European Group on Ethics in Science and New Technologies. (2018). Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems. European Commission, Brussels.
- European Commission. (2020). White Paper on Artificial Intelligence. A European approach to excellence and trust (Vol. COM(2020) 65 final). Brussels.
- Emison, Gerald Andrews. (2006). The complex challenges of ethical choices by engineers in public service. *Science and Engineering Ethics*, 12(2), 233-244. doi: 10.1007/s11948-006-0023-
- Felt, Ulrike, Schumann, Simone, Schwarz, Claudia G., & Strassnig, Michael. (2014). Technology of imagination: a card-based public engagement method for debating emerging technologies. *Qualitative Research*, 14(2), 233-251.
- Felt, Ulrike, Fochler, Maximilian, & Sigl, Lisa. (2018). IMAGINE RRI. A card-based method for reflecting on responsibility in life science research. *Journal of Responsible Innovation*, 5(2), 201-224. doi: DOI: 10.1080/23299460.2018.1457402
- Fisher, Erik, Mahajan, Roop L., & Mitcham, Carl. (2006). Midstream Modulation of Technology: Governance From Within. *Bulletin of Science, Technology & Society*, 26(6), 485-496.
- Fischer, Kerstin, Seibt, Johanna, Rodogno, Raffaele, Kirkegård Rasmussen, Maike , Weiss, Astrid, & Bodenhege, Leon. (2020). Envisioning social robotics. Current challenges and new interdisciplinary methodologies. *Interaction Studies*, 21(1), 145-185. doi: <https://doi.org/10.1075/is.18058.fis>.
- Floridi et al. (2018): An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. In *Minds and Machines. Journal for Artificial Intelligence, Philosophy and Cognitive Science*

- Friedman, B., & Hendry, D. (2019). *Value sensitive design, shaping technology with moral imagination*. Cambridge: The MIT Press.
- Friedman, B., Nathan, L., Kane, S., & Lin, J. (2011). *Envisioning cards*. Seattle: University of Washington.
- Granjou, Céline, & Arpin, Isabelle. (2015). Epistemic Commitments: Making Relevant Science in Biodiversity Studies. *Science, Technology, & Human Values*, 40(6), 1022-1046.
- Groves, Christopher. (2017). Review of RRI tools project, <http://www.rri-tools.eu>. *Journal of Responsible Innovation*, 4(3), 371-374.
- Guston, David H. , & Sarewitz, Daniel. (2002). Real-time technology assessment. 24(1), 93–109.
- Hannibal, Glenda, & Weiss, Astrid. (2020). Envisioning social robotics. Current challenges and new interdisciplinary methodologies. *Interaction Studies*, 21(1), 1-6.
- He, Shiquing, & Adar, Eytan. (2016). VIZITCARDS: A Card-Based Toolkit for Infovis Design Education. *IEEE*, xx(xx), xx.
- Hesjedal, Maria Bårdsen, Åm, Heidrun, Sørensen, Knut H., & Strand, Roger. (2020). Transforming Scientists' Understanding of Science–Society Relations. Stimulating Double-Loop learning when Teaching RRI. *Science and Engineering Ethics*. doi: <https://doi.org/10.1007/s11948-020-00208-2>
- Lee, Hee Rin , Šabanović, Selma, Chang, Wan-Ling, Nagata, Shinichi, Piatt, Jennifer A., Bennett, Casey C., & Hakken, David. (2017). Steps Toward Participatory Design of Social Robots: Mutual Learning with Older Adults with Depression. *Human Robot Interaction*(March), 244-253. doi: <https://doi-org.uaccess.univie.ac.at/10.1145/2909824.3020237>
- Luger, Ewa, Urquhart, Lachlan, Rodden, Tom, & Golembewski, Michael. (2015). Playing the Legal Card: Using Ideation Cards to Raise Data Protection Issues within the Design Process. Paper presented at the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul.
- Okada, A., A., Costa, R., Kowalski, P., Torres, Nakayama, M., & K., Souza. (2016). Open Educational Resources for Responsible Research and Innovation: a case study with Brazilian universities and schools. Paper presented at the International Conference on Responsible Research in Education and Management and its Impact.
- Owen, R., Macnaghten, P.M., & Stilgoe, J. (2012). Responsible Research and Innovation: from Science in Society to Science for Society, with Society. *Science and Public Policy*, 39(6), 751-760.
- Šabanović, Selma. (2010). Robots in Society, Society in Robots. *International Journal of Social Robotics*, 2(4), 439-450. doi: [10.1007/s12369-010-0066-7](https://doi.org/10.1007/s12369-010-0066-7)
- von Schomberg, René. (2013). A Vision of Responsible Research and Innovation. In R. Owen, J. Bessant & M. Heintz (Eds.), *Responsible Innovation. Managing the Responsible Emergence of Science and Innovation in Society* (pp. 51-74). Chichester/UK: John Wiley & Sons.
- Schwaninger, I., Güldenpfennig, F., Weiss, A. et al. What Do You Mean by Trust? Establishing Shared Meaning in Interdisciplinary Design for Assistive Technology. *Int J of Soc Robotics* (2021). <https://doi.org/10.1007/s12369-020-00742-w>
- Seibt, Johanna. (2016). Integrative social robotics: A new method paradigm to solve the description and the regulation problem? In J. Seibt, M. Nørskov & S. Schack Andersen (Eds.), *What Social Robots Can and Should Do* (pp. 104–115): IOS Press.
- Seibt, Johanna, Flensbord Damholdt, Malene, & Vestergaard, Christina. (2020). Integrative social robotics, value-driven design, and transdisciplinarity. *Interaction Studies*, 21(1), 111-144.

- SPARC-The Partnership for Robotics in Europe. (2017). Robotics 2020: Multi-Annual Roadmap For Robotics in Europe Horizon 2020 Call ICT-2017 (ICT-25, ICT-27 & ICT-28).
- Taylor, Steve, Pickering, Brian, Boniface, Michael, Anderson, Michael, Danks, David, Følstad, Asbjørn, . . . Woollard, Fiona. (2018). Responsible AI – Key Themes, Concerns & Recommendations for European Research and Innovation (Work Package Report of 'Next Generation Internet': Summary of Consultation with Multidisciplinary Experts).
- te Kulve, Haico, & Rip, Arie. (2011). Constructing Productive Engagement: Pre-engagement Tools for Emerging Technologies. *Science and Engineering Ethics*, 17(4), 699-714. doi: 10.1007/s11948-011-9304-0
- Tzafestas, Spyros G. (2018). Roboethics: Fundamental Concepts and Future Prospects. *Information*, 9(148), 1-25.
- Urquhart, Lachlan D., & Craigon, Peter J. (2021). The Moral-IT Deck: a tool for ethics by design. *Journal of Responsible Innovation*, 1-33. doi: 10.1080/23299460.2021.1880112.
- Vanderelst, Dieter, & Winfield, Alan. (2018). An architecture for ethical robots inspired by the simulation theory of cognition. *Cognitive Systems Research*, 48, 56-66. doi: <https://doi.org/10.1016/j.cogsys.2017.04.002>.
- Williams, Bernard. (1981). *Moral Luck*. Cambridge: Cambridge University Press.
- Wynsberghe, Aimee. (2013). A method for integrating ethics into the design of robots. *Industrial Robot: An International Journal*, 40. doi: 10.1108/IR-12-2012-451