

A Method for Rapid Ethical Deliberation in Research and Innovation Projects

Marc Steen, TNO, The Netherlands

Martijn Neef, TNO, The Netherlands

Tamar Schaap, TNO, The Netherlands

ABSTRACT

There are many and diverse methods available that can help researchers, designers, developers, and engineers integrate ethics in their research and innovation projects. In practice, however, they can find this challenging. They may believe that such methods are difficult and time-consuming, or that ethics hinders innovation and creativity. Borrowing from various methods, the authors created a lightweight method that researchers and designers can use to integrate ethics in their projects: rapid ethical deliberation. This research collaborated with four projects to assess this method in practice. The authors found that this method helped project team members in several ways: to envision the innovations they work in very practical terms; to look at these innovation from different normative, ethical perspectives; to look at their projects with fresh eyes and engage in creative and strategic thinking; and to articulate critical questions and associated actions to move their projects forward.

KEYWORDS

Applied Ethics, Consequentialism, Creativity, Strategy, Duty Ethics, Relational Ethics, Virtue Ethics

INTRODUCTION

There is a need for better ethical deliberation within the domain of research and innovation. In the past year, for example, we have seen haphazard efforts to develop mobile apps and data dashboards in countries around the globe to monitor and combat the spread of Covid-19, trying to balance various and often conflicting values, needs and interests; we have seen further misuse of social media to monitor and influence people's behaviours in democratic processes; we have seen increased usage of surveillance and face recognition, sometimes with publics that seem to be largely unaware of the enormous impacts. In such a world, we do need ethical deliberation in research and innovation.

Fortunately, there are many and diverse methods available for '*including, integrating or incorporating ethics*' in research and innovation projects (Reijers et al., 2018: 1438). Reijers *et al.* (2018) distinguish between three categories: 1) *ex ante* methods, which are used *before* projects start or in the early of research an innovation, 2) *intra* methods, which are used *during* research, design, development and implementation, and 3) *ex post* methods, which are typically used *after* projects have delivered results.

We (the authors) work at TNO, a Research and Technology Organization of some 2600 people with a mission 'to create innovations that boost the sustainable competitive strength of industry and well-being of society'¹. TNO has a societal mission and is funded in part by the national government, in part by industry, and in part by international research funding. We view innovation as a process

DOI: 10.4018/IJT.2021070106

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

of bringing about positive change in society through the application of new ideas or technologies, and distinguish between four phases: 1) research, 2) design and development, 3) implementation and application, and 4) usage. TNO employees mostly work in the early phases (research, design and development), but some are also active in later phases (implementation, application and usage). Many projects of TNO are organized in a quadruple helix (Carayannis & Campbell, 2009) (i.e. they involve government, industry, knowledge and civil society organizations). Many projects of TNO deliver demonstrators or prototypes to other parties, which they can further develop and integrate in their systems, services or products. Therefore, TNO and its research and innovation efforts are often part of a larger innovation eco-system (Granstrand & Holgersson, 2020) since its outputs are inputs for other parties' innovations.

We set ourselves the goal to develop and test a method that would support colleagues in improving their capabilities to integrate ethics in their projects, especially regarding their projects' results and outcomes, and their impacts in society². We are therefore particularly interested in what Reijers et al. referred to as *intra* methods (2018: 1450-1452)³; methods that researchers and designers can use themselves and which can help them to steer their projects.

We believe that many researchers and designers would be willing and able to integrate ethics in their projects, but are currently lacking methods that are easy to use and that deliver practical results. From years of informal observation we get the impression that many people in research and innovation find the idea of 'doing ethics' challenging; they expect that ethics is time-consuming without it offering measurable impact. Additionally, they perceive ethics as fuzzy, complex or abstract. Moreover, they expect that integrating ethics in their projects will limit, hinder, or obstruct creativity and innovation.

Given our goal to promote the integration of ethics in research and innovation, we need a method that: 1) is practical, 2) that delivers practical results and provides clarity, 3) that fits in current working processes and helps the people involved to move their projects forwards, and 4) that promote creativity and innovation. There is evidence that engagement with ethics can promote creativity. Feng et al. (2018) found that increased ethical leadership promotes employees' creativity; this trend continues until a threshold is reached—after that, employees feel micromanaged and restricted, and their creativity decreases. We feel that the implementation of a less rigid method to apply ethics to innovative projects will inspire new approaches and research ideas. This need for user-friendly, useful and lightweight methods to integrate ethics in research and innovation is salient. This is especially true for certain disciplines such as 'AI ethics' (Hickok, 2020), where people look for ways to go beyond abstract principles—to move from theory to practice, or, as Morley et al. states: 'from what to how' (2019).

We are familiar with various *intra* methods to integrate ethics in research and innovation like *Ethical Parallel Research* (Van Gorp & Van der Molen, 2011), *Value Sensitive Design* (Friedman, Kahn, & Borning, 2006), *Ethical Impact Assessment* (Jansen et al., 2017), and *Disclosive Ethics* (Brey, 2000, 2010) and used elements from these methods (see Reijers et al (2018) for a more complete overview and more extensive discussion of these methods. We will briefly discuss these four methods.

In *Ethical Parallel Research* (Van Gorp & Van der Molen, 2011) ethicists collaborate with researchers and developers, in a five-step process, with the ethicist in the lead: gather information, reflect on ethical issues, prepare a discussion with researchers and developers, have the discussion and make decisions, and report about ethical issues and decisions made—this collaboration typically takes place over a period of time. *Value Sensitive Design* (Friedman et al., 2006) requires involving relevant stakeholders and organizing a relatively elaborate 'tripartite' methodology, with three stages: a conceptual stage in which relevant values and relevant stakeholders are identified and studied, an empirical stage in the ways in which values are (not) embedded in current or future design choices are studied, and a technical stage in which the ways in which properties of the technology at hand hinders or promotes relevant values. *Ethical Impact Assessment* can be understood as one element in a more comprehensive effort to integrate ethics in various stages and elements of research and innovation (Jansen et al., 2017: section 5) and can consist of six stages: threshold analysis, formulate a plan, identify ethical impacts, evaluate ethical impacts, and formulate and implement remedial

actions—with multiple pages of items and questions for each stage. Finally, *Disclosive Ethics* (Brey, 2000, 2010) promotes the ‘moral deciphering of embedded values and norms in computer systems, applications and practices’ (Brey, 2000); it recognises that ‘[m]any practices involving computer technology are morally opaque’ (Brey, 2010: 49)—in other words, it assumes that the people involved *implicitly* participate in *some* form of ethical deliberation (Steen, 2015), and invites and enables them to discuss these ethical issues *explicitly*.

While these methods in themselves are very valuable, we believe that in their current form, with their current depth and width, researchers and designers find it difficult to actually use them in practice. Rather, they need a usable, lightweight method. This is in line with Reijers *et al.* first recommendation to focus *intra* methods ‘on the integration of ethics in the day-to-day work of [research and innovation] practitioners, especially with the disclosure of ethical issues in design’ (2018: 1457).

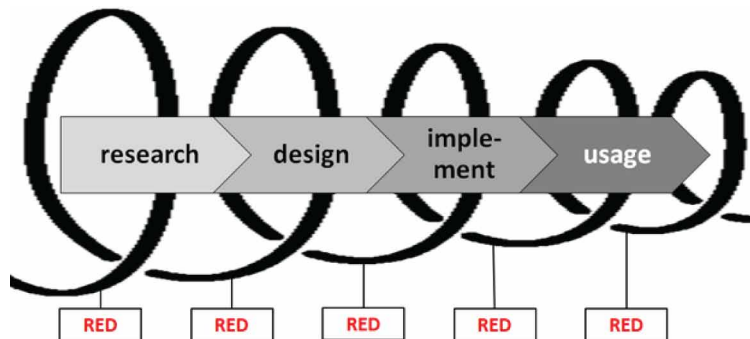
Rapid Ethical Deliberation (RED)

Based on the methods mentioned above—*Ethical Parallel Research*, *Value Sensitive Design*, *Ethical Impact Assessment*, and *Disclosive Ethics*—and based on years of experience with efforts to integrate ethics in research and innovation projects (e.g., Royakkers & Steen, 2017; Steen & Van de Poel, 2012), we developed a method: *Rapid Ethical Deliberation* (RED)⁴. We hope that researchers and designers find this method usable, useful and can actually use it in their projects. Moreover, we aimed to develop a method for ethical deliberation that also promotes creativity and also helps the people involved to reflect on their project more strategically. To that end, we took inspiration from *Design Thinking* (Cross, 2011; Dorst, 2015; Dorst & Royakkers, 2006), an approach to innovation that involves organizing an iterative and participative process of problem-setting and solution-finding—a process that, for instance, enables participants to move fluently from exploring possible solutions to re-examining and re-thinking and modulating the project brief, a process of joint inquiry and imagination (Steen, 2013a). We wanted to move away from the stereotype of ethics as a barrier to innovation.

Furthermore, *Design Thinking* enables participants to reflect on their project and to zoom-in and zoom-out their perspectives: to envision problems and potential solutions in very practical terms rather than having abstract discussions about abstract ideas; and also to question their project and talk to clients and stakeholders to question assumptions and explore new directions when that is needed. Moreover, we intend RED to be a method for Responsible Innovation, that is, to promote anticipation, reflexivity, inclusion and responsiveness within innovation projects (Stilgoe, Owen, & Macnaghten, 2013). Regarding inclusion and responsiveness, we recognize that RED would benefit from further development to include stakeholders in ethical deliberation, e.g., civic society organizations (Steen & Nauta, 2020), so that the dialogues would include voices from society, from ‘outside the lab’ (Doorn, Schuurbiens, Van de Poel, & Gorman, 2013). Such further development, however, is outside our current scope (also, it would introduce an entirely new set of questions, related to power and participation: which are relevant stakeholders, which are prominent values, etc.).

Importantly, our method is based on ‘doing ethics’ in the form of organizing an iterative and participative process of ethical deliberation. Our method focuses on *asking questions and facilitating creativity*, rather than, for example, looking for definitive answers or following an ethics checklist (Hickok, 2020). Our method invites participants to adopt four different and supplementary normative ethical perspectives: consequentialism, deontology, virtue ethics, and relational ethics⁵. These are (roughly) the same four traditions that Van de Poel and Royakkers (2011) follow in their textbook. The mobilization of different normative perspectives is a distinct feature of the RED method. Its purpose is to enable participants to clarify, discuss, navigate, prioritize and balance conflicting values. This follows the second recommendation of Reijers *et al.* for *intra* methods: to base considerations ‘on a normative theoretical framework that explicates how certain technology design choices can be identified as ethical, or how “ethics” is mediated by technology design (2018: 1457). This would address a problem that some see in methods like ETHICS (Stahl, 2007) or VSD (Jacobs & Hultgren, 2018), namely that they lack ways to discuss normative directions or establish ethical commitments.

Figure 1. Rapid Ethical Deliberation can be applied in different phases of research and innovation, preferably as an integral part of an iterative process



While we are certainly *not* looking for a method that provides definitive answers or prescribes ‘good or evil’, we *are* looking for a method that gives participants some handholds and vocabulary to discuss ethical issues also normatively.

Moreover, we propose that RED can be a valuable tool in different phases of research and innovation, especially if it is used on different moments as part of an iterative research and innovation process—see Figure 1. RED can be useful in the *early phases*, for example, in project preparation and definition or the ‘fuzzy front-end of innovation’ (Gassmann & Schweitzer, 2014) (as an *ex-ante method*), in design and *development*, for example to explore, envision, design, or evaluate different solutions or options (as an *intra method*), and in the *end of a project*, for example, to evaluate projects’ results (as an *ex-post method*)—where it is critical to understand, in the context of an Research and Technology Organization (RTO), that its results can very well be input for other parties’ innovation processes; in that sense, an *ex-post* evaluation for the RTO is an *ex-ante* assessment for those other parties.

Please find below a first version of the *Rapid Ethical Deliberation* (RED) method and workshop format:

- **Intake:** Several days before the workshop (below) a meeting of approximately 30 minutes, possibly online⁶, with one or two people of the research and innovation project (‘project team members’) and one or two people who will facilitate the process (‘facilitators’). The goal of the ‘intake’ is twofold. Facilitators learn about the context, goal and scope of the project, and have some between ‘intake’ and ‘workshop’ to learn more. Project team members learn about the RED method, including the division of roles: project team members are responsible for the content, including follow-up actions; whereas facilitators are responsible for the process (and can, optionally, be involved in follow-up actions).
- **Workshop:** A meeting of approximately 60 minutes, possibly online, ideally with the same people: one or two project team members and one or two facilitators⁷.

First, the project team members describe their project’s main goal and frame it in terms of design thinking (problem-setting and solution-finding). The facilitators can ask questions like: In one or two sentences, what problem does your project aim to solve?, What is the project’s overall goal?, or Can you walk us through one of the solutions you are thinking about or working on, in very practical terms?

Then the facilitators invite the project team members to identify potential ethical issues, for example, regarding the project’s potential outcomes and their impacts in society, both desirable and undesirable, or in terms of values that are at stake, that the project aims to protect or that may be at risk. They can draw from lists of values, such as the list of Friedman et al. (2006), who

mention human welfare, ownership and property, privacy, freedom from bias, universal usability, trust, autonomy, informed consent, accountability, courtesy, identity, calmness, and environmental sustainability, or the list of the European Commission's High Level Expert Group on AI (2019), who mention respect for human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability.

Then follows an interactive discussion about the issues that were brought up. Participants are invited to discuss these issues using four different ethical perspectives:

Consequences and impacts:

- *What are potential positive consequences or impacts of this innovation and its implementation? These can happen in processes or services (internal) or in society and people's daily lives (external).*
- *What are potential negative consequences or impacts of this innovation and its implementation? These can happen in processes or services (internal) or in society and people's daily lives (external).*
- *How are positive and negative consequences and impacts divided over different people or groups?*
- *What could be possible unintended and undesirable side effects of this innovation?*

Duties and rights:

- *Does the organization that wants to implement this innovation have any duties or mandates relating to the innovation? If so, what are these duties or mandates?*
- *Do you or do we have any duties on behalf of this organization? If so, what are these duties?*
- *Does this innovation impact on people's fundamental rights, for example, human dignity, freedom, equality, solidarity (Charter of Fundamental Rights of the European Union).*
- *Are there (informal) rules, regulations or legislations which this innovation needs to comply?*

Virtues and flourishing:

- *Looking at the people who would use this innovation, which virtues (or dispositions) are relevant? Think of classical virtues like justice, courage, self-control and practical wisdom, professional virtues like watchfulness or helpfulness (for police officers), or 'technomoral' virtues like honesty, humility, care, civility, flexibility and perspective (Vallor, 2016).*
- *How can this innovation help (or hinder) them to cultivate or express these virtues? Or which virtues are at stake or could be at risk when they use this innovation?*
- *And how can this innovation help them to find appropriate 'means' for relevant virtues? What type of learning or training would be beneficial or necessary?*
- *Zooming-out to the level of society, how can this innovation promote people's flourishing and wellbeing?*

Relationships and interactions:

- *Which relationships, interactions, or treatments would be affected by this innovation?*
- *Think of different relationships and interactions, e.g. between people who use the innovation and people who are affected by it (e.g., police officers and citizens, respectively)*
- *In what ways could these relationships, interactions, or treatments change? For better or for worse?*

- *How could the innovation affect the quality of relationships between people, groups of people or organization? Think of power differences, communication, collaboration, empathy and care.*

After these discussions, the facilitators invite the project team members to articulate conclusions, for example, insights that they developed during the workshop, and also questions that they will want to further explore and address and actions that they will want to do. This last step is key in that it enables the participants to articulate what they can do to move their project forward. It provides both closure, looking back at the workshop, and opening, looking at their project with curiosity and creativity.

- **Reflection:** *It is recommended to organize a third step, in which participants and facilitators can reflect on the workshop, both regarding content and regarding process.*

As we (the authors) were looking for a method that would fit in an organization that is typically not very accustomed to organizing elaborate ethical deliberation, we developed RED to suit researchers and designers who may have little experience with integrating ethics in their work, who have busy schedules. The RED workshop functions as a *pressure cooker* and as an *appetizer*: if participants feel it useful, they can organize more elaborate ethical deliberation activities. Moreover, RED is designed as a flexible tool. It can be applied with more or less structure (or improvisation) and with more or less depth and width, depending on the context and its restraints, and the participants and their skills.

RESEARCH APPROACH AND QUESTIONS

Our research approach is explorative and descriptive. First, we wanted to explore what happens when we organize RED workshops, both regarding content and process. Second, we were especially interested in whether RED workshops can promote creativity, both in terms of enabling fresh perspectives and generating creative ideas or solutions, and in terms of enabling critical reflection on the project more strategically—notably in terms of moving iteratively between solution-finding and problem-setting.

In order to address these questions, we organized a series of four RED workshops with participants from four different research and innovation projects. For the first research question, we looked at the proceedings of the workshops and described both content and process. For the second research question, we focused on ‘aha’ moments; moments in which participants looked at the situation, problem, or possible solutions in creatively and generated new, creative or strategic ideas or options.

In order to find researchers and developers who could help us understand RED’s applicability, we approached colleagues from different projects and invited them to participate in an intervention in which we collaboratively apply this method. We targeted and selected projects in our organization which (we expected) may benefit from ethical deliberation, and found four. All four projects are concerned with research and innovation in the domain of public safety and police work, and focus on communication, sharing information and collaboration. While these projects are conducted in close collaboration with customers and other stakeholders, we involved only researchers and designers from our own organization in the RED workshops (see our remark above about involving stakeholders).

We conducted this study as *action research* (Guertler, Kriz, & Sick, 2020; McPhee, Hoppe, & Lindhult, 2019; Ollila & Yström, 2020) by organising and executing intake and workshop meetings with the project team members. We made notes of the meetings, had these reviewed by the participants, and then discussed these notes amongst the authors. Our approach employs a *case study* research design (Yin, 1994) with four cases⁸. We justify our relatively small number of cases in that we view these cases as ‘revelatory cases’ in that we had a relatively unique ‘opportunity to observe and analyze a phenomenon previously inaccessible to scientific investigation’ (Yin, 1994: 40). We

were able to do these workshops and test the RED method because we were ‘insiders’ and already had trust and rapport with the people we engaged with. Notably, our approach can be understood as *social-constructionist* (Easterby-Smith, Thorpe, & Lowe, 2002). Instead of discovering some objective truth, we were interested in the ways in which project team members express and discuss issues and ideas within their project; we were interested in their (subjective) experiences, embedded within their specific and practical contexts.

Case Studies

In this section, we provide observations from the intake and workshop meetings with team members of four research and innovation projects. We focus on significant (‘aha’) moments and supply some background information on the projects, which can help to better understand and these moments. Please note that the workshops were with co-workers in research and innovation roles; *not* with people from other organizations, for example, from the police.

Case A: *Pedo-Hunters and Civil Arrests*

This case is concerned with citizens who track down putative pedo-sexual individuals⁹, mostly online, for example, in private chat rooms, and make appointments with them to meet physically. They are referred to as *pedo-hunters* (*pedo-jagers* in Dutch); a diverse group that includes concerned citizens who want to stop pedo-sexual individuals, sometimes with very personal motivations, and vigilantes, who might use violence against putative perpetrators when they meet them. The police is troubled by the work of pedo-hunters, especially those who meet alleged pedo-sexual individuals and conduct *civil arrests*, because these meetings often lead to violence and vigilantism. Citizens are legally allowed to conduct civil arrests when they catch suspects ‘red-handed’, provided that they transfer suspects to the police immediately and use force only when necessary and proportionally. We discussed this case from the perspective of the police, with the aim to envision new, better policies or processes. The discussions of *duties* and of *virtues* were especially useful and sparked several ‘aha’ moments.

Many pedo-hunters act from feeling a *civil duty* to combat pedo-sexual individuals’ harmful and criminal behaviours. In addition, the police have a formal duty to protect citizens and to combat crime. We focussed on civil arrests. One of the problems is that citizens can only *reactively* call the police: when the suspected pedo-sexual arrives at the agreed-upon location. In the time that it takes for the police to arrive, however, this meeting between pedo-hunters and suspects can often turn violent. One solution would be that citizens can *premeditatively* ask for police assistance; for example, by planning such a meeting with the suspect in collaboration with a police officer. There is, however, currently no way to collaborate in this manner. Consequently, citizens who plan on conducting civil arrest are left to do so on their own at the potential expense of their safety or the safety of others. A discussion of citizens’ *informal* duties and the police’s *formal* duties led to the idea to organize *better handovers from civil arrests to police arrests*. The police’s formal duties would then match and supplement the informal duties of citizens. We also discussed virtues. Justice as an overall virtue that is needed in such situations. In addition, police officers would need to cultivate courage and avoid its deficiency, that leans towards passivity, when conducting such arrests, and citizens would need to cultivate courage and avoid its excess that leans towards rashness. Another idea for this solution would be that the police organizes trainings for conducting these new types of civil/police arrests properly. A key result of this discussion was that one team member will discuss this idea with the police, to explore ways to modify processes and enable civil/police arrests with risks for violence and vigilantism.

Case B: *Dashboard for Emergency Room Operators*

During the minutes it takes for police officers to respond to an emergency call (112), the police may want to collect additional information about the incident, such as backgrounds of the people involved. For this task, the police created a Real-Time Intelligence Centre (RTIC). RTIC operators use diverse data sources, both within the police and in the public domain, such as social media, interpret these

data, and relay relevant information, via the emergency room operator, to the police officers who rush to the incident. This case is concerned with the police's exploration of options to automate parts of this process in the RTIC. One option is to use algorithms that link additional data to a case during an information search. The ambition is to improve effectiveness and speed, while preventing risks of bias, such as not taking into account alternative hypotheses in a case or selectively choosing databases from which to search. We discussed this case with the goal to explore ideas for functional requirements for such a system and for procedures for the emergency room operators. Discussions of *consequences*, *virtues* and *relationships* gave rise to several 'aha' moments.

An exploration of positive and negative consequences led to a discussion about the need to enable emergency room operators to become (more) aware of their own biases and search strategies, so that they can avoid following the system's output unthinkingly. Instead, more awareness may lead to critical thinking, which could lead to more efficient and effective police work. In addition, an exploration of virtues led to a discussion of the need for operators to act both swiftly and *decisively*, and carefully and *thoroughly*. They need to find appropriate means for, and balances between, these virtues; they need to work *decisively* within restraints of carefulness, and *thoroughly* within restraints of swiftness. They need to strike appropriate balances in different situations, such as street fights, cases of domestic violence, missing persons cases, or traffic accidents. To address this, project team members developed the idea that the system could support operators to 'wisely' allocate on different tasks: *searching* information, *interpreting* information, and *communicating* to police officers on the street. Finally, the project team members explored different relationships and speculated that the operators need to provide information that is both *trustworthy* and *actionable* for the police officers they relay it to. As a result of this discussion, the project team will further explore ways to make the system a tool for promoting critical thinking and, for example, use insights from psychology and human-computer interaction to create different 'information profiles' for different roles.

Case C: Neighbourhood Watch Groups and Early Detection

Citizens have become increasingly active in *Neighbourhood Watch* groups, where they share information and offer each other support, typically using WhatsApp as a medium. Some police officers collaborate with such groups, for example, by joining *Admin Groups*, which consist of administrators of different *Neighbourhood Watch* groups. Administrators can choose to relay messages from 'their' *Neighbourhood Watch* group to an *Admin Group* and, if needed, collaborate there with this police officer.

This case is concerned with an idea for an innovation in which the police can monitor *Neighbourhood Watch* groups automatically and for the 'early detection' of incidents, that is, *before* somebody calls the police. Imagine that a citizen writes about a 'suspicious van in our street' in a *Neighbourhood Watch* group. If the police were monitoring this group, an algorithm from their Real-Time Intelligence Centre (RTIC) would detect that. An RTIC operator might then call the concerned citizen to ask for details, such as a license plate. This citizen, maybe with help from others in the group, relays this information to the police, where the RTIC operator can search for additional information. If they find anything suspicious, they can send a police officer to further investigate. We discussed this case with the aim to envision new processes for RTIC operators. The discussions of *duties* and *rights*, and of associated *virtues*, were especially useful and caused several 'aha' moments.

Currently (without such 'early detection'), the police's duty to respond to citizens' emergency calls is relatively clear: they have a duty to be accessible for emergencies (via 112), and to respond adequately and timely. However, for 'early detection' their duties are less clear. A key question emerged: *What duty or task of the police would this 'early detection' be part of? Law enforcement, criminal investigation, emergency relief and assistance, communication and prevention?* These tasks have different mandates and processes. In very general terms, the police need to be *watchful* and *helpful*¹⁰ and protect citizens' safety. Another key question arose about these *Neighbourhood Watch* groups: are they *private* spaces (because WhatsApp is owned by US corporation Facebook) or *public* spaces? This

distinction affects the police's duties of citizens' rights¹¹. We also discussed several relevant virtues for both police officers and citizens in *Neighbourhood Watch* groups. Finding appropriate means for courage is required during incidents. Another relevant virtue is self-control; both police officers and citizens need appropriate restraint in their usage of social media since conversations on social media quickly escalate and polarize. A key result of this discussion was that project team members plan to discuss with the police about where such 'early detection' would fit in the police's tasks and processes; for example, an RTIC operator might be able to classify each incident, as enforcement, investigation, assistance or prevention.

Case D: Risk Assessment for Subversive Criminality

A starting point for this case is the Multi-Agency Vulnerability Assessment Support Tool (MAVAST). This tool was developed to enable experts to assess individual vulnerability for violent radicalisation and identify potential interventions to prevent or mitigate violent radicalisation¹². This current case involves exploring the possibility of re-using elements from this tool in another project in another domain; namely for the assessment of risks for young people to slide towards subversive criminality—types of criminality in which legal and illegal activities are blended, where people in business or public administration become complicit in criminal activities. Several decisions went into the development of MAVAST, which will be upheld: MAVAST does *not* give numerical scores (e.g., between 0 and 1, as is typical), instead it displays the underlying indicators, both available and missing. As a result, professionals who use MAVAST will need to actively interpret its output, which avoids the risk people depending on an algorithm without using their expertise, knowledge of the situation, and critical thinking. We discussed how a Regional Information and Expertise Centre (RIEC) could use this innovation. RIECs facilitate sharing information¹³ between organizations in order to combat subversive criminality, for example, between public prosecution, police, municipalities, provinces, inspection agencies, and Youth Services. The discussion in the workshop aimed to explore ways to modify and utilize a tool like MAVAST in this other context. Our explorations of *consequences*, of *duties*, and of *relationships* gave rise to several 'aha' moments.

The positive outcomes of using such a tool would be that young people are prevented from sliding towards subversive criminality. A discussion arose about the project's key underlying assumptions. Is the tool meant to identify 'vulnerable people' and offer them support, or is it meant to identify 'potential perpetrators' (who may actually be victims of criminality or adverse circumstances) and to repress or arrest them? Different partners within a RIEC may have different views on this. It also became clear that a tool for risk assessment only makes sense within a wider context of risk *management*, with adequate resources and competencies. Otherwise there is a risk for the tool to become a gadget, offering the illusion to users of having data or being in control, without actually promoting agency. Further discussion will be needed to unpack the impacts of such a tool on various relationships and collaboration, notably between frontline professionals and potential perpetrators, between users of the tool and the experts at RIEC, and between different organizations and people within RIEC. A key outcome of this workshop was the project team members' plan to talk to their stakeholders about underlying assumptions, e.g., in terms of vulnerability and support 'versus' criminality and repression, and about ways to align different goals.

DISCUSSION AND CONCLUSION

We developed a method for *Rapid Ethical Deliberation*, which people can use in their research and innovation projects. We applied this method in four projects and studied what happened, regarding both content and process. Based on our observations, we found that in all four cases it helped the project team members in several ways¹⁴:

- To envision the innovation that they work on very *practical* terms and to envision its practical implementation and usage, and associated potential impacts in very practical terms.
- To look at their project and the innovation they work on from different ethical perspectives, which enabled them to express, navigate and balance different normative orientations.
- To look at their projects with fresh eyes ('aha' moments); this sparked their creativity and helped to look at their projects strategically, for example, to ask questions about key assumptions.
- To articulate questions that they can further investigate or discuss with others, notably with clients; this helped them to experience the workshops as productive, and move their projects forward.

Looking at these results, we propose that RED is usable and useful for researchers and designers. This contrasts with some other methods for integrating ethics, which are associated with fruitlessly discussing relatively abstract ideas and with hindering creativity and stifling innovation. We were happy to see several 'aha' moments during the workshops: moments in which project team members took on fresh perspectives, generated creative ideas or asked fundamental questions—see Table 1.

In Case A, the idea emerged to combine citizens' (informal) duties and police officers' (formal) duties and develop ways to better organize civil/police arrests, rather than accepting that there is currently no process. In Case B, the idea was developed to make the system promote and facilitate RTIC operators' abilities for critical thinking, so that they can use the system responsibly. In Case C, the need to clarify 'early detection' vis-à-vis different police tasks became clear, rather than leaving that key question unaddressed. In Case D, the need to discuss the overall goal of this risk assessment (vulnerability and support 'versus' criminality and repression) became clear, rather than letting this key question unexamined.

Based on these cases and feedback that we received from participants about RED, we can answer our second research question: that ethical deliberation can promote creativity and strategic thinking. Ethical deliberation can enable researchers and designers to steer their research and innovation projects—rather than functioning as a barrier to innovation, which people tend to associate ethics with. Sometimes researchers and designers focus too much on practical situations and on current restrictions so that it is difficult for them to be creative or reflect critically and strategically on their project.

Looking back at some of the methods that RED draws from, we can see that it enables researchers and designers to engage in ethical deliberation: similar to *Ethical Parallel Research* (Van Gorp & Van der Molen, 2011), it enables participants to discuss values and interests of different stakeholders, similar to *Value Sensitive Design* (Friedman et al., 2006), it enables them to identify and evaluate ethical issues and find ways to deal with these, similar to *Ethical Impact Assessment* (Jansen et al., 2017) and *Disclosive Ethics* (Brey, 2000, 2010), it enables them to make implicit values more explicit and to discuss these critically. RED does a little bit of all of this, in relatively short time, and with less depth and less breadth.

Table 1. Several 'aha' moments (indicated 'x') occurred during the workshops for the different projects

	Case A: Pedo-hunters and civil arrests	Case B: Dashboard for RTIC operators	Case C: Neighbourhood Watch groups and early detection	Case D: Risk assessment for subversive criminality
Consequences		X		X
Duties and rights	X		X	X
Virtues	X	X	X	
Relationships		X		X

Moreover, we would like to point out that although a method like RED enables people to make decisions and move their project forward, these decisions must *not* be understood as final. Rather, these decisions are part of an iterative process of research and innovation, and therefore amenable for validation and improvement. RED is best used in a context of experimentation and learning (van de Poel, 2016, 2018).

Looking ahead, we would like to propose that other people can try-out the RED method, preferably in different stages of research and innovation. We hope that the RED method is flexible regarding the balance between structure and improvisation in that it provides enough structure to organize a workshop, and enough room for improvisation to adopt the workshop to the participants and context. We can imagine variations that offer more depth, for example with elaborate analyses of consequences, duties, virtues and relationships, and more width, for example, through engagement of more diverse people and roles, notably the involvement of customers, suppliers, citizens or societal stakeholders.

When organizing RED workshops, it is critical to maintain the spirit of ethical deliberation, and organize it as part of a participative, iterative process that promotes curiosity and creativity (Steen, 2013b), in the spirit also of Responsible Innovation and its key dimensions of anticipation, reflexivity, inclusion and responsiveness (Stilgoe et al., 2013). There is, however, a risk of this method becoming a shallow exercise. This draws attention to the key role of workshop facilitators; they will need to create a space for openminded reflection, for critical and constructive questioning, for curiosity and creativity, for anticipation, reflexivity, inclusion and responsiveness—so that the workshop has sufficient depth and width (not a shallow exercise). Ideally, a workshop like this is organized as an opportunity to cultivate relevant virtues, not only classical, ‘cardinal’ virtues like justice, courage, self-control and practical wisdom, but also contemporary ‘technomoral’ virtues (Vallor, 2016), like *flexibility*, a ‘reliable and skilful disposition to modulate action, belief, and feeling as called for by novel, unpredictable, frustrating, or unstable technosocial conditions’ (p. 145) and *perspective*, ‘a reliable disposition to attend to, discern and understand moral phenomena as meaningful parts of a moral whole’ (p. 149). Indeed, Vallor reminds us that ‘the ultimate engineering task is the fragile, endless, and sublime human project of using the culture we produce to make ourselves into the beings we wish to become’ (*op. cit.*: 159).

REFERENCES

- Brey, P. (2000). Disclosive computer ethics. *SIGCAS Comput. Soc.*, 30(4), 10–16. doi:10.1145/572260.572264
- Brey, P. (2010). Values in technology and disclosive computer ethics. In L. Floridi (Ed.), *The Cambridge Handbook of Information and Computer Ethics* (pp. 41–58). Cambridge University Press. doi:10.1017/CBO9780511845239.004
- Carayannis, E., & Campbell, D. F. J. (2009). ‘Mode 3’ and ‘Quadruple Helix’: Toward a 21st century fractal innovation ecosystem. *International Journal of Technology Management*, 46(3-4), 201–234. doi:10.1504/IJTM.2009.023374
- Cross, N. (2011). *Design thinking*. Berg. doi:10.5040/9781474293884
- Doorn, N., Schuurbijs, D., Van de Poel, I., & Gorman, M. E. (2013). Early engagement and new technologies: Towards Comprehensive Technology Engagement? In N. Doorn, D. Schuurbijs, I. Van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 233–251). Dordrecht, The Netherlands: Springer Science+Business Media.
- Dorst, K. (2015). *Frame Innovation: Create new thinking by design*. MIT Press. doi:10.7551/mitpress/10096.001.0001
- Dorst, K., & Royakkers, L. (2006). The design analogy: A model for moral problem solving. *Design Studies*, 27(6), 633–656. doi:10.1016/j.destud.2006.05.002
- Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). *Management Research: An Introduction* (2nd ed.). Sage.
- Feng, J., Zhang, Y., Liu, X., Zhang, L., & Han, X. (2018). Just the Right Amount of Ethics Inspires Creativity: A Cross-Level Investigation of Ethical Leadership, Intrinsic Motivation, and Employee Creativity. *Journal of Business Ethics*, 153(3), 645–658. doi:10.1007/s10551-016-3297-1
- Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology*, 32(2), 185–193. doi:10.1007/s13347-019-00354-x
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. doi:10.1177/1077800405284363
- Friedman, B., Kahn, P., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction and management information systems* (pp. 348–372). M.E. Sharpe.
- Gassmann, O., & Schweitzer, F. (2014). *Management of the Fuzzy Front End of Innovation*. Springer. doi:10.1007/978-3-319-01056-4
- Granstrand, O., & Holgersson, M. (2020). Innovation ecosystems: A conceptual review and a new definition. *Technovation*, 90-91, 102098. doi:10.1016/j.technovation.2019.102098
- Guertler, M. R., Kriz, A., & Sick, N. (2020). Encouraging and enabling action research in innovation management. *R&D Management*. 10.1111/radm.12413
- Hickok, M. (2020). *Lessons learned from AI ethics principles for future actions*. AI and Ethics., doi:10.1007/s43681-020-00008-1
- Jacobs, N., & Hultgren, A. (2018). Why value sensitive design needs ethical commitments. *Ethics and Information Technology*. Advance online publication. doi:10.1007/s10676-018-9467-3
- Jansen, P., Sattarov, F., Douglas, D., Reijers, W., Gurzawska, A., Alexandra, K., & Warso, Z. (Eds.). (2017). *Outline of an Ethics Assessment Framework: Main results of the SATORI project*. Academic Press.
- McPhee, C., Hoppe, M., & Lindhult, E. (2019). Editorial: Action Research. *Technology Innovation Management Review*, 9(4), 3–6. doi:10.22215/timreview/1228
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*. Advance online publication. doi:10.1007/s11948-019-00165-5 PMID:31828533

- Ollila, S., & Yström, A. (2020). Action research for innovation management: three benefits, three challenges, and three spaces. *R&D Management*. 10.1111/radm.12407
- Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O’Sullivan, D., & Gordijn, B. (2018). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. *Science and Engineering Ethics*, 24(5), 1437–1481. doi:10.1007/s11948-017-9961-8 PMID:28900898
- Royakkers, L., & Steen, M. (2017). Developing Tools to Counteract and Prevent Suicide Bomber Incidents: A Case Study in Value Sensitive Design. *Science and Engineering Ethics*, 23(4), 1041–1058. doi:10.1007/s11948-016-9832-8
- Stahl, B. C. (2007). ETHICS, Morality and Critique: An Essay on Enid Mumford’s Socio-Technical Approach. *Journal of the Association for Information Systems*, 8(9), 479–490. doi:10.17705/1jais.00138
- Steen, M. (2013a). Co-design as a process of joint inquiry and imagination. *Design Issues*, 29(2), 16–29. doi:10.1162/DESI_a_00207
- Steen, M. (2013b). Virtues in participatory design: Cooperation, curiosity, creativity, empowerment and reflexivity. *Science and Engineering Ethics*, 19(3), 945–962. doi:10.1007/s11948-012-9380-9
- Steen, M. (2015). Upon opening the black box and finding it full: Exploring the ethics in design practices. *Science, Technology & Human Values*, 40(3), 389–420. doi:10.1177/0162243914547645
- Steen, M., & Nauta, J. (2020). Advantages and disadvantages of societal engagement: A case study in a research and technology organization. *Journal of Responsible Innovation*, ●●●, 1–22. doi:10.1080/23299460.2020.1813864
- Steen, M., & Van de Poel, I. (2012). Making values explicit during the design process. *IEEE Technology and Society Magazine*, 31(4), 63–72. doi:10.1109/MTS.2012.2225671
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. doi:10.1016/j.respol.2013.05.008
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press. doi:10.1093/acprof:oso/9780190498511.001.0001
- van de Poel, I. (2016). An Ethical Framework for Evaluating Experimental Technology. *Science and Engineering Ethics*, 22(3), 667–686. doi:10.1007/s11948-015-9724-3 PMID:26573302
- van de Poel, I. (2018). Design for value change. *Ethics and Information Technology*. Advance online publication. doi:10.1007/s10676-018-9461-9
- Van de Poel, I., & Royakkers, L. (2011). *Ethics, technology, and engineering: An introduction*. John Wiley and Sons.
- Van Gorp, A., & Van der Molen, S. (2011). Parallel, Embedded or Just Part of the Team: Ethicists Cooperating Within a European Security Research Project. *Science and Engineering Ethics*, 17(1), 31–43. doi:10.1007/s11948-009-9187-5 PMID:19936967
- Yin, R. (1994). *Case study research* (2nd ed.). Sage.

ENDNOTES

¹ <https://www.tno.nl/en/about-tno/>

² We are currently less interested in ethical issues that may occur within projects, e.g., regarding ways in which researcher treat subjects in trials (although these issues are important).

³ We are currently less interested in *ex ante* methods, which are typically used by *experts* in responsible innovation or foresight or by ethicists (Reijers et al., 2018: 1447-1450), rather than by researchers and designers themselves, i.e. the people we want to target and collaborate with (our colleagues), or in *ex post* methods, which are typically applied *after* projects have delivered results (*op cit.*: 1452-1454), i.e. the phases in which projects and their outcomes cannot be easily influenced anymore (as per the so-called Collingridge control dilemma: at a project’s start there are opportunities to steer it, but also a lack of

knowledge to base this steering on; whereas at a project's end more knowledge has been generated or become available, but opportunities to steer it have become minimal).

We are aware of the tension (or irony) that exists between calling our method 'Rapid' and claiming that it can facilitate a process of 'Ethical Deliberation'—which needs time and room for reflection and dialogue. We nevertheless chose to use the term 'Rapid' to make our method attractive to people who work in tech; they are likely to be familiar with rapid prototyping and iterative methods like Agile or Scrum. Moreover, we position RED as an accessible, first step in a promoting more ethical deliberation. Our intention is to enable people, over time, to engage in more extensive ethical deliberation. We do not claim that doing one RED is sufficient; indeed, that could easily derail into *ethics washing* (Florida, 2019).

We assume readers are familiar with these traditions and will not discuss too many details; this is consistent with the way in which these perspectives are presented in the method: without theoretical or conceptual background. Please note that we use the term 'relational ethics' to refer to what others call 'ethics of care' or 'feminist ethics'.

This method was developed and tried-out during the Covid-19 pandemic, hence the usage of online communication, rather than face to face meetings.

We have also worked with larger groups, where four groups of people, or two to four people each, work in parallel on the four ethical perspectives and share their insights.

Flyvbjerg (2006) refutes several common misunderstandings about case studies; he argues that case studies can provide 'concrete, practical (context-dependent) knowledge', which can be more valuable than 'general, theoretical (context-independent) knowledge, and that one can 'generalize on the basis of [even] an individual case'.

Sexual activities with children under 16 years are illegal in The Netherlands. In addition, 'grooming' is illegal; i.e. initiating contact via phone or internet with underage children and initiating to meet in person with intentions to commit fornication.

These words, *watchful* ('waakzaam') and *helpful* ('dienstbaar'), are displayed on Dutch police cars and buildings, and can be understood as key professional values, duties or virtues for police officers.

The use of 'early detection' services involves rights to privacy and resulting requirements, such as 'informed consent' before the police would participate in such groups.

Developed in *Pericles*, a Horizon 2020 project funded by the EU (<https://project-pericles.eu>; https://project-pericles.eu/wp-content/uploads/2020/03/3-Vulnerability-Assessment-Tool_landscape_compressed.pdf).

For privacy and data protection, see: https://www.riec.nl/binaries/riec/documenten/jaarverslagen/2019/06/23/riec-liec-jaarverslag-2019/Jaarverslag_RIECLIEC_2019_DEF.pdf, page 26.

Please note that we were interested in project team members' experiences during the innovation process. We were *not* interested in, e.g., the innovation project's *outcomes*—which were not yet available because the projects were still under development or underway. Moreover, it would have been difficult to study the effects of method X on outcomes Y; there are all sorts of variables that influence innovation processes 'in the field'.

Marc Steen works as senior research scientist at TNO, The Netherlands Organisation for Applied Scientific Research; his research interests include Human-Centred Design, Responsible Innovation, and Applied Ethics in the domain of Big Data, Algorithms and Artificial Intelligence.

Martijn Neef works as a senior innovator at the Defence, Safety and Security unit of TNO, The Netherlands Organisation for Applied Scientific Research.

Tamar Schaap works as a data scientist and researcher at TNO, The Netherlands Organisation for Applied Scientific Research.