



Circles of Care for Safety: A Care Ethics Approach to Safe-by-Design

Lieke Baas · Suzanne Metselaar ·
Pim Klaassen

Received: 30 July 2021 / Accepted: 27 July 2022 / Published online: 16 August 2022
© The Author(s) 2022

Abstract Safe-by-Design is an approach to engineering that aims to integrate the value of safety in the design and development of new technologies. It does so by integrating knowledge of potential dangers in the design process and developing methods to design undesirable effects out of the innovation. Recent discussions have highlighted several challenges in conceptualizing safety and integrating the value into the design process. Therefore, some have argued to design for the *responsibility* for safety, instead of for safety itself. However, this idea has not been developed further. In this article, we develop an approach to Safe-by-Design, grounded in care ethics, which builds on the idea of designing for responsibility and can deal with the complexity that is inherent to the conceptualization of safety. We describe five ways in which care ethics contributes to the conceptualization of Safe-by-Design: (1) It suggests the development of ‘circles of care’

in which stakeholders share the responsibility for safety; (2) it recognizes the importance of considering safety as something that is situated in the surroundings of a technology, instead of as a property of the technology itself; (3) it acknowledges that achieving safety is labour that requires an ongoing commitment; (4) it emphasizes that the way in which we relate to technology impacts its safety; and (5) it recognizes the role of emotions in assessing safety. All these elements combined lead to a broader understanding of safety and a philosophically more substantial and practically more appealing conceptualization of Safe-by-Design.

Keywords Safe-by-Design · Care · Responsible research and innovation · Safety engineering · Risk · Uncertainty

Introduction

Emerging technologies such as nanotechnology and synthetic biology may contribute substantially to innovations in food production, medicine, energy and manufacturing, and through this to human well-being and environmental sustainability. Simultaneously, however, they are also a source of risks and uncertainty. In an attempt to handle safety issues pro-actively, the concept of Safe-by-Design has been introduced [1–4]. Although the idea of designing

L. Baas (✉)
Medical Humanities, Julius Center for Health Sciences
and Primary Care, University Medical Center Utrecht,
Utrecht, Netherlands
e-mail: l.baas-3@umcutrecht.nl

S. Metselaar
Department of Ethics, Law & Humanities, Amsterdam
University Medical Center, Amsterdam, Netherlands

P. Klaassen
Athena Institute, Vrije Universiteit Amsterdam,
Amsterdam, Netherlands

products, processes, materials, or technologies¹ to be safe(r) per se is not new, the concept of Safe-by-Design as presently conceived is.

In a particularly insightful study on Safe-by-Design, Van de Poel and Robaey argue that, at least where emerging technologies that are characterized by high levels of uncertainty are concerned, it would be wise not only to design for safety, but also *for the responsibility for safety* [5]. They argue that it is ineffective to design technologies in ways that decrease (future) users' freedom to handle them in presently unforeseen manners. Certain risks may only become apparent after the design stage, and in such cases, users are in a unique position to handle the technology adaptively and flexibly, thereby potentially increasing safety levels. This is particularly relevant for technologies that may be used by many people, such as sunscreen with nanoparticles.

Van de Poel and Robaey argue that we should learn to handle these technologies safely, instead of taking their safety for granted, similar to how users have learned to dispose batteries safely. Because technologies have the ability to change society in both material and moral dimensions, they consider it undemocratically to deny users the possibility to shape technologies during the use phase. If Safe-by-Design also entails designing for the responsibility for safety and not exclusively for ultimate safety, significant weaknesses can be avoided.

Although the proposal to design for the responsibility for safety is very promising, it focuses primarily on rule-like prescriptions, rather than relating to what responsibility might mean on the level of the actual practice. As such, it leaves several questions unanswered, including ones concerning the temporal, material, and affective dimensions of what it means to take care of safety and to organize the responsibility for it.

¹ A terminological point: We use the term *technology* as shorthand for sociotechnical assemblages, which encompass interactions between a technological artefact and its surroundings. Insofar as the products, processes or materials that can theoretically be the upshot of research and innovation that manifests a Safe-by-Design approach, these are developed by way of applying scientific knowledge and hence also fit the standard definition of technology. For that reason, from here on, we only speak of technology — but we use it in the broadest sense possible.

In this article, we aim to expand on the work by Van de Poel and Robaey, and develop an approach to Safe-by-Design that fills this gap. Our approach to Safe-by-Design is grounded in care ethics; it centres around building 'circles of care' for safety, in which all actors involved in any stage of a technology's life cycle share the responsibility of caring for its safety.

The idea of 'circles of care' was introduced by care ethicist Joan Tronto as a way of no longer viewing care solely as a dyadic relation, but as an on-going process in which multiple actors care together [6]. In a more speculative fashion, the extension of such a notion of care to issues of technoscience also features in the work of María Puig de la Bellacasa [7, 8], who regards care as an ongoing commitment to the making of futures.

Because the ethics of care recognizes the importance of concrete relationships in evaluating our moral obligations, as well as our interdependence in living the 'good life', it is particularly suitable for ethically guiding innovation practices in which many different stakeholders are involved. Furthermore, the concept of *care* in care ethics is, more than core concepts of other ethical theories such as autonomy, justice, or equity, capable of dealing with uncertainties, as it is actively involved with the values and virtues that we consider important for the future [9]. Thus far, care ethics had not explicitly been adopted in discussions on Safe-by-Design, but we will show that Safe-by-Design can be conceived of as a practice of caring for new technologies and the world into which they are introduced, including future generations [10, 11].

To do so, we proceed as follows: in '[Safe-by-Design](#)', we provide further background information on the concept of Safe-by-Design, and '[Care Ethics in the Age of Technoscience](#)' elaborates on the significance of care ethics for understanding technoscience. In '[A Caring Perspective on Safe-by-Design](#)', subsequently, we bring Safe-by-Design and care ethics together and we develop our care ethical conceptualization of Safe-by-Design. In '[Conclusions and Discussion](#)', finally, we formulate our conclusions and discuss these in light of topical developments.

Safe-by-Design

Significant contributions to our understanding of Safe-by-Design have been made in the context of

several projects funded by the European Commission, including Nanoreg2 and Prosafe,² and through a Special Section of the journal *NanoEthics* (2017, 11:3) dedicated to Safe-by-Design and a Special Issue on it in the *International Journal of Environmental Research and Public Health* (2020, 18:4). The contributions to both make clear that Safe-by-Design shows quite some interpretative flexibility [3].

However, common to all understandings is the idea that Safe-by-Design promotes the value of safety as important to consider when designing new technologies, and that safety (or risk) cannot be regarded simply as *outcome* of innovation processes for two reasons: first, because risk levels cannot always be measured with sufficient accuracy [12] and second, because safety is essentially context-sensitive and socially constructed, as it always has to be determined for whom or for what something has to meet which safety standards, and in which context [3, 5]. This makes it difficult, if not impossible, to evaluate whether a technology is per se safe.

At the same time, views on Safe-by-Design diverge such that we for instance find it described as “(experimental) process design focusing on procedural and technical risk management” [13], as well as “a specific elaboration of Responsible Research and Innovation, with an explicit focus on safety in relation to other important values in engineering such as well-being, sustainability, equity, and affordability” [3, and see also 4, 14].

Furthermore, some consider Safe-by-Design as a topic at home in the field of nanotechnology [3], whereas others apply it to all engineering disciplines [4], and yet others deem it particularly apt for understanding those technologies characterized by high levels of uncertainty [13]. However, for the present purposes, a more-encompassing comparative analysis will not be necessary. For a better understanding of Safe-by-Design, we will focus on an assessment of ‘safety’.

In the philosophy of technology, safety is often defined as being ‘concerned with avoiding certain classes of events that it is morally right to avoid’ [15, p. 45]. In the context of engineering, this usually translates to absence of hazards and risks, which includes at the very least the prevention of human

injuries and death resulting from the intended use of the technology [16]. Thereby, the concept of safety is closely related to risk and uncertainty.

As a result of this connection with risk, it has been argued that the definition of safety is paradoxical, as it is ‘defined and measured more by its absence than by its presence’ [17, p.4]. Further, James Reason argues that assessing safety based on such a definition is unreliable, for instance because low rates of incidents or even the total lack of incidents need not reflect a situation of absolute safety, as it may also result from a lack of reporting.

Moreover, this definition of safety leaves us with the question of what it is that is ‘morally right to avoid’. Apart from preventing human death or injuries, literature on designing for safety also discusses animal health, ecosystem integrity, and sustainability, to name just a few values considered worthy of protection [18–20]. Safety is thus a concept that comprises multiple different values. It should therefore be regarded as a relational value, the precise extension and significance of which have to be decided in a context-sensitive way, in which it is considered for who or what something has to meet particular safety standards. In certain cases, this may for instance require weighing the safety of one species against that of another, or the safety of an individual animal against the group or species [21].

Furthermore, where different values may already play a role in the identification of risks, their effect becomes even more pronounced in the deciding on the acceptability of risks. In practice, different stakeholders often come to different conclusions, in which laypeople tend to value the context in which the risks exist more than professionals do [22].

A concept that is closely related to safety is security. Whereas safety aims to protect against unintended hazards resulting from the use of a technology, security aims to protect against intended hazards, such as a terrorist attack. Although the concepts are slightly different, safety measures may well benefit security and vice versa [16].

Further difficulties in designing for safety arise from the complexity that is inherent in technological development. New and emerging technologies often have long development trajectories involving numerous types of actors and are oriented towards the future [23]. Furthermore, emerging technologies often combine knowledge from various fields with engineering

² See www.nanoreg2.eu and www.h2020-prosafe.eu/.

principles [24]. This makes it increasingly difficult for researchers and engineers trained in a particular discipline to understand the technology as a whole [20, 23]. As a result, risks become harder to assess precisely and uncertainty becomes more pervasive.

Uncertainty can take several different shapes, which affect Safe-by-Design in different ways, as has been discussed by Van de Poel and Robaey [5]. They recognize five types of uncertainty: (1) risk, for situations in which the potential consequences and their probability are known; (2) scenario uncertainty, for situations in which not all failure mechanisms that might lead to undesirable outcomes are known; (3) ignorance, for situations in which both the potential failure mechanisms and undesirable consequences are unknown; (4) indeterminacy, for situations in which causal chains to the future are still open; and (5) normative ambiguity, for situations in which there is disagreement about values and norms. Strategies to design for safety thus need to recognize these types of uncertainty and find a way to deal with them.

Apart from these forms of epistemic uncertainty, designers have to deal with ontological uncertainties as well. With ontological uncertainties, we mean uncertainties arising from natural variability [23]. Ontological uncertainty may result from the unpredictability of human behaviour, similar to how Van de Poel and Robaey define the concept of indeterminacy, but in the case of living organisms, it may also result from evolutionary forces, which create individual differences and cause organisms to change unpredictably over generations [23, 25]. This type of uncertainty cannot be reduced by doing more research and obtaining more knowledge.

In current innovation practices, it is often the case that either technological risks and safety are understood in utilitarian-consequentialist terms, as are the cost–benefit analyses to which these risks are usually translated, or the precautionary principle is adopted, as is done in many innovation policies of the European Union [23, 26–29]. However, both these methods are unsatisfactory. The precautionary principle is criticized for being too restrictive, and cost–benefit analyses lose in effectiveness in situations with a high degree of uncertainty, as is often the case with complex technological development. Furthermore, not all elements relevant to the safety and acceptability of a technology can be quantified. For instance, citizens' subjective experience of safety

is not incorporated in such risk–benefit analyses, as the 'GMO backlash' has shown [19, 30]. Moreover, the introduction of new technologies into society can lead to value change or even the emergence of new values, which may impact over time how safety is conceptualized, weighed against other values, or understood to appropriately feature in cost–benefit analyses [31, 32].

Safe-by-Design thus needs to find different, additional ways of conceptualizing safety and handling risks. Current discussions surrounding the concept have already engaged with some of these challenges. For instance, many Safe-by-Design strategies include conducting more (fundamental) research, both inside and outside of the lab, in order to reduce uncertainties [19, 33]. A different approach has been taken by Van de Poel and Robaey [5], who argue that it may be better to design for the responsibility for safety rather than for safety itself, as elaborated on in the introduction. In what follows, we will elaborate on the idea of designing for the responsibility for safety, by developing a care ethics approach to Safe-by-Design, in which stakeholders share the responsibility in a circle of care.

Care Ethics in the Age of Technoscience

At the heart of care ethics is its critique of the idea that ethical problems can be solved by means of rules, regulations or abstract principles. Turning away from those, it focuses on concrete relationships and (the allocation of) responsibilities over these relationships. Care ethics departs from the idea that all human beings are interdependent; we need, receive, and give care to others. Instead of understanding care ethics as a normative theory, it can also be conceived as a moral perspective, which draws attention to other ethical aspects and questions [34].

Care ethics originally developed as a critical, feminist response to 'masculine' obligation-based approaches, which focus on universal principles as governing the good life. Care ethics, on the other hand, emphasizes relationships, embodied experience, and particularities as what is morally most significant.

Care ethics has subsequently evolved into many different forms, all of which share at least two assumptions [35, p. 134]:

1. The main characteristic of human existence is *relationality*.
2. Moral reasoning is characterized by moral *sensitivity, attentiveness, and connectedness*.

Thus, there are multiple forms of conceptualizing care and care ethics, each emphasizing different aspects; whereas some authors focus more on care as an act, others regard care as a virtue or a value. In this paper, we use Hamington's definition, which understands care as 'an approach to personal and social morality that shifts ethical considerations to context, relationships, and affective knowledge in a manner that can only be fully understood if care's embodied dimension is recognized. Care is committed to flourishing and growth of individuals, yet acknowledges our interconnectedness and interdependence' [35, p.3]. Simultaneously, we would like to approach care ethics as a kind of virtue ethics here: in order to give and receive care in the way it is described by Hamington, we need virtuous characters, and to see care as a primary virtue related to a host of other virtues [36].

Hence, our social, personal, and political life — including the technologies that are or will become part of it — require our caring responsibilities. This caring responsibility is seen as a virtue or a disposition that can be learned or fostered.

To further illustrate the meaning of care and care ethics, and in particular what the concept can contribute to responsible innovation, the next sections will elaborate on Tronto's explanation of five phases of care, with particular attention for her idea of democratic care; the way care has been conceptualized in the context of engineering and design; and how care contributes to thinking about techno-scientific futures.

Five Phases of Caring

Tronto defines five phases, or dimensions, of care: caring about, taking care of, care-giving, care receiving, and caring with [6, 37]. The first phase, *caring about*, describes the recognition of a need of another. Meeting this need requires care, specifically through the virtue of *attentiveness*. The next phase is *taking care of*, which goes one step further by assuming the *responsibility* for the meeting of this need, thereby also determining how the need

should be met. The third phase, *care-giving*, concerns the actual meeting of the need that requires care. This phase thus involves the actual, often physical, work of caring and thereby requires *competence*. The fourth phase in the caring process is *care-receiving*. This phase describes the response of the care-receiver towards the need being met, and requires the virtue of *responsiveness*.

The fifth phase of care, *caring with*, was added by Tronto more recently, in *Caring Democracy. Markets, Equality, and Justice* [6]. This work is a plea to rethink the fundamental values and commitments of democracy from a caring perspective: rather than prioritize production and economic life, the *allocation of caring responsibilities* should be our main priority. The work does not only argue that a caring democracy is a better democracy, but also that democratic care is better care.

Tronto reinforces the idea that care should not be confined to the realm of the domestic, but rather be the central concern of political life. Here, she focuses on the (power) mechanisms related to gender, race, class, and market that misallocate caring in our present democracies; care is not distributed equally among citizens. Whereas some citizens are much more on the receiving end of care, Tronto speaks of 'privileged irresponsibility', i.e. the privilege of not having to care for others as much due to one's social-economic status, others bear the burden of giving care without sufficiently being taken care of themselves. Democratic caring allows for a care process that is complex and multidimensional, in which caring responsibilities are shared among individuals and institutions, and allocated in a way that allows to achieve freedom, equality, and justice. In such a care process, it is essential that people, also as caregivers, recognize their own vulnerability and care needs. Tronto argues that making our democracy more inclusive in terms of freedom and equality requires making caring more just. Therefore, she expands her model of care by adding a fifth phase: *caring with*, which focuses on the distribution of care within society and requires both *solidarity* and *trust*.

Care in Engineering and Design

Presently, care ethics does not have a prominent place in engineering or technology ethics. Nonetheless,

several characteristics of engineering practice fit well with the ethics of care, and several authors have connected stages of the design process to Tronto's phases of care [38–40]. For instance, care plays an important role in the everyday work in laboratories. Kerr and Garforth identified care for nature, materials, colleagues, and careers in bioscience laboratories, which influences researchers' perspectives on what constitutes a successful scientific topic [41].

Furthermore, Hamington argues that the relational and responsive approaches in design thinking and in care ethics show similarities, and he has identified three elements associated with both: inquiry, empathy, and culture change [42]. Hamington emphasizes that design thinking is a form of *inquiry*, and recognizes a *relational epistemology* herein in two ways. The first concerns a 'holistic understanding of the target of concern' (p. 3). The process of knowledge gathering is described as an activity during which facts and data, as well as implicit knowledge are gathered. The other aspect of a relational epistemology is recognizable in the interactions within the design team. Design is often the result of teamwork, and the exchange of tacit knowledge between team members is essential for the development of an engineering solution as well as the professional development of individual engineers.

He further argues that *empathy* plays a crucial role in both care and design thinking. It is an essential component of providing care, as empathy leads to the motivational displacement that is required to recognize needs. For comparable reasons, empathy is also a central element of human-centred design. It serves as a means of obtaining a deep insight and understanding of the future users of the design, thereby allowing to obtain rich knowledge. In line with these elements, Hamington argues that both design thinking and care ethics foster a *culture change* towards a 'relational and interactive human-centred community', in which there is room for innovation and taking risks.

Thinking with Care

Along similar lines of reasoning, Puig de la Bellacasa connects care ethics to thinking about technoscientific futures, by exploring what care can mean for 'thinking and living with more than human worlds' [7, p.4]. Building upon the work of Bruno Latour and Donna Haraway, Puig de la Bellacasa views care as ongoing, thick

involvement. It aims to give insight into care as a speculative commitment to the 'ongoing material remaking of the world' [7, p. 28]. Building on Latour's concept of 'matters of concern', she also looks into the ways in which care can be a part of science and sociotechnical assemblages, by introducing the notion of 'matters of care'. The concept *matters of concern* is developed on the basis of the insight that sociotechnical assemblages always incorporate social and political interests. It brings the mode of fabrication to attention, as well as researchers' own concerns in framing problems. Because the affective connotations of care are stronger than those of concern, the concept *matters of care* implies a more active stance and more involvement than matters of concern, Puig de la Bellacasa argues.

Furthermore, she emphasizes the importance of *thinking with care*, which is described in terms of the embeddedness of thought in the world one cares for, which entails recognizing and acknowledging one's own involvements in perpetuating values. As knowledge is situated, thinking and knowing are inconceivable without the relations that form the world. As all relations require care, thinking and knowing require care too [43]. Knowing, so construed, does not revolve around 'prediction and control', but around being and remaining 'attentive to the unknown' [7, p. 100].

A Caring Perspective on Safe-by-Design

The first two sections have provided a description of current debates on Safe-by-Design and care ethics, particularly insofar as pertinent to considerations on the responsible design and governance of emerging technologies. In this section, we will show how care ethics can provide a valuable contribution to the conceptualization of Safe-by-Design, by elaborating on several elements related to both safety and the design process that care ethics draws attention to. Most importantly, we will direct attention to building circles of care, which allow to share the responsibility for safety. Furthermore, we will elaborate on how care ethics highlights the importance of looking at the situatedness of a technology, of approaching the care for safety as an ongoing commitment, of the way in which researchers relate to their objects of study, and of recognizing the reality and importance of emotions.

Building Circles of Care for Safety

Tronto argues that relations of care do not necessarily need to be one-on-one relationships, i.e. a dyadic relation between one caregiver and one receiver [6]. Instead, she argues that just as research improves by triangulation, care improves when it is done by multiple care-givers. In doing so, multiple individuals as well as institutions can together form a so-called ‘circle of care’. Within the circle, the caring responsibility and care work are shared among the various actors involved in the different stages of a product’s life cycle. Sharing these responsibilities also facilitates breaking down hierarchical relationships, which in turn allows to make the caring process more democratic, according to Tronto [6].

The caring process taking place in the circle of care should integrate all five dimensions of care that were previously described, which requires that the related moral elements attentiveness, responsibility, competence, responsiveness, justice, and trust are present in the circle. Furthermore, since what constitutes ‘good’ care is always dependent on the specific situation in its context, several specifics of the circle of care can only be defined in a given situation. This implies, for instance, that the question of how such a circle of care should look like cannot be answered once and for all, but has to be revisited in the context of any sociotechnical design process. Although care may improve when it is done by multiple people, these circles cannot be endlessly large. Therefore, the question remains who or what should be the care-receiver and who should be the care-givers, and how they should share their caring responsibilities.

The importance of incorporating various actors in a circle of care can be illustrated with an example from the medical field, with the development of innovative therapies such as gene therapy. Caring for the safety of patients receiving gene therapy requires the involvement of several actors. This includes the physicians, nurses, and potential other caregivers directly involved in administering the gene therapy product, as well as all the researchers involved in (pre-) clinical studies into gene therapy, and the committees and governing institutions approving such trials and treatments. However, in addition to these more usual suspects, there is a role for sponsors, patient organizations, and the media as well, in the creation and management of (unrealistically) large expectations. Gene

therapy is often promised to cure various genetic diseases, even though such promises have (thus far) rarely been achieved. Such promises and expectations are notorious for influencing patients’ perspective on and assessment of risks and benefits [44, 45].

An example of how a circle of care could look like can be found in an article by Spruit, who offers an analysis of how care can complement the precautionary principle for maximizing the safety of workers who work with nanomaterials [46]. In so doing, Spruit describes several relations of care, which can together be regarded as a circle of care surrounding the workers. Most prominently is the responsibility of the employers to care for their employees. These employers do not always have the relevant expertise, and have to rely on expert advice for making decisions regarding safety. Furthermore, in the context of the case analysed by Spruit, organizations such as the Dutch National Institute for Public Health and the Environment (RIVM) and (then) government-funded knowledge institute TNO set up guidelines and create tools to support working with nanomaterials safely.

A circle of care thereby also highlights that design for safety should incorporate more than designing the technological artefact at issue, and should extend to designing the use practices in which it becomes embedded as well. As was argued by Jasanoff, McGonigle, and Stevens, “what technologies eventually become, and how we shape ourselves to live with technology, depend not only on choices of material design but also on how we design the social, political, and educational systems in which they are embedded” [47, p.72]. This requires that a circle of care incorporates these systems when designing for safety.

Situated Safety

As said, context plays an important role in the ethics of care. It has often been argued that it is impossible to define what constitutes good care in general, as that is dependent on the specific situation and the relations involved [37]. Care ethics thereby emphasizes the importance of looking at the situatedness of a technology and how this may influence what issues of risk and safety surface and what types of uncertainty are at play.

Awareness of the importance of context can also be recognized in the literature on Safe-by-Design and safety engineering, albeit in a different way.

For instance, in literature focussing on biotechnology, attention to context often implies having attention for the interaction an organism may have with its environment, whether that is in the lab or in an ecosystem. It focuses on questions around, for instance, the availability of food, toxicity, and the possibility of unintended gene transfer taking place. To give an example, the use of an antibiotic resistance gene as a marker may be considered to be safe when the bacteria are contained in the laboratory and handled with the proper lab safety procedures, whereas releasing such modified microorganisms into the environment may have serious public health consequences.

The way in which care ethics emphasizes the situatedness of both the process of technology development and of the technology itself, goes further than looking at biological interactions between organisms or their physical context. This can, for instance, also include its political or relational context, and the power dynamics at play [6]. To stick with the case of biotechnology, that context comprises a plethora of actors and factors — with some variation from one geographical location to another. Wherever one looks, one will find clearly discernible differences between different (types of) actors on dimensions ranging from emotions and attitudes towards different applications of biotechnology (e.g., food, industrial manufacturing, medicine, and vaccine production) to power to influence sociotechnical developments. Including such a variety of actors underlines again the importance of adopting an inclusive and democratic approach to engineering, which is inherent to the caring approach in a circle of care for safety.

Adopting a more inclusive approach to technological risk assessment also allows to incorporate a wider variation in types of knowledge, such as users' experiential knowledge in dealing with a technology. As was discussed before, laypeople and experts often differ in their views on risks, and properly designing for safety thus also requires the design of deliberation structures in which decisions can be made with different stakeholders.

Safety as an Ongoing Commitment

An important aspect of care is that it is not a one-time activity, but an ongoing commitment to meeting the needs of another — or several human or

non-human others. Applying this to Safe-by-Design means that care for the safety of a technology is not finished once it is introduced into society, but continues for as long as the technology is there. This goes further than anticipating on every step in a technology's life cycle, which many already consider to be an essential element of Safe-by-Design [2, 19]. The concept of a circle of care helps in establishing this continued care for a technology and its surroundings, as different actors may play a role in providing or receiving care in different phases.

Seeing safety as an ongoing commitment, which may require new actions at a later, initially unforeseen point in time may help in overcoming two important challenges of Safe-by-Design. First, some of the uncertainties related to emerging technologies are irreducible, as these technologies open up unknown territory [23]. Many risks may only become apparent (long) after the technology is introduced. Even when aiming for thorough and inclusive anticipation, it is naive to think they can fully be anticipated. The circle of care helps in identifying these risks when they emerge, and allows to act on them if needed.

Second, the entrenchment of new (radical) technological innovations does not seldomly cause values to change, or new values to arise [32]. As a result, the interpretation of safety, or its relative importance in relation to other values may be different than it was at the time of the design of the technology [32]. Such changes may occur independently of the technology in society, and they can also result from the introduction of the technology itself. As a result, a technology that may have been considered safe at the time it was introduced may be considered insufficiently safe at a certain point in the future. Continuing to care for a technology, even long after it has been introduced on the market, increases the chances of picking up on this problem, and being able to re-design the technology, its contexts of usage, or any other feature of the sociotechnical assemblage it is part of to be considered safe again.

Relating with Care: Minding Your Language

Engineering is a form of knowledge creation, and many forms of engineering, especially in fields such as nanotechnology and synthetic biology, are working at the vanguard of currently available knowledge

[42]. Puig de la Bellacasa aimed to develop a notion of care that is valuable for thinking and knowing. She argues that thinking and knowing are ways of relating to the world, and as all relations require care, they do too. She argues that the way in which research is conducted and how its research objects are represented has world-making effects. Therefore, she emphasizes the importance of being aware of and caring for these effects of our thinking, by being critical about how research relates to its objects of study and to the world, thereby also recognizing and acknowledging one's own involvement in perpetuating values.

For Safe-by-Design, this means that it is important to take into account how fields of research and engineering represent their engineering questions and products, what models they use and particularly how these impact safety. One example of what should be taken into account is the role of language and metaphors in models and communication about research, as these shape expectations.

An example of how language impacts expectations of a new technology in a way that is relevant for their safety comes from innovation in the medical field again. Many researchers tend to speak of gene “therapy” in situations in which therapeutic value is absent, such as in early phase clinical trials in which evidence of efficacy is lacking, or applications aimed at genetic enhancement for cosmetic ends rather than treatment of a disease. It has been argued that using gene “therapy” to describe all these techniques does not only impair public communication and misrepresent the ethical aspects of such technologies, but also impacts their risk assessment [54].

Another example of the impact of language comes from the field of synthetic biology. In this field, three main metaphors can be distinguished, describing organisms as books, as engines or machines, and as computers [48]. Moreover, several authors have emphasized that the entire concept of engineering biology is a metaphor itself, on which the entire field is built [49, 50].

These metaphors can influence safety in at least three ways. First, they influence how the (engineered) organism is understood, which impacts the extent to which its behaviour is understood more or less correctly. For instance, basic cell physiology is often described as the cell's ‘chassis’. This description

suggests that biological parts have known, stable, and (hence) predictable properties, making the behaviour of the organism as a whole predictable as well. However, in reality, these parts interact with their environment, their behaviour is context-dependent, and they are subject to evolutionary forces [51]. The difficulties arising from such metaphors can be summarized as the ‘problem of environment’ and the ‘problem of development’ [52].

Second, the safety strategies developed are also based on these metaphors, which means that different metaphors may lead to different safety strategies. For instance, the field regards evolution and interaction with the environment as threats to reliable design, and reliability and predictability are regarded as essential components of designing for safety. As a result, many safety strategies focus on eliminating these elements, for instance through the development of a ‘minimal genome’, which contains only the genes absolutely necessary for survival, thereby limiting the potential of interaction with the environment [19, 53]. Boudry and Pigliucci have argued that engineers may be more successful in achieving their aims when they choose to embrace adaptation and evolution, instead of aiming to design it out [49]. Introducing yet more metaphors, they suggest that design is not that different from evolution, as both can be described as a process of *tinkering* in order to find something that *works* under particular circumstances.

Third, regardless of whether a metaphor accurately captures characteristics of an organism, assumptions resulting from their use have implications for the experience of safety. Boldt points out that machine metaphors do not only hide aspects, such as evolutionary change and interdependence in ecosystems, but also reinforce a belief in the potential of synthetic biology in solving societal problems [54]. And solving problems, of course, is something diametrically opposed to *causing* or *introducing* problems, such as novel safety risks.

Metaphors may be very helpful in mapping unknown territories, but it is important to remain critical of them, and they may also be misinterpreted [51]. In particular with an increasingly large circle of care surrounding technologies, there may be differences between the various stakeholders in how the metaphors are understood.

Recognizing Moral Emotions and Affect

The ethics of care has a central place for emotions. However, emotions are usually absent in the conventional methods of risk assessment [55]. The inadequacy of this was clearly shown by the so-called GMO backlash. Even though GMO foods passed all safety tests and were regarded to be safe by experts, their introduction on the market faced — and still faces — large public resistance, an emotional reaction that has often been described as a ‘yuck’ feeling [30]. Preston and Wickson have summarized that this reaction represents concerns around, amongst others, freedom of choice for non-GM food producers and consumers, or the domineering relationship between humans and nature [30].

Nussbaum has argued how such emotional reactions are in fact judgements of value, and thereby provide insight into what it is that we regard as valuable [56]. In doing so, emotions might also give unique insights into the values that underlie risk perceptions. In the example above, for instance, one of the values that the ‘yuck’ feeling alerts us to is freedom.

Roeser even argues that emotions are essential for making rational decisions about the moral acceptability of technological risks [51]. She makes use of a cognitive theory of emotions, which says that emotions are required to have moral knowledge, instead of viewing emotions and cognitions as being mutually exclusive. As an example Roeser mentions how the moral emotion of sympathy may alert us to issues of equity. In appraising the acceptability of risks, Roeser goes on to argue, considerations such as voluntariness or the availability of alternatives play an important role. Considering voluntariness, she states that people who are forced to do something against their will often experience feelings of anger or frustration. Further, she argues that someone who is not upset after an injustice has been committed to them would probably be considered irrational. Considering the availability of alternatives, she compares the cases of driving a car and nuclear energy. In situations without the availability of public transport, a car might be a necessary means of transportation. In contrast, she argues, there are alternatives to nuclear energy that are not fully exploited yet. It is therefore not irrational that people experience more fear of the risks of nuclear energy than of driving a car [55]. Thus, emotions cannot be left out of full assessments of the acceptability of risks.

Next to the emotions of the users or public, the emotions of engineers and designers may also have a significant impact on the safety of a technology. For instance, their enthusiasm about their project may cause them to miss certain risks. Particularly in combination with the use of metaphors, as discussed in the previous section, researchers’ emotions may lead them to overestimate technologies’ benefits and underestimate their risks. This is comparable to the phenomenon of therapeutic optimism in clinical research, in which hope for a certain health outcome impacts healthcare professionals’ framing of negative outcomes and balancing of uncertainties [57].

Further, it has been described that emotions can impact both preclinical and clinical research into medical innovations such as gene transfer. If researchers are emotionally invested in their preclinical studies aiming to inform, for instance, the development of gene transfer treatment for cancer, they can subconsciously score responses from treatment and control animals differently [55]. In addition, Kimmelman argues that it is plausible that the US Recombinant DNA Advisory Committee (RAC) has let optimism about the potential future success of gene transfer influence their decision on the weighting of harms and benefits of the first human trial of gene transfer for haemophilia [22].

Conclusions and Discussion

In this article, we aimed to put forward a care ethics approach to Safe-by-Design, which builds on the idea of designing for responsibility and is capable of dealing with the complexities that are inherent to the conceptualization of safety. Although there is no one single definition or framework for Safe-by-Design, the several different interpretations all describe Safe-by-Design as an approach to engineering and risk management that aims to design technologies to be as safe as possible, which is done by anticipating and incorporating responses to potential safety threats that may occur throughout the entire life cycle of a product or technological application, while it is still in its design phase.

However, as has been argued by Van de Poel and Robaey, there are several difficulties with the current conceptualization of Safe-by-Design [5]. The existing approach wrongly considers safety as an outcome or a

property of the technology, and puts the responsibility for safety entirely on the shoulders of designers and engineers, which is both undemocratic and ineffective. Moreover, considering the high level of uncertainty surrounding emerging technologies, it is very difficult to anticipate future safety issues that may come along with these technologies themselves.

Approaching the concept of Safe-by-Design from a care ethics perspective draws attention to the fact that achieving safety is a process that requires work and commitment, and thereby focuses on the responsible subjects (i.e., the stakeholders involved), instead of on the safety of the object (i.e., the technology itself). As it further emphasizes that the responsibility for safety has to be shared by all stakeholders involved with the technology, from its very first conceptualization to its disposal as waste, it highlights that safety is not only the burden of engineers and designers, but of all those involved, including institutions and individual users. This approach builds on the work by Van de Poel and Robaey, by thinking through what designing for responsibility could look like, an idea they launched but did not elaborate on in much detail.

Additionally, this approach draws attention to the importance of addressing the situatedness of a technology in conceptualizing safety and efforts to achieve it, thereby recognizing a broader range of factors that impact safety. This situatedness also concerns the way in which researchers and engineers relate to their objects of study, and the language, metaphors, and models they use to describe them. Last, a care perspective emphasizes the importance of incorporating emotions in the assessment of technology's safety, as they may point to values that underlie conceptions of risk, and are thereby essential for a full understanding of safety.

By combining all these elements, a care ethics approach broadens the understanding of safety. It does so by both allowing to incorporate more values, for instance fairness or autonomy, and by widening the scope of what is considered to be the object of safety, including not only the sociotechnical assemblage, but also the context into which it is introduced and the language and models surrounding it. Hereby, the conceptualization of safety comes closer to ways in which it is often understood by the general public, and more removed from the narrower understanding as a toxicological value that is often employed by professionals [22].

The five components care ethics contributes to Safe-by-Design as presented in this paper add more dimensions to Van de Poel and Robaey's proposal to design for responsibility. In their article, Van de Poel and Robaey present several heuristics for designing for responsibility. Several of those already focus on fostering certain behaviour, capacities, and virtues in stakeholders. Our care ethics approach gives more depth to these heuristics, especially by focusing on caring virtues. Moreover, it adds to the heuristics of Van de Poel and Robaey as they focus on what design should *achieve*, whereas our care ethics approach pays attention to how the design process is *brought about*, for instance by drawing attention to the language and metaphors used in underlying theories and to the impact of researchers' and engineers' emotions on their designs.

Furthermore, the design for responsibility approach focuses specifically on indeterminacy, whereas our care ethics approach is valid for other types of uncertainty as well, by considering in a speculative fashion what different technological futures could look like and what safety would mean in these futures.

Moreover, by recruiting the conceptual armoury of care ethics in further developing the notion of Safe-by-Design, we not only provide more moral substance to the latter, but also contribute to wider efforts of bringing Safe-by-Design in contact with the broader notion of Responsible Research and Innovation (RRI) [4], which also builds on key insights from care ethics [58].

As a result of the context dependence of this approach, there will be no one-size-fits-all method for implementing it in practice. However, there are some general characteristics of this approach that would need to be taken into account in order to make it feasible in practice. For one thing, there needs to be attention for the technology's surroundings and how people will behave with the technology. As it is unlikely that all future practices in which the technology will be embedded can be anticipated, this may require cultivating the virtues required for caring, such as those Tronto has put forward: attentiveness, responsibility, competence, responsiveness, justice, and trust. Future engineers' training and education may focus on cultivating these virtues, for instance through programmes intended to build character, such as already exist for promoting research integrity [59]. Building this caring attitude in the stakeholders

involved is an important requirement for establishing successful circles of care.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hale A, Kirwan B, Kjellén U (2007) Safe by design: Where are we now?. *Saf Sci* 45(1–2):305–327. <https://doi.org/10.1016/j.ssci.2006.08.007>
- Noorlander C, Sips A, Höck J, Höhener K, Lehmann HC (2016) NANoREG safe-by-design (SbD) concept. https://rivm.nl/en/About_RIVM/Mission_and_strategy/International_Affairs/International_Projects/Completed/NANoREG/deliverables/20160602_NANoREG_SbD_concept_final.org
- Schwarz-Plaschg C, Kallhoff A, Eisenberger I (2017) Making nanomaterials safer by design? *NanoEthics* 11(3):277–281
- van Gelder P, Klaassen P, Taebi B, Walhout B, van Ommen R, van de Poel I et al (2021) Safe-by-Design in engineering: An overview and comparative analysis of engineering disciplines. *Int J Environ Res Public Health* 18:6329
- van de Poel I, Robaey Z (2017) Safe-by-Design: From safety to responsibility. *NanoEthics* 11(3):297–306. <https://doi.org/10.1007/s11569-017-0301-x>
- Tronto JC (2013) *Caring democracy: Markets, equality and justice*. NYU Press, New York
- Puig de La Bellacasa M (2017) *Matters of care: Speculative ethics in more than human worlds*. University of Minnesota Press, Minneapolis
- Puig de La Bellacasa M (2015) Making time for soil: Technoscientific futurity and the pace of care. *Soc Stud Sci* 45(5):691–716. <https://doi.org/10.1177/0306312715599851>
- Groves C (2009) Future ethics: Risk, care and non-reciprocal responsibility. *J Glob Ethics* 5(1):17–31. <https://doi.org/10.1080/17449620902765286>
- Adam B, Groves C (2011) Futures tended: Care and future-oriented responsibility. *Bull Sci Technol Soc* 31(1):17–27
- Stilgoe J, Owen R, Macnaghten P (2013) Developing a framework for responsible innovation. *Res Pol* 42(9):1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- Reimhult E (2017) Nanoparticle risks and identification in a world where small things do not survive. *NanoEthics* 11(3):283–290
- Bouchaut B, Asveld L, Hanefeld U, Vlierboom A (2021) Value conflicts in designing for safety: Distinguishing applications of Safe-by-Design and the inherent safety principles. *Int J Environ Res Public Health* 18:1963. <https://doi.org/10.3390/ijerph18041963>
- Kallhoff A (2017) Safer by design and trump rights of citizens. *NanoEthics* 11(3):291–295
- Hansson SO (2006) Safe design. *Techné* 10(1):45–52
- Hansson SO (2009) Risk and safety in technology. In: *Philosophy of technology and engineering sciences*. North-Holland, pp 1069–102. <https://www.sciencedirect.com/science/article/pii/B9780444516671500434>
- Reason J (2000) Safety paradoxes and safety culture. *Inj Contr Safety Prom* 7(1):3–14. [https://doi.org/10.1076/1566-0974\(200003\)7:1;1-V;FT003](https://doi.org/10.1076/1566-0974(200003)7:1;1-V;FT003)
- Guan Z, Schmidt M, Pei L, Wei W, Ma K. (2013) Biosafety considerations of synthetic biology in the International Genetically Engineered Machine (iGEM) competition. *BioScience* 63(1):25–34. <https://doi.org/10.1525/bio.2013.63.1.7>
- Robaey Z (2018) Dealing with risks of biotechnology: Understanding the potential of Safe-by-Design. <https://doi.org/10.13140/RG.2.2.13725.97769>
- Schmidt M (2009) Do I understand what I can create? Biosafety issues in synthetic biology. In: Schmidt M, Ganguli-Mitra A, Kelle A, de Vriend H (eds) *Synthetic biology: The technoscience and its societal consequences*. Dordrecht, Springer, pp. 81–100
- Hjorth R, van Hove L, Wickson F (2017) What can nanosafety learn from drug development? The feasibility of “safety by design”. *Nanotoxicology* 11:305–12. <https://doi.org/10.1080/17435390.2017.1299891>
- Kimmelman J (2004) Valuing risk: The ethical review of clinical trial safety. *Kennedy Inst Ethics J*. Johns Hopkins University Press 14:369–93
- Sollie P (2009) On uncertainty in ethics and technology. In: Sollie P, Düwell M (eds) *Evaluating new technologies: Methodological problems for the ethical assessment of technology developments*. Dordrecht, Springer, pp 141–158
- Bayer K (2014) Brief note on the development of biotechnology. *Food Technol Biotechnol* 52(1, SI):13–15
- Cardinale S, Arkin AP (2012) Contextualizing context for synthetic biology — Identifying causes of failure of synthetic biological systems. *Biotechnol J* 7:856–866
- Groves C (2015) Logic of choice or logic of care? Uncertainty, technological mediation and responsible innovation. *NanoEthics* 9(3):321–333
- Hansson SO (2007) Philosophical problems in cost-benefit analysis. *Econ Philos* 23:163–183
- Hansson SO (2020) How extreme is the precautionary principle? *NanoEthics* 14(3):245–257. <https://doi.org/10.1007/s11569-020-00373-5>
- Swierstra T, Rip A (2007) Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. *NanoEthics* 1:3–20

30. Preston CJ, Wickson F (2016) Broadening the lens for the governance of emerging technologies: Care ethics and agricultural biotechnology. *Technol Soc* 45:48–57. <https://doi.org/10.1016/j.techsoc.2016.03.001>
31. Swierstra T (2015) Identifying the normative challenges posed by technology's 'soft' impacts. *Etikk i praksis Nord J Appl Ethics* 9(1):5–20
32. van de Poel I (2018) Design for value change. *Ethics Inf Technol* 1–5. <https://doi.org/10.1007/s10676-018-9461-9>
33. Schmidt M (2008) Diffusion of synthetic biology: A challenge to biosafety. *Syst Synth Biol* 2, 1–6. <https://doi.org/10.1007/s11693-008-9018-z>
34. Verkerk M (2005) A feminist care-ethics approach to genetics. In: Ashcroft R, Lucassen A, Parker M, Verkerk M, Widdershoven G (eds) *Case analysis in clinical ethics*. Cambridge University Press, pp 133–148
35. Hamington M (2004) Embodied care: Jane Addams, Maurice Merleau-Ponty, and feminist ethics. University of Illinois Press. Available from: https://books.google.nl/books?hl=nl&lr=&id=t4oJHxk9AUC&oi=fnd&pg=PA5&dq=Embodied+Care:+Jane+Addams,+Maurice+Merleau-Ponty+and+Feminist+Ethics.+Chicago,+IL:+University+of+Illinois+Press,+2004.&ots=dg6ZXIHq3g&sig=ddP0fOpljDLI9-d4Ou6cSNSZlYc&redir_esc=y#v=onepage&q&f=false
36. Halwani R (2003) Care ethics and virtue ethics. *Hypatia* 20(18):161–192
37. Tronto JC (1993) Care. In: *Moral boundaries: A political argument for an ethic of care*. New York, Routledge, pp 101–110
38. Kardon JB (2006) Concept of “care” in engineering. *J Perform Constr Facil* 19(3):256–260
39. Pantazidou M, Nair I (1999) Ethic of care: Guiding principles for engineering teaching & practice. *J Eng Educ* 88(2):205–212
40. Nair I, Bulleit WM (2020) Pragmatism and care in engineering ethics. *Sci Eng Ethics* 26(1):65–87. <https://doi.org/10.1007/s11948-018-0080-y>
41. Kerr A, Garforth L (2015) Affective practices, care and bioscience: A study of two laboratories. *Sociolog Rev* 64(1):3–20. <https://doi.org/10.1111/1467-954X.12310>
42. Hamington M (2017) Integrating care ethics and design thinking. *J Bus Ethics* 155:91–103
43. De La Bellacasa MP (2012) “Nothing comes without its world”: Thinking with care. *Sociolog Rev* 60(2):197–216. <https://doi.org/10.1111/j.1467-954X.2012.02070.x>
44. Kimmelman J (2010) *Gene transfer and the ethics of first-in-human research: Lost in translation*. Cambridge University Press. https://books.google.nl/books?hl=nl&lr=&id=NT56AB80JkIC&oi=fnd&pg=PP9&dq=gene+transfer+and+ethics+of+first+in+human+trials&ots=KdRVFP7crU&sig=g9lnuLacWfniGHE4-6PEcr5IvYA&redir_esc=y#v=onepage&q=gene%20transfer%20and%20ethics%20of%20first%20in%20human%20trials&f=false
45. Rosenberg LE, Schechter AN (2000) Gene therapist, heal thyself. *Science* 287(5459):1751. <https://doi.org/10.1126/science.287.5459.1751>
46. Spruit SL (2015) Choosing between precautions for nanoparticles in the workplace: Complementing the precautionary principle with caring. *J Risk* 20(3):326–346
47. Jasanoff S, McGonigle I, Stevens H (2021) Science and technology for humanity: An STS view from Singapore. *East Asian Science, Technology and Society: An International Journal* 15(1):68–78
48. Hellsten I, Nerlich B (2011) Synthetic biology: Building the language for a new science brick by metaphorical brick. *New Genetics and Society* 30(4):375–397
49. Boudry M, Pigliucci M (2013) The mismeasure of machine: Synthetic biology and the trouble with engineering metaphors. *Stud Hist Philos Biol Biomed Sci* 1–9. <https://doi.org/10.1016/j.shpsc.2013.05.013>
50. Kearnes M, Kuch D, Johnston A (2018) How to do things with metaphors: Engineering life as hodgepodge. *Life Sci Soc Policy* 14(22)
51. de Lorenzo V (2011) Beware of metaphors: Chasses and orthogonality in synthetic biology. *Bioengineered Bugs* 2(1):3–7
52. Pigliucci M, Boudry M (2010) Why machine-information metaphors are bad for science and science education. *Sci Educ*. <https://doi.org/10.1007/s11191-010-9267-6>
53. de Lorenzo V (2018) Evolutionary tinkering vs. rational engineering in the times of synthetic biology. *Life Sci Soc Pol* 14(18)
54. Boldt, J. (2018) Machine metaphors and ethics in synthetic biology. *Life Sci Soc Policy* 14, 12. <https://doi.org/10.1186/s40504-018-0077-y>
55. Roeser S (2006) The role of emotions in judging the moral acceptability of risks. *Saf Sci* 44:689–700
56. Nussbaum MC (2001) *Upheavals of thought: The intelligence of emotions*. Cambridge University Press
57. Hallowell N, Snowdon C, Morrow S, Norman JE, Denison FC, Lawton J (2016) The role of therapeutic optimism in recruitment to a clinical trial in a peripartum setting: Balancing hope and uncertainty. *Trials* 17(1):1–13. <https://doi.org/10.1186/s13063-016-1394-1>
58. Owen R, Macnaghten P, Stilgoe J (2012) Responsible research and innovation: From science in society to science for society, with society. *Sci Public Policy* 39(6):751–760
59. Banks S (2018) Cultivating researcher integrity: Virtue-based approaches to research ethics. In: *Virtue ethics in the conduct and governance of social science research*. Emerald Publishing Limited, Bingley, pp. 21–44

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.