Value Sensitive Design meets Participatory Value Evaluation for autonomous systems in Defence

Burken, Christine Boshuijzen Van; Spruit, Shannon; Fillerup, Lotte; Mouter, Niek

DOI
10.1109/ETHICS57328.2023.10155025

Publication date
2023

Document Version
Final published version

Published in
2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology

Citation (APA)
Burken, C. B. V., Spruit, S., Fillerup, L., & Mouter, N. (2023). Value Sensitive Design meets Participatory Value Evaluation for autonomous systems in Defence. In M. Cheong, J. Herkert, & J. Hess (Eds.), *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology: Ethics in the Global Innovation Helix, ETHICS 2023* (2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology: Ethics in the Global Innovation Helix, ETHICS 2023). IEEE. https://doi.org/10.1109/ETHICS57328.2023.10155025

Value Sensitive Design meets Participatory Value Evaluation for autonomous systems in Defence

Christine Boshuijzen-van Burken
School of Engineering and IT
UNSW
Canberra, Australia
c.vanburken@adfa.edu.au

Shannon Spruit
Populytics
Leiden, The Netherlands
shannon@populytics.nl

Lotte Fillerup
Populytics
Leiden, The Netherlands
lotte@populytics.nl

Niek Mouter
Engineering Systems and Services
Delft University of Technology
Delft, The Netherlands
n.mouter@tudelft.nl

2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS) | 978-1-6654-5713-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ETHICS57328.2023.10155025

***Abstract*— We use Value Sensitive Design for the development of an ethical framework for autonomous systems in Defence in the Australian context. Two novel empirical data gathering methods are deployed for mining stakeholder's values, namely Group Decision Room (GDR) and Participatory Value Evaluation (PVE). GDR findings reveal a general concern for environmental values, geo-political and economic stability. A PVE based on these and other values is designed around an autonomous mine counter underwater vessel and an autonomous drone that drops bombs.**

Keywords—value sensitive design, participatory value evaluation, autonomous systems, military technology, group decision room

I. Introduction

In 2021, IEEE presented their ethics by design framework for systems design, captured in the P7000 standards series [1]. The P7000 is a "Model Process for Addressing Ethical Concerns During System Design" and pays explicit attention to stakeholders' values in systems design. P7000 resembles the 'value sensitive design' (VSD) methodology, initially developed by Friedman and Kahn in the 1990's. VSD is a proven concept for including ethical considerations into technology development, and explicitly calls for stakeholder engagement for mining of stakeholder values. It is an iterative three partite design methodology existing of conceptual, empirical and technical investigations [2]–[4]. We use VSD to inform the development of an ethical framework for military autonomous systems. VSD is enhanced with two novel methods for eliciting values, namely a Group Decision Room (GDR) and participatory value evaluation (PVE).

In this short paper we report on the initial stages of an ongoing project that aims to build an ethical framework to assist developers of autonomous systems in Defence with thinking through the ethical aspects of their technologies. It is based on VSD. Friedman and Hendry state that "the design process engages reciprocally with and, […] co-evolves technology and social structure. Social structures are viewed broadly and may include policy, law, regulations, organizational practices, social norms, and others." [5, p. 68]. Our prospective ethical framework, as social structure, will become part of a VSD good practice for autonomous systems design, potentially informing policy and acquisition. Autonomous systems for military purposes do not operate in a social void, but they are being deployed in the military and on behalf of society. Aligning values serves a pragmatic purpose: the uptake of the systems in the military; a societal purpose: it increases societal support for military endeavors with these technologies; and ethical purpose: improvement of the system in an ethical sense. The context for our project is Australia. Our approach to developing an ethical framework does not merely rely on values as expressed by academics, policy makers and pressure groups in autonomous systems debates (see e.g. [6]–[9]), but includes the voice of developers, industry, users of such technologies and the 'silent majority' of Australian citizens.

This research received funding from the Australian Government through Trusted Autonomous Systems, a Defence Cooperative Research Centre funded through the Next Generation Technologies Fund. The views and opinions expressed here are those of the researchers, and do not necessarily reflect the views of the Australian Government or any other institution.

II. Methodology

VSD is the overall design methodology we are using, and we deploy several methods each with a distinct role in data gathering and framework building.

The VSD methodology has been used in all sorts of contexts, e.g. windmill parks [10], care robots [11], refugee logistics [12], suicide bomber countering technologies [13], and many more (see [14] for an overview). VSD does not prescribe the use of specific tools or methods. We operationalize VSD in this study in three ways, 1) we use Critical Systems Heuristics (CSH) as a way to select relevant stakeholders, 2) we use a Group Decision Room (GDR) to elicit values from stakeholders and explore how they inform design choices, and 3) we use Participatory Value Evaluation (PVE) as a method to invite the wider society (unorganized affected stakeholders) to make value trade-offs and value prioritizations.

A. Critical Systems Heuristics (CSH)

VSD does not provide a standardized way of including stakeholders. Ulrich's framework of CSH includes a method for stakeholder selection [15]. In CSH a distinction is made between the affected and involved stakeholders, and the voice of affected stakeholders is prioritized over the voice of the involved stakeholders. The concerns of affected stakeholders are used to select additional relevant stakeholders (e.g. if it is found that health effects are a concern for affected stakeholders, we could invite medical expert as additional stakeholders) and in this way the affected stakeholders can in part influence the design process indirectly as well as directly. Involved stakeholders related to this study are politicians or industry partners whereas affected stakeholders could be soldiers or citizens of hostile nations, or Australian citizens when systems are employed in the homeland (e.g. for defending national territory). This distinction between affected and involved stakeholders is important, as it directly influences the design process by taking into account what is

978-1-6654-5713-2/23/$31.00 ©2023 IEEE

most important to (a) certain (group of) stakeholders. How an autonomous technology functions, depends on which, or whose, basic point of concern is chosen. If the autonomous system is designed in such a way that it maximizes the value of protection of soldiers that work with the technology, the system will behave differently than when it is designed with the highest legal concerns in mind (e.g. by valuing false negatives over false positives in target detection).

We identified several stakeholders relevant for the design of autonomous systems in Defence: soldiers that work with or alongside autonomous systems, developers and software engineers, national citizens of the countries that develop autonomous systems for Defence, citizens of hostile nations that might be most likely encountering autonomous military systems activity, politicians, lawyers, societal and non-governmental organizations, the international community, industry and businesses. Table 1 shows at what point in the research which stakeholders participated. We note that not all stakeholders that we identified participated in the research, either due to our university's human research ethics restrictions, or because no response was received, or because the invitation was rejected.

In an initial round of semi-structured interviews, we invited involved and affected stakeholders. From these we learned about new affected stakeholders and invited them to the GDR. Members of the Australian Defence Force were not included in this study due to human research ethics limitations, although they are deemed an important stakeholder.

B. *Group Decision Room (GDR)*

A GDR was conducted to understand stakeholders' values and how they can inform and be incorporated into the design of autonomous systems in Defence. A GDR can best be described as an anonymous computer mediated focus group. It is characterised by five principles: anonymity, automatic recording, working in parallel, efficiency and structured discussion. In a standard GDR set-up, participants physically sit together in a room, and each individual contributes anonymously to the discussion through an online interface. We used an anonymous remote set-up where each participant joins an online meeting room (video conferencing software Teams) and interacts anonymously through an online collaboration tool (whiteboard Mural) as well. Participants used the chatbox function when their input did not match the structure. The software tools allowfor a structured discussion to come to focussed inputs, where participants can simultaneously share and develop their ideas. In a GDR, hierarchy, personalities (introvert versus extrovert), power-relations, etcetera, are likely to have less influence in the discussion than in a focus group where a group interacts through face-to-face discussion. It therefore partially solves a critique on Participatory Design, namely that participatory design may be too optimistic about "genuine participation and power relations" [16, p. 16]. An ideal number of participants in a GDR is between 15-20 people.

C. *Participatory Value Evaluation (PVE)*

PVE is a value elicitation method developed in the economical sciences. It aims at mapping values in a large and diverse group of citizens. PVE has been used for large scale value mining regarding, e.g., flood protection [17], COVID-19 measures [18], [19], healthcare funding [20] and climate change mitigation [21]. The essence of a PVE is that citizens can give advice to a decision-maker. Participants are effectively placed in the seat of a decision-maker. In an online environment, they (a) see which options the decision-maker is considering, (b) the concrete impacts of the options, and (c) they have to make choices within given constraint(s). Subsequently, citizens are asked to motivate their choices. Individuals' preferences over (the impacts of) options can be determined by feeding these choices into behaviourally informed choice models and for instance can be used to rank options in terms of their desirability.

The PVE for VSD used in this study is a novel application of PVE. Whereas most PVE's until now have focused on choosing between policy options for supporting governmental decision, this PVE aims to inform design choices to support a more inclusive design process. It allows for inclusion of unorganized affected stakeholders. The options presented in this PVE all maximise a specific value, and participants will have to make a value trade-off (they prioritize certain values over others) in the PVE platform. The rationale for using PVE for VSD is that it allows us to consult large groups of citizens about their values and value trade-offs. In this way the development of autonomous systems in Defence can be better aligned with Australian societal values. One of the additional strengths of the PVE is that it creates opportunities for informing about and actively engaging society with this topic. For instance, by providing information about what the actual risks of autonomous systems are.

D. *Research Steps*

The following research steps for value elicitation from stakeholders to autonomous systems in Defence are suggested. This paper reports on step 1, 2 and 3, although step 3 (the online PVE survey) is designed, but not yet conducted. Each research step provides input for the framework and informs the next step, see Fig 1.

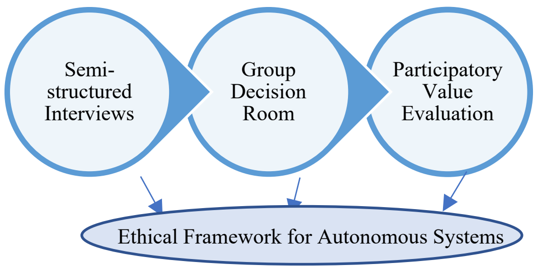

Fig. 1. Overview of research steps for Value Sensitive Design framework.

III. Results

A comprehensive overview of the stakeholders that were selected for each research step can be found in Table 1.

TABLE I. Overview of stakeholder representatives

	Involved stakeholders	**Affected stakeholders**
Interviews	ethicists, (industry) engineers, military experts (ex-Defence), Human-machine teaming expert	International community representative
GDR	(academic and industry) engineers, lawyers, ethicists, military experts (ex-Defence)	Elderly people, international community representatives
PVE	any stakeholder	Australian residents over 18 years old

A. Semi-structured interviews results

A total number of 18 participants from industry, NGOs and academia were invited to partake in the semi-structured interview. A total of seven (7) participants semi-structured interviews were conducted. All interviews were held online with participants in Australia. The interviews served to gather values and to identify potentially overlooked stakeholders that should be invited into the GDR. One finding regarding overlooked stakeholder groups for deciding on values around autonomous systems related to age groups. The generation under 25 years old grows up in a platform era and they experience autonomous systems continuously. Their values are informed by multiple experiences with autonomous systems in their daily lives. The older generation of (60+ years old), which makes up a large portion of Australian society, should be actively included as well, as they have a different 'digital life pattern' and potentially would like to see different values included in autonomous systems. Autonomous systems learn from human behaviour data, but seniors may have a different pattern of behaviour than what a system is trained on, or, seniors may interpret the behaviour of an autonomous system differently than a teenager.

We probed participants about threats to Australia in the broadest sense, as an expressed threat reveals concern for a value(s) that is at stake. Rather unexpectedly, we found that participants held an unanimous concern for climate change, suggesting the importance of environmental values. Another often mentioned value was geo-political stability in the region and Australia's economic position.

B. GDR results

A total number of 46 people was invited into the GDR, of which 17 people eventually participated on one occasion in a two hour session. The demographics are listed in Table 2.

TABLE II. GROUP DECISION ROOM DEMOGRAPHICS

Number of participants in GDR per age bracket[a]				
<29 y	*30-39 y*	*40-49 y*	*50-59 y*	*60 < y*
3	6	2	4	1
Number of participants in GDR per profession				
Software engineering	*Hardware Engineering*	*Law*	*Ethics*	*Other*
3	2	4	3	5

[a] One participant did not list their age

The GDR session was conducted simultaneously via Teams and Mural. Participants took on a self-chosen nickname based on a fruit or an animal to maintain anonymity. The GDR consisted of the following exercises that were done via typing on sticky notes and moving them around on the virtual Mural board:

a) Threats to Australia

List threats to Australia, and vote and prioritize them

b) The role of autonomous systems in threats

List an autonomous system that can mitigate the threat

c) Braindump Values

List all the values relevant to any autonomous system

d) What values are relevant in autonomous systems?

Match up an example autonomous system with most important values for that example system. Indicate clashes between values and distinguish between 'must have' and 'nice to have' values. Translate the value into a norm, required behaviour or a design requirement for the system.

1) Threats to Australia

Findings from the GDR were similar to the interview findings when it came to threats: climate change, armed conflict and economic instability (due to aging population, lack of skilled workforce, economic inequality) were mentioned often. However, a lack of social cohesion and weak government and leadership was mentioned too.

2) The role of autonomous systems in threats

The participants saw a role for autonomous systems in countering the threats of climate change and armed conflict. For mitigating climate change threats, they proposed autonomous systems that assist in agriculture, surveying the environment, disaster response and smart coordination of energy grid.

Mitigation of armed conflict threats by autonomous systems were suggested in the form of systems for counter missile and mine defence, for dropping bombs on targets, systems that detect civil or protected physical or digital infrastructure and systems that detect foreign threats.

3) Braindump values

Participants named a total of 51 values that can roughly be divided into 7 categories, namely learning, transparency, trustworthiness/reliability, explainability, accountability, security, protection and costs. Participants voiced their concerns mostly about the trustworthiness/reliability of the autonomous systems as well as the possibility of systems being compromised by adversaries. Lastly, the ability of systems and people to learn and people's ability to change their minds based on the context was important. Therefore, the value of learning and adaptability was brought forward and discussed.

4) What values are relevant in autonomous systems?

Participants chose to mostly work with three example systems, namely "drones that plant trees and attend to them", "drones that drop bombs on targets" and "autonomous mine counter vessel". In the design of tree planting drones, values like integrity, privacy and buildable were mentioned most often. Regarding the practical implementation of the drones, participants expressed concern for the needs and concerns of stakeholders, the geographic limits in its operation, its multi-use function, and keeping the drone low-cost.

For the design of the mine counter vessel participants considered reliability, protection of marine life and non-combatants, transparency, accuracy, cost and, safety important values. Practical solutions for implementing values were a risk registration, human in the loop, clear communication with operator, dedicated authority, programming so it always complies with laws of armed conflict and just war principles, thorough testing, heaps of sensors, self-destruct of valuable data in case systems gets lost and redundancy via sensors.

The values that were most important in the design of the autonomous drones that bombs targets were trust, distinction, control, reliable and accountable, international humanitarian law (IHL), explainability, awareness of limits, traceability and, security. Practical solutions for implementing values were drones to be programmed to always comply with IHL, have a designated human to account for final outcomes, track decision making via a log, ping a human before engaging, have demonstrated the ability to strike intended target and not strike non-targets, training of staff, training the system on data specific to conflict for identification of friend and foe, release videos for third party review and, set criteria and limitations for use. The participants named control through a human in the loop and response time a clear design tension. While they considered multiple technological solutions, they concluded this tension was mainly caused by a distrust in the abilities of autonomous systems rather than the technological impossibilities. Participants expected that the use of autonomous systems for military purposes would cause resistance from the public, while they felt that for combating climate change there would be less resistance. Therefore, they believed that machine and human learning should be of high importance to increase reliability in autonomous weapons used in warfare.

C. *PVE – set up*

We discuss the design of the PVE here (in the results section), because it is the result of the GDR. We chose two example autonomous systems to elaborate on in the PVE, namely:

1) An autonomous mine counter vessel
2) An autonomous drone that drops bombs

These two examples were chosen because, firstly, they were examples about which we gathered sufficient input from the participants in the GDR session, which reflects a wide range of values that can be weighed and traded off. Secondly, the two examples resemble both offensive and defensive autonomous systems, and in this way we can see whether participants make different value trade-offs when it comes to offensive and defensive situations.
Next, we took the values that were listed in the GDR for each of the examples and translated them into design requirements. Some of the design requirements where already listed down in the GDR sessions, while others were developed by the researchers.
We designed an PVE-experiment with two different designs. One PVE in slider mode (see Fig.1), where participants can choose to maximise choices on particular values. And a PVE in 'pick-mode' (see Fig. 2) where participants can choose design features that implicitly maximise values.

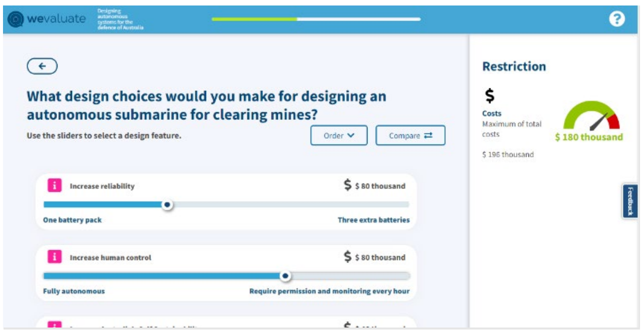

Fig. 1. Example of a PVE with 'slider mode' for making design choices.

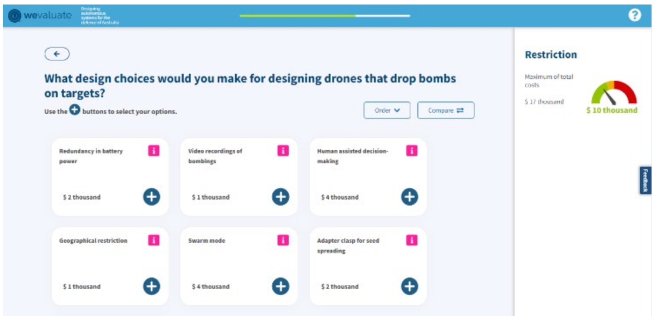

Fig. 2. Example of a PVE with 'pick-mode' for making design choices.

Constraints are provided through a fictional budget represented by a meter, which can point to green (costs are within budget), orange (costs are close to the maximum budget), and red (costs exceed maximum budget). To test the assumptions underlying the PVE, we consulted several experts working on autonomous systems to check whether the costs meter and the design requirements we drafted were within reasonable bandwidths.
Our reporting here excludes results from the PVE because our university's human research ethics committee is still pending.

IV. Discussion

The research approach presented in this paper provides a promising approach to operationalizing VSD. It makes it possible to give voice to an often-overlooked group of stakeholders, namely unorganized affected stakeholders. This group often do not get to sit at the table when it comes to design decisions. Furthermore, this group often only encounter the consequences of design choices when technology is implemented in society.

The combination of a GDR session with PVE, which is also novel, helps to come to a PVE design that is informed by the experience and knowledge of stakeholders and speeds up the PVE design process. The CSH element encourages researchers to be open minded, ensuring all relevant societal voices are heard to develop a PVE that is recognizable and understandable for a wide audience.

Limitations to our approach were as follows. First, we note the exceptional circumstances that preceded and unfolded during this research that may have influenced the findings. The unanimous concern for environmental values may be aggravated due to Australia's recent 'Black Summer' bushfire season (2020), followed by unprecedented floods (2022) that have impacted many Australians directly or indirectly. In addition to the natural disasters, there was a rise of global tensions when Russia invaded Ukraine in February 2022, with Australians closely following the potential affects in the Indo-Pacific region, and not the least because of its AUKUS pact, which is a security pact between Australia, USA and the UK established in 2021.

Secondly, some of the values that are mentioned in the results section resonate with existing frameworks that lists principles for the responsible use and development of AI. This is partly because we provided an example list that the participants could draw from during the GDR. The example values were generated from existing literature [7], [22]–[24]. We acknowledge that a distinction exists between values and principles, but for the purpose of this study we ignore this, since we adopt Friedman and Kahn's rather loose definition of a value which may include principles. They state that a value is something that "[…], a person or group of people consider important in life" [25, p. 349]. between Values that were mentioned by participants and that coincide with typical AI principles found in literature are traceability, accountability, control, trust.

Thirdly, we note that some values are mentioned for all example systems, whether they be under water, aerial, for military or non-military purposes, whereas others are unique to the respective systems. E.g., human in the loop was not mentioned for the drone that plants trees and neither was explainability. On the other hand, privacy was mentioned in the drone that plant trees, but not in the military examples. We note that one GDR with limited number of participants (17) is insufficient to draw broad conclusions from, but they give us some insights on how a larger group of participants may appreciate values differently, depending on its area, context and purpose of use. Suggestions for further research is to tailor principles for AI and autonomous systems development to the context of use, rather than develop broad general principles meant to apply to a multitude of user contexts.

Finally, members of the Australian Defence Force and people outside Australia were excluded from participating in this research, as this would complicate human research ethics approval processes and cause delays beyond the available funding timeline. This leaves important stakeholder groups out, which is a limitation. We suggest that future research should include members from the Australian Defence Force and ideally citizens of foreign nations.

References

[1] IEEE, "IEEE 7000-2021 - IEEE Standard Model Process for Addressing Ethical Concerns during System Design," IEEE Computer Society, IEEE Standards ISBN 9781504476874, 9781504476881, 9781504479356, Sep. 2021. [Online]. Available: https://standards.ieee.org/standard/7000-2021.html

[2] B. Friedman, "Value-sensitive design," *ACM Interactions*, vol. 3, no. 6, pp. 17–23, 1996.

[3] B. Friedman and P. Kahn, "Human Values, ehics and design," in *The human-computer interaction handbook*, 2003, pp. 1177–1201. [Online]. Available: https://brandorn.com/img/writing/tech-ethics/human-values-ethics-and-design.pdf

[4] B. Friedman, "User Autonomy: Who Should Control What and When? A CHI 96 workshop," *ACM SIGCHI Bulletin*, vol. 30, no. 1, pp. 26–29, 1998.

[5] B. Friedman and D. G. Hendry, *Value sensitive design: Shaping technology with moral imagination*. Cambridge, MA: Mit Press, 2019.

[6] ICRC, "ICRC Position Paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach," *International Review of the Red Cross*, vol. 913, 2021, Accessed: Apr. 13, 2022. [Online]. Available: http://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913

[7] NATO, "NATO Review - An Artificial Intelligence Strategy for NATO," *NATO Review*, Oct. 25, 2021. https://www.nato.int/docu/review/articles/2021/10/25/an-artificial-intelligence-strategy-for-nato/index.html (accessed May 28, 2022).

[8] T. Walsh, "Autonomous Weapons: an open letter from AI & robotics researchers." Future of Life Institute, 2015. Accessed: Nov. 06, 2020. [Online]. Available: https://futureoflife.org/open-letter-autonomous-weapons/

[9] T. Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds & Machines*, vol. 30, no. 1, pp. 99–120, Mar. 2020, doi: 10.1007/s11023-020-09517-8.

[10] I. Oosterlaken, "Applying Value Sensitive Design (VSD) to Wind Turbines and Wind Parks: An Exploration," *Sci Eng Ethics*, vol. 21, no. 2, pp. 359–379, Apr. 2015, doi: 10.1007/s11948-014-9536-x.

[11] A. Van Wynsberghe, "Designing robots for care: Care centered value-sensitive design," *Science and engineering ethics*, vol. 19, no. 2, pp. 407–433, 2013.

[12] C. Boshuijzen-van Burken, R. Gore, F. Dignum, L. Royakkers, P. Wozny, and F. L. Shults, "Agent-based modelling of values: The case of value sensitive design for refugee logistics," *JASSS: Journal of Artificial Societies and Social Simulation*, vol. 23, no. 4, 2020.

[13] L. Royakkers and M. Steen, "Developing Tools to Counteract and Prevent Suicide Bomber Incidents: A Case Study in Value Sensitive Design," *Science and Engineering Ethics*, vol. 23, 2016.

[14] T. Winkler and S. Spiekermann, "Twenty years of value sensitive design: a review of methodological practices in VSD projects," *Ethics Inf Technol*, no. 23, pp. 17–21, 2021, doi: 10.1007/s10676-018-9476-2.

[15] W. Ulrich and M. Reynolds, "Critical systems heuristics," in *Systems approaches to managing change: A practical guide*, Dordrecht: Springer, 2010, pp. 243–292.

[16] M. Van der Velden and C. Mörtberg, "Participatory Design Design participatory and Design for Values," in *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, J. van den Hoven, P. E. Vermaas, and I. van de Poel, Eds. Dordrecht: Springer Netherlands, 2015, pp. 41–66.

[17] N. Mouter, P. Koster, and T. Dekker, "Participatory value evaluation for the evaluation of flood protection schemes," *Water Resources and Economics*, vol. 36, p. 100188, Oct. 2021, doi: 10.1016/j.wre.2021.100188.

[18] N. Mouter, J. I. Hernandez, and A. V. Itten, "Public participation in crisis policymaking. How 30,000 Dutch citizens advised their government on relaxing COVID-19 lockdown measures," *PLOS ONE*, vol. 16, no. 5, p. e0250614, mei 2021, doi: 10.1371/journal.pone.0250614.

[19] N. Mouter *et al.*, "Stepping into the shoes of the policy maker: Results of a Participatory Value Evaluation for the Dutch long term COVID-19 strategy," *Social Science & Medicine*, vol. 314, p. 115430, Dec. 2022, doi: 10.1016/j.socscimed.2022.115430.

[20] A. H. Rotteveel *et al.*, "If you were a policymaker, which treatment would you disinvest? A participatory value evaluation on public preferences for active disinvestment of health care interventions in the Netherlands," *Health Economics, Policy and Law*, vol. 17, no. 4, pp. 428–443, Oct. 2022, doi: 10.1017/S174413312200010X.

[21] A. Itten and N. Mouter, "When Digital Mass Participation Meets Citizen Deliberation: Combining Mini- and Maxi-Publics in Climate Policy-Making," *Sustainability*, vol. 14, no. 8, Art. no. 8, Jan. 2022, doi: 10.3390/su14084656.

[22] Kate Devitt, Michael Gan, Jason Scholz, and et al., "A method for ethical AI in Defence," Australian Government Department of Defence, Canberra, DSTG-TR-3786, 2021. [Online]. Available: https://apo.org.au/node/311150

[23] OECD, "Recommendation of the Council on Artificial Intelligence," *OECD Legal Instruments*, 2019. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed Jan. 08, 2023).

[24] Australian Government. Department of Industry, Science, Energy and Resources, "AI Ethical Principles," 2019. https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles (accessed Aug. 24, 2020).

[25] B. Friedman, P. Kahn, and A. Borning, "Value sensitive design and information systems," in *Human-Computer Interaction in Management Information Systems: Foundations*, P. Zhang and D. Galletta, Eds. Armonk, NY: M.E. Sharpe, 2006, pp. 348–372.

Acknowledgment

We thank all interviewees and participants in our Group Decision Room. We thank three anonymous reviewers.