INTERNATIONAL DESIGN CONFERENCE – DESIGN 2020
https://doi.org/10.1017/dsd.2020.78

A FRAMEWORK FOR ADDRESSING ETHICAL CONSIDERATIONS IN THE ENGINEERING OF AUTOMATED VEHICLES (AND OTHER TECHNOLOGIES)

J. Millar [1], D. Paz [2], S. M. Thornton [2]✉, C. Parisi [3] and J. C. Gerdes [2]

[1] University of Ottawa, Canada, [2] Stanford University, United States of America, [3] Apple Inc., United States of America

✉ smthorn@alumni.stanford.edu

Abstract

Policymakers have attempted to preemptively address the concern of ethical issues with the regulation of automated vehicles. Unfortunately, both policymakers and designers of these technologies struggle to articulate ethical issues and their resolution. We propose a framework that engineers and designers of automated technologies can apply that allows them to identify and resolve ethical tensions within the design task. We demonstrate the practicability of the framework to the engineering design process through a human-subject study where engineers applied the framework in a workshop.

Keywords: ethics, participatory design, artificial intelligence (AI), value sensitive design, autonomous vehicle

1. Introduction

In the fall of 2016, the US National Highway Traffic Safety Administration (NHTSA) released its first Automated Vehicles (AV) Policy (AVP1.0) (NHTSA, 2016a). AVP1.0 asked all "individuals or companies manufacturing, designing, testing, and/or planning to sell automated vehicle systems" to voluntarily provide NHTSA with a "Safety Assessment Letter" indicating how they had satisfied each of fifteen design requirements. One of those requirements was the treatment of "ethical considerations," or conflicts among values, during the design of AV systems. During the public comment period following AVP1.0's release, some manufacturers and professional engineering organizations raised a key concern regarding the "ethical considerations" requirement. They pointed out that it was unclear how to go about addressing ethical considerations during the design of automated vehicles; this novel requirement, though supported in principle, had no corresponding engineering framework with which to accomplish it in practice (NHTSA, 2016b; Verhalen, 2016). Others argued in support of the discussion proposed by the policy, stating that, "Because automated vehicles promise such a broad and deep human impact, companies should consider the ethical dimensions of them in comparably broad and deep terms" (Kenner, 2016). A year later, while revising the AV Policy, NHTSA removed the ethical considerations requirement, prompting some calls from state DOTs and consumer groups to have the "ethical considerations" reinstated in future versions of the Policy (Kyrouz, 2018). In the absence of this requirement, the public discussion envisioned in AVP1.0

never happened, leaving a fundamental question open: How does one address ethical considerations during the engineering of automated vehicles (or other automation technologies, for that matter)?

Ethical issues in robotics and artificial intelligence are surfacing regularly. In response, many have proposed methodologies and toolkits to help decision-makers and technology designers better understand the ethical implications of potential products in order to adjust their design to be more ethically robust. Mediation analysis and Value Sensitive Design (VSD) have been suggested as potential methodologies to assist with ethical evaluation of technologies (Gerdes et al., 2019; Millar, 2016), but require additional tools to support their use in producing thoughtful and justifiable ethical design decisions. The Markkula Center for Applied Ethics provides an Ethics Toolkit (Vallor, 2018) which includes a broad set of tools to assist with ethical decision-making in regards to technology development. The toolkit includes a sample design workflow: idea, sketch, feasibility assessment, prototype, design of user-interface, development, testing, release and feedback. The Open Roboethics Institute developed an ethics toolkit, known as the Foresight in AI Ethics (Moon et al., 2019), to help with developing an ethics roadmap for artificial intelligence projects. Having a general approach for identifying and addressing a broad range of ethical issues in engineering practices will likely help anticipate those ethical issues earlier in the development cycle, and help to mitigate the negative and/or enhance the positive social impacts those technologies produce (Friedman et al., 2008). Such an approach could also increase the likelihood that emerging technologies are aligned with human values by embedding values-based considerations in the engineering design process. While these methodologies and toolkits are likely useful in assisting engineering design, to the authors' knowledge, there are no studies of engineers using these techniques to support such claims.

In this paper we set out to accomplish two goals. First, we outline a general ethical considerations framework ("the framework" hereafter)—that includes a set of concepts, methodologies and tools for addressing ethical considerations in the design and engineering of automated systems, including automated vehicles. Second, we report the results of a qualitative empirical study we conducted, using a Grounded Theory approach, to test the effectiveness of the framework for helping engineers identify ethical issues associated with autonomous vehicles and start working toward solutions that help address those ethical issues. The purpose of the study was to gauge the effectiveness of the framework as a means for equipping engineers to identify ethical challenges in their design tasks and apply ethical considerations in the design of automation and artificial intelligence technologies. We also wanted to gauge how useful engineers considered the framework, and whether or not they felt the framework provided enough guidance for them to apply it independently in their daily engineering tasks. Finally, we wanted to understand what ethical supports engineers might need in order to better apply the framework in their daily work.

This research is situated within a broader set of activities that has emerged in the past few years, which focus on the ethical engineering of robotics and artificial intelligence.

1.1. The framework

We prototyped the framework over the course of a year at Stanford University's McCoy Family Centre for Ethics in Society, the Center for Automotive Research at Stanford (CARS), and a large technology firm (TecFirm hereafter). We introduced multi-disciplinary groups of participants (engineers, designers, and other industry stakeholders) to the framework using a formal workshop format that we describe in more detail below.

1.2. Designing the framework

The framework includes a set of concepts, methodologies and tools for addressing ethical considerations during the engineering of a technology. It has its roots in Value Sensitive Design (VSD), which is an open-ended design framework that helps to analyze technology in terms of the human values that technology expresses (Friedman et al., 2008). As a framework, VSD prompts the designer to focus on a broad set of stakeholders impacted by the technology under consideration (e.g. users, policy makers, the environment, the public), the values attached to those stakeholders (e.g. privacy, trust, profit), and the value tensions that can arise between different competing values and associated stakeholders (Friedman et al., 2008). For example, an engineer might value access to users' location data in order to improve some feature in a mapping application, which will be in tension with the users' expectations of privacy

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

when using the mapping application. As an open-ended framework, VSD does not restrict designers to, or set out to define, a particular process or set of tools to use when implementing VSD in a design environment; designers are free to deploy VSD in any number of unspecified ways (Friedman and Kahn, 2003). Thus, VSD can be used to, among other activities: analyze an existing technology to identify the values embedded in it; produce a set of design/engineering requirements for a technology under development; or explore the impact of a particular technology on society.

Our framework adds a set of concepts, methodologies and tools to the underlying VSD framework. VSD's open-endedness can make it somewhat unwieldy in engineering contexts, in part because open-ended frameworks are difficult to embed in engineering design processes efficiently, reliably and consistently. Because engineers value efficiency, reliability and consistency in their workflows (the history of standardization is a history of the success of these three properties), any new practice that does not score high on these three properties can be a tough sell that engineers will likely, and reasonably, reject. In order to produce a workable engineering framework, we treated those three properties as constraints in our design of the framework. Satisfying those three constraints required "taming" VSD with a relatively well-defined process to help "close" the design space to a certain extent, thus the introduction of a set of concepts, methodologies and tools to act as a sort of semi-standardized wrapper around VSD.

Another of our important design goals for the framework was that the concepts, methodologies and tools we defined must be relatively simple to use, requiring little knowledge of ethics (as an academic knowledge domain). Our goal is not to use the framework to make philosophers out of engineers. Rather, we aim to empower engineers to anticipate and address ethical considerations specific to their technology by providing a framework that does a lot of the heavy ethical lifting. That said, the framework does not include a set of formulas used to answer ethical questions. The framework is designed to: help expose relevant ethical issues related to a particular technology upstream in the engineering design process; and prompt engineers to identify and reason through design options in more detail, and with a more informed, nuanced and critical eye, than they would have otherwise done.

1.3. Prototyping the framework - the workshop

In order to prototype the framework, we introduced our target users to the set of concepts, methodologies and tools—the framework—via a highly interactive four-hour workshop. The workshop thus served as a "How-To" guide to working with the framework. The workshop and the prototype framework, both of which we report on in this paper, are the result of seven prototyping workshops, five of which were conducted with various teams at TecFirm, the other two with student and industry groups at Stanford University, between Fall 2017 and Summer 2018.

2. Methodology

To begin validating the framework (following the seven prototyping workshops), we ran a single IRB-approved workshop at Stanford, using a qualitative Grounded Theory approach for the study design, with participants recruited from various manufacturers associated with the automated vehicle industry.

The workshop consisted of a pre-workshop activity, an introductory presentation, followed by a five-step series of rapid (15-20 minutes each) design activities, each paired with a group discussion/debrief, and each intended to focus participants' efforts on either a particular aspect of the framework or a technique used to move the design efforts forward. Finally, the workshop concluded with a one-hour semi-structured group interview that took the form of an audio-recorded open-ended discussion.

2.1. Participants

The workshop was run exclusively with Affiliates of the Center for Automotive Research at Stanford (CARS). The participants were a diverse group of individuals with various backgrounds of employment. A majority of the participants were engineers but there were other occupations, such as managers and computer scientists. There was no data collected on the participants' age or gender. A total of 16 participants attended the workshop. The group was divided into subgroups of three or four participants. The subgroups were selected by separating employees who worked at the same company so that the participants would be less familiar with other members of their subgroup. This was intended to prevent the participants from assigning "roles", and to promote a diversity of views, within subgroups.

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

2.2. Materials and physical setup

This study was performed in a collaboration room at Stanford's d.school. The collaboration room was set up as an open workspace with movable tables organized in a semi-circle. The tables were standing tables with high stools placed at each one allowing the participants to sit, stand or walk around depending on their preferences. The participants sat at tables in groups of three or four and each group was equipped with a white board, markers, pens, post-it notes, and the handouts provided to them by the researchers. These handouts included a Framework Workbook, Pre-workshop Activity Book, and Data Package, each outlining the various steps involved in the corresponding part of the workshop and providing instruction to participants on how to complete each step. The standing tables also allowed for an easy transition between working at the table and writing on the whiteboard. The event included breakfast and lunch and the participants were compensated for their time with a $25 Amazon gift card.

2.3. Procedure

This study was conducted like a workshop that would be run at a company or as a standalone event. Participants were asked to work on a design challenge while being guided through the framework. The entire workshop duration was four hours with a break after three hours. However, the participants were free to walk around, grab food and such as needed through the entirety of the event.

2.3.1. Pre-Workshop design activity

The workshop began with a pre-workshop activity to get the participants immediately thinking about the design challenge. The premise of the pre-workshop activity was to give the participants something to reflect on, in a pre/post reflection activity, once the entirety of the workshop was over. The pre-workshop activity was designed to capture participants' intuitions about how to approach the design challenge prior to their framework introduction. The researchers did not explain the framework or what they were going to be doing throughout the day. The participants were handed a packet that described the scenario and challenge. The scenario involved a car approaching a pedestrian crosswalk (Figure 1).

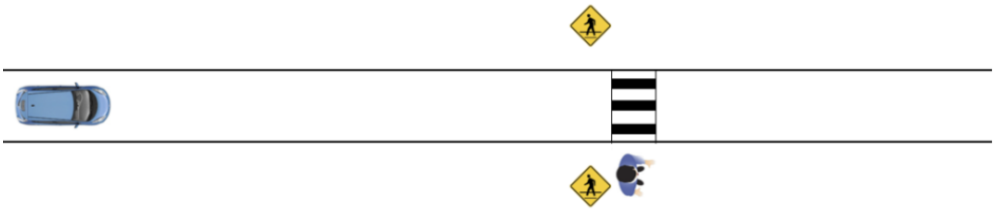

Figure 1. Scenario presented to participants during the workshop

In the design scenario, the speed limit is 25 mph, the vehicle is 100 ft away from the crosswalk, and there is a pedestrian at the edge of the crosswalk with a 20% probability of entering the crosswalk. The participants were tasked with determining what the vehicle should do in this particular situation. Along with providing the participants a document describing the scenario, they were given an information package. The information package included: California pedestrian law for pedestrian crosswalks (CA Veh Code § 21950), vehicle stopping distance curves, and injury curves.
Participants were instructed as follows:

> *In your group, take 15 minutes to discuss one or two candidate solutions to the design challenge in as much detail as possible. You can also describe general approaches you might take for addressing this design challenge.*

Groups were given roughly 15 minutes to complete the Pre-Workshop Design Activity, which was followed by a 5-minute all-group discussion during which participants described the general approaches to problem-solving they discussed in their groups.

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

2.3.2. Introduction to the framework – slide presentation

One of the researchers gave a brief presentation outlining the framework. This was to explain what it is, what it might contribute to the engineering design process, and a brief overview of what they would be doing throughout the workshop. The presentation lasted approximately 20 minutes.

2.3.3. Stakeholders

VSD (the underlying design methodology we incorporated into our framework) is premised on the idea that the ethical implications of a technology are rooted in that technology's impact on different stakeholders—the various individuals and/or groups of individuals (e.g. users, designers, engineers, retailers) whose goals and values are somehow affected by that technology (Friedman and Kahn, 2003). As such, each of those stakeholders adds a perspective from which to analyze the ethical implications of a technology. By analyzing a technology from multiple stakeholder perspectives, engineering teams can uncover value tensions (i.e. ethical issues) that result from different configurations of the technology. The first step in analyzing the technology, therefore, is to generate a comprehensive list of stakeholders, both direct and indirect, who are (potentially) impacted by the technology.
During the Stakeholders activity, participants were instructed to:

> *Write down a list of all direct and indirect stakeholders you feel are implicated by the design challenge under consideration.*

Groups were given 5 minutes to complete this task and were encouraged to generate a list of at least 20 stakeholders. The subsequent group discussion (10 minutes) was used to expand significantly on the list. Groups were encouraged to add stakeholders to their lists that other groups had identified in order to build as comprehensive a list as possible. In our pilot workshops we found that expanding the list of stakeholders helped open up the design space, thus opening up subsequent avenues for investigation and analysis that would otherwise remain opaque.

2.3.4. Goals and values

Participants were then asked to focus on three to five stakeholders and list those particular stakeholders' goals and values that related to the design scenario. This step is meant to challenge the participants to think deeply about the problem and identify what is important to each of the stakeholders that their group identified. The researchers described the goals as more concrete (e.g. crossing the street, getting home) and the values as more abstract (e.g. mobility, legality). The participants were instructed specifically to:

> *Select between three and five Key Stakeholders from the previous step and write down some goals and values for each of them.*

The participants were instructed to think about the goals and values in terms of the context of the design task and then describe each goal and value in as much detail as possible. Once each of the groups created a list of their own, they were asked to share key goals and values with the group during a 10-minute group discussion.

2.3.5. Tensions

After having identified a number of goals and values participants identified value tensions between various goals, values and stakeholders. Identifying value tensions is a good way for the workshop participants to understand and clearly see how the goals and values of the stakeholders involved often conflict with one another, leading to ethical issues. For this section, the framework workbook included a fill in the blank tool that participants were instructed to use in describing up to four value tensions:

> *[value 1] conflicts with [value 2] because …*

A simple example of a value tension is the pedestrian's value of mobility conflicting with the AV passenger's value of mobility. Pedestrians value mobility in that they need it to get where they are going in a timely manner, while the AV passenger values mobility for a similar reason. Yet neither stakeholder can maximize their own mobility without sacrificing the other's, hence the tension. This is

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

how the participants were asked to think about the values they came up with in the Goals and Values step. At this point, the participants only needed to come up with the potential conflicts and in the following step they were asked to think about resolving them.

2.3.6. Dissolution, compromise and tradeoffs

In the next step of the workshop, participants tried to resolve one of their value tensions using one of three approaches: dissolution, compromise or tradeoff. Participants were given the following instructions:

> *Focusing on the value tensions you identified in the previous step, use this space to begin exploring how you might resolve them.*

The framework workbook offered brief explanations (another design tool!) of each of the three proposed methods, while acknowledging that others could be used. Dissolving the tension was described as figuring out a solution that satisfies both of the values. An example of this might be building a pedestrian bridge over the street such that both the pedestrian and the vehicle both maintain mobility. A compromise was described as when each of the competing values is satisfied to some extent, but not fully, while a tradeoff was described as involving the selection and satisfaction of one value at the expense of the other. The participants were challenged to play with each of the three strategies in attempts to resolve one of the tensions they identified in the previous step. This was the last critical thinking step prior to Rapid Ideation, so it was important that the participants were able to work through the potential resolutions prior to designing an implementation of their preferred resolution.

2.3.7. Rapid Ideation

The penultimate step in the workshop was Rapid Ideation, which refers to creating many ideas in a short amount of time. The participants were instructed to:

> *Use this space to sketch out or describe some ideas for at least one solution you and your group feel passionate about, and get ready to pitch your group's preferred solution to the class.*

The Rapid Ideation part of the workshop was where the groups were able to come up with some concrete prototypes to the crosswalk scenario. In guiding participants to think broadly about their prototypes, the researchers emphasized that solutions could be technical, but could also focus more on formulating general values-based design principles, processes, or follow-up stakeholder engagements that could be prototyped while working towards a technical solution. Thus, participants were encouraged to consider a variety of solution types that could be useful when designing and progressing towards a technical solution. In a group discussion the participants were asked to describe their prototypes to other groups. This discussion lasted roughly 10 minutes and jumpstarted the final reflection/debrief portion of the workshop.

2.3.8. Reflection/Debrief – group discussion

During the final step in the workshop, we conducted a thirty-minute, audio-recorded, open-ended group reflection/debrief interview using a series of question prompts designed to elicit feedback about both the framework, and the workshop format used to introduce it. Participants were asked to reflect on the Pre-Workshop Design Activity and describe if/how engaging with the framework during the workshop changed the way they approached the design challenge. One researcher facilitated the discussion with a list of predefined questions but also allowed the participants to discuss freely anything they felt was relevant. At the end of the workshop the researchers collected all of the design artefacts, including workbooks and data packets, and photographed the whiteboards that the participants used to generate and work through ideas throughout the workshop.

3. Results

Researchers analyzed both the participant packets and the audio recording transcripts. The transcripts were analyzed using a Grounded Theory approach, which is common in the social sciences for analyzing unstructured interview data. In a Grounded Theory analysis, researchers extract regularly occurring

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

ideas, themes or concepts from the data ("themes" hereafter) that are meant to represent conceptual trends expressed by study participants. Theme extraction is accomplished through a systematic "coding" process involving multiple researchers and multiple passes through the data. In our analysis, two of the researchers independently coded the interview transcripts to identify frequently occurring phrases, ideas or sentiments that resulted in candidate themes (each identified as subsection headers and discussed in detail below). The candidate themes were then compared across researchers to produce a final list of themes. In a final pass, the researchers used the finalized list of themes to independently re-code the transcripts. The researchers then independently compared the two coded transcripts to resolve disagreements between coders. Each researcher considered the disagreement, and decided whether the quote representing a theme was a better fit in the theme originally selected or that of the other researcher. The researchers then collaborated to compare their independent theme categorizations and resolve any remaining disagreements. The following five themes were extracted from the data.

3.1. The workshop imparted transferrable skills to participants

One of the researchers' primary goals in designing a workshop around the framework was to build enough capacity so that participants felt they could apply the framework to their daily work. During the post-workshop interview, several participants explained that they felt like the framework was something they could bring back to their workplaces. One participant explained, "I liked that this felt, the way it was conducted here, this felt lightweight enough that I could see doing parts of it, you know, either in a small group, or just saying, 'Hey, let's…carve out some time and think about [the design problem] from this perspective.'" When comparing this experience to other workshops that try to engage participants in thinking about empathy when designing, the same participant commented that the workshop defines a process that "seems really different…you're thinking systemically. And, [this] seems more doable, like the payoff is bigger for an engineer."

Another participant felt that the workshop provided a way to facilitate conversations with stakeholders, explaining, "One thing that I kept wondering about was that we're going off of the values…that exist today among those stakeholders and I think it would have been fun to also think about what values might shift. What are some of the new goals and values that we might see because of autonomous vehicles being on the road? I feel like that could bridge a little bit when I think about having conversations like that with the customers; how it might get them to think a little more broadly about impact of what we are doing, and what changes might happen." The idea of sharing the results of the workshop with a customer illustrates the value of the workshop and how the skills could transfer directly into some participants' workplaces.

3.2. The benefits of focusing on a broad set of stakeholders and values

One theme that clearly emerged in the interview was the impact of focusing on a broad set of stakeholders and values. This included how the participants grew the list of stakeholders, the positives and negatives of considering a broad list of stakeholders, and the influence those considerations had on scoping the complexity of the design problem. Participants expressed excitement that the framework helped them "go beyond the obvious" and think about the stakeholders that took more time to identify. One participant found this part of the process to be "eye opening." Initially, participants felt that "there's a concept of some [stakeholders] being maybe more crucial than others" without "completely thinking of the whole problem at first," but "as a result of the exercise, we get to see, maybe there's another stakeholder that is equally or more important and, this can help shape what solution we arrive at."

Another participant described how it seemed important to get a variety of stakeholders involved in the workshop because it engaged "different people that have different points of views and different perspectives, and you can see, like, how the solution kind of broadened or is different than what you might have initially started off with in your own head or perspective." During pilot workshops the researchers noticed that separating individuals with similar jobs (or from the same company) by placing them in separate workshop groups was a useful strategy for generating good discussions during the workshop. Anecdotally, the diversity of participants in each workshop group seemed linked to the diversity of the stakeholder list generated during group breakout discussions. One participant explained, "you bring in the values of the participants, too, to try and figure out what do I see versus what do you see. I think it's really a

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

compelling example." Additionally, through the workshop, one participant thought critically about what people say is important versus what they truly believe, stating, "Yes, we can imagine what we think is important, but if you can actually ask them, they might surprise you and say something else. And another point is, they may tell you something, but then what they actually do in life may be yet different again from what they tell you is important." Overall, the ability to interact with a diverse group of people to come up with a broad list of stakeholders and their values was positively received and allowed participants to think about the problem in a different light. Participants expressed appreciation that the framework motivated activities that are "different than just plain brainstorming," in that "it brings a focus to project management in—who are the stakeholders? And then you add a new dimension, which may not be present in many project management ideologies. What are their values? Because generally people have to suppress their values and just stick with whoever [is] the top person, whatever they're saying. So, this [process] allows us to spill everything out on the table, focus down, zoom out, focus down, zoom out."

3.3. Interdisciplinary discussions

The workshop structure provides time and space to have (non-typical) interdisciplinary discussions. The participants expressed appreciation for working in small groups because "no matter what the scene," "you'll end up with different conclusions in the end. Then you can compare and then see if there's a synergy." Many of the participants described how typical conversations in their workplace involve the same people repeatedly focusing on the same problems rather than getting people from multiple departments to focus on the problem. In reflecting on the workshop, one participant described how they felt the framework "highlights the importance of teamwork, because a system like this, teamwork with team members from different functions, [contrasts with what] we always tend to say, you know…I have the answer, you know, a team of one, I can solve it…So with exercises like this, or with actual problems, teamwork and diversity of the team…their expertise really helps." Moving forward, it is also clear that "autonomous vehicles [are] not just an engineering problem"; "we need to look at the problem holistically, and that's where the [framework] is going to be hugely helpful." The workshop encouraged interdisciplinary discussions which seems to have prompted the participants to want to take the framework back to the workplace and use it to structure broader interdisciplinary discussions.

3.4. Application of the framework adds ethical value

Participants described the workshop's design outcomes as justified, reasonable and well-considered. One explained that the framework is "a systematic way to capture stuff that you really want. I mean…this is a totally different outcome I think." Many of the participants had participated in various design workshops but expressed that the workshop provided different insights and allowed them to approach a problem with different considerations. As one participant explained, "I've been involved in stakeholder analysis in my past, it's after working here [with] a framework, especially calling all tensions, and a separate step to figure out the compromise and tradeoff was really good, because we always talk about different stakeholders, and then we may pick one element of one group, and then implicitly prioritize one over the other. But, actually making each topic as a step, made us realize what they offered and weigh more, what competence they were having, right? When somebody asked them why, which I think is a much better process." Stepping through the process with the workshop participants seems to have provided structure, but also room to creatively approach and think through the problem. In describing the results of the framework one participant noted, "it seems like [a] good engineer ultimately can get more creativity and bigger, better solutions." Overall, the responses were very positive and demonstrated how the workshop added ethical value to solving this particular engineering problem but also could be applied to others to generate solutions that focus on a broad set of stakeholders, the underlying value tensions and well-considered tension resolutions.

3.5. Task hesitation in carrying out the design task

Researchers also found that participants identified areas where there was a lack of conceptual or role clarity or lack of expert guidance, which led to task hesitation or discomfort in carrying out the design task. This is a relatively broad theme as it captures moments when participants expressed having

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

struggled with the workshop and framework, sometimes because it was making them think differently and they needed more conceptual support. Other times participants made it clear that the workshop prompts or process could be improved in order to help overcome a lack of task clarity or role clarity within the overall framework.

An example of conceptual difficulty was explained by a participant, "I thought it was difficult to come up with many values, it was easier to come up with multiple stakeholders, for each stakeholder what values you associate with them, that was really difficult…personally I struggled with that." This confusion around the nature of values was identified early in our framework piloting phase, which prompted us to encourage participants to describe stakeholders' "goals and values". The fact that participants continue to struggle with this concept suggests more work is needed to develop clarity around the concept.

In the beginning of the workshop, task hesitation stemming from a lack of role clarity was apparent. One participant recalled, "I just caught myself trying to figure out, like, you know, the affiliation from your [name tag]." But the participants did not have job titles or companies listed on their name tags. She explained that they normally would "try to delegate certain types of exercise, or, like, thinking exercises to a certain person. Like, 'oh, you would know about this, and you would know about that.' As opposed at the beginning, he was quite anonymous. We don't [all] know what [the others] do. So, it was more, sort of a collective, and even effort."

Another area where the participants felt there could have been more guidance provided during the workshop was narrowing down stakeholders when deciding which ones to focus on as they moved through the process, explaining "there was no logical process to narrow down the stakeholders, we kind of worried about personal morals or bias, but there was [no] way to figure out which of the stakeholders we should consider first." Additionally, participants would have liked to see more "examples of the differences between at least compromise and trade off, and maybe more examples of dissolving a tension, just to give us more to go on, too." We plan to address this feedback in a future iteration of the framework and workshop.

4. Discussion

The purpose of the study was primarily to gauge the effectiveness of the framework as a means for equipping engineers to identify ethical challenges in their design tasks and apply ethical considerations in the design of autonomous vehicles. Throughout the study, especially during the final interview, it became clear that applying the framework was successful in building some ethics capacity among engineers. The participants continuously spoke about how useful the workshop was in guiding them through the design process. In particular, helping them come up with a list of stakeholders and the values they hold. In the pre-workshop design activity, some of the teams wrote down values that they were considering when designing how the autonomous vehicle should approach the crosswalk. Some groups even included stakeholders (pedestrian and vehicle occupant) but the list was substantially shorter than the one generated once the framework was applied. The ability to generate a list of people affected by a technology and then consider their values are useful in equipping the design engineers with the tools they need to identify ethical challenges. Being able to consider multiple perspectives and how different peoples' values conflict is the first step in being able to design ethically.

Overall, participants considered the framework a useful tool. They brought up different ways that they would apply it in their everyday work. Some suggested running the workshops with multiple AV companies and others liked the idea of taking pieces of the framework and applying it in "bite sized pieces" so that it can be a quick exercise to get a variety of people thinking about a design problem. These types of comments demonstrated that the framework provides a structure the participants found useful enough to want to apply to their everyday jobs. This was one of the other questions that the research aimed to answer: how useful do engineers consider the framework? The results of the workshop illustrated that the framework did in fact provide a useful structure that could be easily transferred back to the workplace and used for various design challenges.

However, it is important to note that there were some aspects of the framework with which participants would have liked more concrete guidance. In particular, the participants would have liked more guidance in narrowing down the stakeholder list. The participants came up with lists in their groups and then came together as a whole and made an even larger list. After that, the researchers

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press

asked them to pick three of the stakeholders. They could have picked three based on which they felt were most relevant or interesting or important or according to various other metrics. Though the lack of guidance in this area was intentional to encourage participants not to pick the most obvious stakeholders (thus encouraging more varied design considerations), it does make sense that the participants would want more guidance in narrowing down the stakeholder list because the lack of guidance forced them to make uncomfortable decisions. Still, practice making uncomfortable, yet justified, ethical design decisions could be a good skill in that it could empower engineers with the capacity to make those decisions independently in their workplaces.

The workshop answered the initial questions that the researchers proposed when developing the study: the framework appears to be a useful tool in helping engineers anticipate ethical issues in design, while providing enough structure that participants considered relevant and helpful enough to use in their everyday work. Hopefully, more designers and engineers will use the framework when approaching design problems, enabling them to think and design in response to the ethical challenges their technologies raise.

Acknowledgement

The researchers thank the Center for Automotive Research at Stanford, the McCoy Family Center for Ethics in Society (Stanford), and the Social Sciences and Humanities Research Council (SSHRC Canada) for providing critical funding and support for this research. In addition, we thank the amazing staff at the Dynamic Design Lab (Stanford) for organizing, and the Hasso Plattner Institute of Design (d.school) for hosting, the workshop.

References

Friedman, B., Kahn, P.H. and Borning, A. (2008), "Value sensitive design and information systems", In: Himma, K.E. and Tavani, H.T. (Eds.), *The handbook of information and computer ethics*, Wiley, Hoboken, NJ. pp. 69-101.

Friedman, B. and Kahn, P.H. Jr. (2003), "Human Values, Ethics, and Design", In: Jacko, J.A. and Sears, A. (Eds.), *The human-computer interaction handbook*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 1177-1201.

Gerdes, J.C., Thornton, S.M. and Millar, J. (2019), "Designing Automated Vehicles Around Human Values", In: Meyer, G. and Beiker, S. (Eds.), *Road vehicle automation 6, AVS 2019, Lecture notes on mobility*. Springer, Cham, pp. 39-48.

Kenner, S. (2016), *Apple's Comments on the Federal Automated Vehicle Policy*. [online] November 22, 2016. Medium.com. Available at: https://www.regulations.gov/document?D=NHTSA-2016-0090-1115 (Accessed: 13 Nov 2019).

Kyrouz, M. (2018), *Public Comments on NHTSA's 2017 Automated Driving Systems: A Vision for Safety*. [online] Medium.com. Available at: https://medium.com/smart-cars-a-podcast-about-autonomous-vehicles/public-comments-on-nhtsas-2017-automated-driving-systems-a-vision-for-safety-2-0-526ca67b1955 (Accessed: 30 Oct 2019).

Millar, J. (2016), "An Ethics Evaluation Tool for Automating Ethical Decision-Making in Robots and Self-Driving Cars", *Applied Artificial Intelligence*, Vol. 30 No. 8, pp. 787-809.

Moon, A.J. et al. (2019), *Foresight into AI Ethics (FAIE): A toolkit for creating an ethics roadmap for your AI project*. [online] Open Roboethics Initiative. Available at: http://www.openroboethics.org/ai-toolkit (Accessed: 4 November 2019).

NHTSA (2016a), *Federal Automated Vehicles Policy: Accelerating the Next Revolution in Road Safety*. [online] U.S. Department of Transportation. Available at: https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf (Accessed: 30 Oct 2019).

NHTSA (2016b), *What is NHTSA's Approach to Ethical Considerations*. [online] NHTSA.gov. Available at: https://www.nhtsa.gov/vehicle-manufacturers/automated-driving-systems#automated-driving-systems-faq (Accessed: 30 Oct 2019).

Vallor, S. (2018), *An Ethical Toolkit for Engineering/Design Practice*. [online] Markkula Center for Applied Ethics at Santa Clara University. Available at: https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit (Accessed: 2 November 2019).

Verhalen, J.K. (2016), *The National Society of Professional Engineers' Public Comment on Docket ID No. NHTSA-2016-0090-0001*. [online] National Society of Professional Engineers. Available at: https://www.nspe.org/sites/default/files/resources/pdfs/NSPE-Public-Comment-NHTSA-2016-0090-0001.pdf (Accessed: 30 Oct 2019).

https://doi.org/10.1017/dsd.2020.78 Published online by Cambridge University Press