Jun 23rd, 9:00 AM - Jun 28th, 5:00 PM

Thought Experiments In The Ethics Of Designing For Future People.

Peter Buwert
Edinburgh Napier University, United Kingdom

Matt Sinclair
Edinburgh Napier University, United Kingdom

Citation

Buwert, P., and Sinclair, M. (2024) Thought Experiments In The Ethics Of Designing For Future People., in Gray, C., Ciliotta Chehade, E., Hekkert, P., Forlano, L., Ciuccarelli, P., Lloyd, P. (eds.), *DRS2024: Boston*, 23–28 June, Boston, USA. https://doi.org/10.21606/drs.2024.518

DRS BOSTON 20 DRS 24 JUNE 24 — 28

The Non-Identity Problem: Thought experiments in the ethics of designing for future people

Peter Buwert*, Matt Sinclair

Edinburgh Napier University

*Corresponding e-mail:p.buwert@napier.ac.uk

doi.org/10.21606/drs.2024.518

Abstract: The Non-Identity Problem (NIP) is a philosophical puzzle which challenges our intuitive assumptions and reasoning around the question of our moral obligations towards 'future people'. This paper explores the significance of the NIP for design, an activity which is necessarily both future-oriented and ethical in nature. Through examination of two thought experiments proposed by philosopher Derek Parfit, this paper makes two contributions to the field of design ethics. Firstly, it raises the profile of the NIP as a topic of interest and for further study in design ethics research. The second is to propose that philosophical thought experiments can play a practical role in equipping designers for real-world challenges. When employed as thinking devices to disrupt our existing ways of thinking, thought experiments open up spaces of creative disequilibrium in which to nurture, exercise, and strengthen mental capacities for approaching the ethically complex challenges of future-oriented design practice.

Keywords: Non-Identity Problem; Thought Experiments; Design Ethics; Design Futures

1. The ethical challenge of future oriented design

Design, by its very nature, is necessarily a future oriented activity. The act of designing is a present-moment imagining of possibilities for future states of being which are other than those which currently exist. Design is, in a fundamental sense, an act of the production of potentialities – the imagining of possibilities for what is not yet, but could be. As such, there is a future-oriented ethical dimension deeply and symbiotically embedded within designing (Dilnot,2011; Fry,2009; Yelavich & Adams,2014). All design activity (even speculative and abortive design conceptualisation which is never implemented) invokes questions of responsibility and obligation towards inhabitants of the future. Whenever the design act brings new potentialities into being, choices must be made: which will be actualised, and which will remain as un-realised possibilities? As the designer makes choices between competing potentialities, questions of ethics arise inherently through the consequences and responsibilities

incurred by setting a process in motion to realise one possible future, as opposed to another (Buwert,2016,2017).

Design choices shape the future, and have very real repercussions for those who will inhabit these altered future states. We could rework Anne-Marie Willis' neat articulation of ontological design: "we design our world, while our world acts back on us and designs us." (2006, p.70), to emphasise this future orientation: 'when we design today, we design our future worlds, and these future worlds act back on their inhabitants and design them'. This basic principle of ontological design futuring is at play in all design activity. Taking the implications of this principle seriously, this paper is motivated by examination of the underlying ethical questions: How ought we to design for the future? What obligations and responsibilities do designers have to care for those who will inhabit the futures which result from our decisions in the present? And how can we equip designers to productively engage with the first two questions within their everyday design practice?

'The Non-Identity Problem' (NIP) is a philosophical puzzle posed by British philosopher Derek Parfit (1942-2017), challenging our intuitive assumptions and reasoning around the question of our moral obligations towards 'future people'. Parfit's work is characterised by his regular use of vivid thought experiments which provide accessible entrance points to what can often be complex and counter-intuitive arguments, allowing the reader to quickly grasp the key principles at stake and test out their thinking. Through an examination of two thought experiments presented by Parfit in his discussion of the NIP, this paper makes two main contributions towards the exploration of the questions raised above. The first is simply to raise the profile of the NIP as a topic of interest and for further study in future-oriented design ethics research. The second is the more general proposal that philosophical thought experiments like the NIP can play a practical role in equipping designers for real-world challenges, when employed as thinking devices to nurture, exercise, and strengthen mental capacities for approaching the ethically complex challenges of future-oriented design practice.

2. Thought experiments

Thought experiments are imaginary scenarios carefully crafted – through the constraining of narrative variables – in order to highlight a certain core principle or challenge (Miscevic,2022; Sorensen,1999). The thinker is invited to play through a given scenario in their mind and consider the consequences of the unfolding action and any choices they must make within the constraints set out, before making conclusions. Thought experiments are commonly used by philosophers as a device for presenting, testing and critiquing complex nuanced philosophical proposals and arguments. But we also find them throughout daily life, wherever we mentally explore and test possibilities and counterfactuals. In their simplest forms we employ thought experiments whenever we ask a "what if…?" or "would you rather…?" question asking how things could potentially be different under alternative conditions, or if different choices were made.

Good thought experiments are remarkably engaging and accessible, requiring no prior knowledge of theoretical concepts or terminology, only an active imagination and a critical mind. Perhaps the best-known ethical thought experiment is the Trolley Problem (Foot,1967), which asks us to imagine making a choice: take action to save the lives of five people in the path of a runaway tram at the cost of the death of one person, or do nothing, resulting in the deaths of the five and the survival of the one. Successive variations of this formula probe our moral intuitions and principles, as we are asked to consider alternative scenarios: Would we push one person in front of the tram to save five lives? Would it change our choice if we knew the identities of those at risk? If you would sacrifice one person on the tram-track to save five, would you harvest the organs of a healthy bystander to save the lives of five gravely ill hospital patients? (Edmonds,2014; Thomson,1976)

Sometimes a thought experiment may point to an acceptable and relatively conclusive 'answer', but more often than not, the power of the thought experiment is to reveal more detail or better understanding of the range of competing and as yet unresolved approaches to addressing or reframing the question. The Trolley Problem example does not reveal universal moral principles which can be relied upon to know which action is 'right' when faced with an unfolding ethical crisis. Rather, its enduring power and appeal is found in the way it lays bare: the insufficiencies of established normative ethical theories in offering us satisfactory guidance beyond what our gut intuitions tell us is right; the wildly inconsistent nature of our intuitive moral reasoning; and the key significance of contextual factors in influencing our reasoning as to the right thing to do (Königs,2023; Plunkett & Greene,2019; Walsh,2011).

Thought experiments are fictions not intended to resolve or even reflect the complex nuanced realities of the real world. Some would criticise them in this regard as being reductionist word-games, suggesting we cannot extrapolate meaningful learning for the real-world from such unrealistically simple linguistic abstractions (Dennett,2013; Fodor,1964; Harman,1986; Scruton,2014). But to criticise in this way is to miss the point: they are purposefully reductionist precisely in order to help us to isolate relevant factors which play some specific role in our reasoning processes, and to consider each of these in turn. A carefully crafted thought experiment can defamiliarise the taken-for-granted, open space to question assumptions, and provide opportunities for reflection and recalibration of our underlying ways of thinking. This is why thought experiments are particularly useful and popular within the philosophy of ethics, and why they have specific potential to be useful and relevant to design practice as we wrestle with the wicked (Chan,2023; Coyne,2005) challenges of designing for the future.

Thought experiments can be particularly useful when established paradigms are challenged by new thinking. Despite its near-ubiquity in commercial design practice, human-centred design (HCD), for example, has been increasingly critiqued as unable to consider or address the systemic global issues we face today. Norman (2005) was one of the first to acknowledge the blind spots of HCD, drawing attention to its tendency to improve situations for one group at the expense of others (predominantly, those who have not engaged in a monetary transaction related to the use of that which has been designed). Norman also identifies that human

needs are not fixed, and that designs that are appropriate today may not be so in future. However, Norman then discounts these concerns as less serious than the needs of users engaged in tasks whose complexity cannot be addressed by conventional HCD approaches.

The global consequences of the climate crisis have also led to the criticism that HCD ignores externalities experienced by non-human actors and systems. This is especially apposite in the context of the NIP, which, as alluded to later here, invites us to consider those who are not-yet-born as non-human. The NIP challenges the conceptual possibility of using HCD approaches to design for future people.

Wood's Meta Design emerged from an understanding that HCD's intention of reducing complexity to increase usability leads to a focus on the individual consumer, and that design should instead create a discourse that considers the relations between things (2008). Planet-centric Design (Huber,2021), Xenodesign (Schmeer,2019), More-than-Human Design (Tarcan et al,2022) and Responsible Design (UAL,2019) have further built on this foundation, requiring designers to consider the longer-term impacts of their work and to de-centre the human while considering other species, environments and ecosystems as stakeholders. Boehnert et al's Six Principles of Responsible Design (2022) are of particular relevance to this paper, requiring designers to engage with planet-centric and pluriversal future timescales, and explicitly calling for engagement with the ethical consequences of their work. As design practice increasingly expands beyond the narrow focus of HCD, and engages with the systemic nature of the interventions it proposes and enables, it therefore becomes increasingly necessary for designers to become familiar with the ethical thought that precedes their engagement.

In order to explore this potential relevance and usefulness of thought experiments in design ethics, two of Derek Parfit's thought experiments in relation to the NIP will now be examined. Variations of these thought experiments have been employed by Buwert as part of design ethics training workshops over a number of years alongside other philosophical thought experiments such as the 'Trolley Problem' (Foot,1967), the 'Drowning Child' (Singer,1997), and the 'Veil of Ignorance' (Rawls,1999). In these workshop sessions, participants explore and discuss each thought experiment in small groups, considering in each case what they believe the correct course of action to be, examining points of disagreement and what this reveals about our moral reasoning, while reflecting on how this connects to real-world design practice (Buwert et al.,2021) The following examination of Parfit's two cases is not a full exegesis of the philosophical implications of each example, but rather a reflection of common ways in which participants have engaged with these examples. At first glance, the philosophical challenge of the NIP may appear to have little relevance towards design. However, on closer engagement a profound and potentially unsettling significance emerges for the very real ethical implications of design activity as it shapes the future.

3. The Non-Identity-Problem

The NIP forces us to mentally work through the implications of the apparently banal observation that our actions today shape the course of the future. Parfit (1976a,1976b,1984) presents the NIP by constructing a series of thought experiments that guide us to consider the implications of the realisation that our present-day actions shape not only the material and social conditions of those who will live in the future, but also their very identities. Every decision made today, however small, opens up and closes off various potential futures. Although perhaps seemingly negligible in their individual impact, the cumulative effect of tiny differences in our daily lives multiplied across a population quickly snowball to result in different combinations of people meeting in different places at different times, forming different partnerships, and having different children who go on to live different lives. Over the course of a few generations the individual identities of the members of a population will be entirely different than they might otherwise have been. The NIP challenges us to consider what our moral obligations are to these future people whose existence and identity are as yet undetermined, but are in some part formed by our decisions and actions in the present. How can we act with the best interests of future people in mind, when we do not know who these people are or what their needs and desires might be?

Questions of our moral obligations towards 'possible future people'1 (who do not yet exist, but whose future lives and conditions of existence we can imagine) are a central theme which Parfit repeatedly returns to within his work, presenting a number of thought experiments examining the challenge of the NIP (Parfit,1976a,1976b,1984). Here we consider just two of these as they appear in Parfit's 1976 consideration of the "person-affecting principle": the principle that in acting towards the future "We should do what harms people the least and benefits them the most" (1976b, p.371). This pair of thought experiments present two variations of the same underlying challenge dressed up in the guise of different analogies. By working through each and then comparing the variations, the thinker is required to consider how, and on what grounds, their intuitions and reasoning might differ between the two cases. The following discussion of the two thought experiments scenarios as set out by Parfit, is drawn from and informed by the discussions, interpretations, intuitions, concerns and reflections of our participants across multiple workshop sessions.

3.1 Thought Experiment 1: The woman who intends to become pregnant

Parfit presents his first case in this way:

> "The first involves a woman who intends to become pregnant as soon as possible. She learns that she has an illness which would give to any child she conceives now a certain handicap. If she waits for two months, the illness would have passed, and she would then conceive a normal child.

[1] Parfit makes a technical distinction between 'possible' and 'future' people (Parfit, 1976b), but for simplicity's sake here we will refer only to 'future people' as a single category of those future generations of people not yet born.

> Suppose she decides not to wait – suppose that she knowingly conceives a handicapped rather than a normal child. Has she thereby harmed her child or affected him for the worse?" (Parfit,1976a, pp.100-101)

This is a textbook delivery of a thought experiment. A narrative scenario is sketched out in just enough detail to engage us in considering the specific challenge set before us. While as part of our rational deliberation we may consider all the different potential outcomes resulting from the scenario, in our search for an answer we cannot add or take anything away. Thought experiments are restricted fictions. We have to play by the author's rules. For example, we cannot suggest the mother should procure a miracle medicine which will instantaneously cure her of her illness.2

The mental task set to us is simply to consider whether the mother – if she does knowingly conceive a child whose life may foreseeably be more difficult – has harmed or affected her child for the worse. There are two possible outcomes here: either the woman has harmed her child, or she has not harmed her child. Our answer will depend on two things: the more obvious of the two is our definition of 'harm' or 'affected for the worse'. The second, less obvious, is the question of which child we are talking about.

We should pause here to recognise that Parfit, writing in 1976, uses language (specifically the terms "handicapped" and "normal") in ways that we would not today. The social model of disability recognises that disability is not an attribute of an individual, rather it is the physical and social world which disables individuals (Bogdan & Biklen,2013). To help clarify the case under examination and avoid this problematic language, let us instead specify that the child conceived now would be born with polydactyly, the most common genetic limb abnormality (Umair et al.,2018). In our case, the child will be born with six fully functional fingers on each hand. Polydactyly need not be a disability, but may be experienced as such. For example: the child may be made fun of or singled out throughout their lifetime for being different, and may at times be disabled by a world which designs objects and experiences almost exclusively for a five-fingered population.

It feels intuitively obvious that if we can imagine two lives, one of which is less difficult than the other, this less difficult life would be 'better', and conversely a life featuring obstacles which will make it more difficult would be 'worse'. On the question of harm, we could then argue that if the woman chooses to conceive the child who will have the more challenging life, she will have decided to affect the child for the worse. When discussing this scenario in workshops, the intuitive feeling of the majority of participants is normally that it would be preferable for the woman to wait and conceive a child in two months' time when her illness has passed. It is generally felt that on balance this produces a better outcome: a child is born without a physical characteristic which may make life difficult for them. And so, on these

[2] Parfit does have another thought experiment which draws our attention to a different aspect of the NIP in which prospective mothers have a chance of suffering from one of two medical conditions which may detrimentally affect babies conceived: one condition can be cured by taking a medicine, the other condition will pass within two months. See the case of the "Medical Programmes" in (Parfit,1984, p.367)

grounds, the intuitive first impression answer to the question of whether the mother who chooses to knowingly conceive the six fingered child has 'harmed or affected them for the worse' is: yes.

However, upon closer consideration things are not as clear cut. Regarding the question of the identity of the child, two different potential children are mentioned in the scenario: the child conceived now, and the child conceived in two months' time. Due to the basic facts of human biology, we are not talking about two alternative states of being for the same child (one six-fingered, one five-fingered) but two entirely different children conceived from different egg/sperm combinations. And we are only asked to consider the case of the first conception. If this child is not conceived now, it will never exist.

The choice is therefore between either a good life worth living, albeit with certain foreseeable difficulties, or no life at all. Seen in these terms it is no longer obvious that this child would be 'harmed' by being brought into existence. Parfit's thought experiment challenges us to consider whether an 'imperfect' life is better or worse than no life at all. It is of course possible to imagine scenarios in which a life is characterised by constant unbearable pain and misery and to argue that some lives may not be worth living. Noting the possibility of these extreme cases, participants do generally agree that a life which is 'imperfect' but still worth living is preferable to no life at all, and that we would not normally say that a child brought into being under less than perfect conditions has been 'harmed'.

The core challenge of the Non-Identity Problem is the question of what our responsibilities and obligations are to those future people who have not yet been born and who therefore as of yet have no identity and existence, only imagined potential. How should we understand the person affecting principle that "we should do what harms people the least and benefits them the most" (Parfit,1976b, p.371), when we may now have established that the bar for harming people who do not yet exist is set very high?

3.2 Thought Experiment 2: Two social policies

Let us now turn to Parfit's second case:

> "Suppose we have a choice between two social policies. These will alter the standard of living – or, more broadly, the quality of life. The effects of one policy would, in the short term, be slightly better, but, in the long term, be much worse. Since there clearly could be such a difference between two policies, we need not specify details. It is enough to assume that, on the "Short Term Policy," the quality of life would be slightly higher for (say) the next three generations, but be lower for the fourth generation, and be *much* lower for several later generations." (Parfit,1976a, pp.101-102)

Should we choose to pursue the 'Short Term Policy' which benefits people now and in the near future but significantly affects more distant future people for the worse, or an alternative 'Long Term Policy' which rations and restricts benefits available to us now and in the near future, in order to secure greater benefits to those in the more distant future?

While Parfit specifically chooses not to specify the nature of the social policy, discussion in our workshops typically focuses on natural resource usage and climate change.3 In this context, a majority of participants generally express fairly strong intuitive leanings towards pursuing a longer-term policy which restricts benefits now (affecting us for the worse), in favour of benefits for distant possible future people. This is the logic upon which the broad sustainability narrative is constructed: a degree of pain and sacrifice is necessary now to ensure long term benefits (survival) for the future of humanity. A majority (though not all) generally agree that we should, for example, make sacrifices now to reduce emissions which cause the global atmosphere to heat, in order to avoid climate change, ecosystem collapse, and mass extinctions. We deny our present and near-future selves benefits as we seek to undo damage, stabilise and sustain our natural environment. In this way we hope to benefit distant future people by affording the opportunity to thrive on a sustainable planet earth. Conversely, it is generally felt by eco-conscious participants that if we choose to benefit ourselves by pursuing a Short Term Policy of resource consumption to maintain or enhance our current lifestyles, future generations will experience very real harm as temperatures and sea-levels rise, ecosystems collapse and mass extinctions ensue.

This is the point at which our thought experiment takes a mischievous, subversive and counterintuitive turn, with uncomfortable repercussions for design. Parfit's pair of scenarios challenge us to test and cross-check our reasoning across the two cases before us. Consideration of the first case leads many to conclude that a life which is difficult in some regards, but still worth living, is preferable to no life at all. Therefore people who come into being under less-than-perfect conditions should not be considered to have been harmed or affected for the worse, as they would not otherwise have existed. Now, in presenting the social policy case, Parfit challenges the thinker to follow the same line of reasoning to its rational conclusion:

> "The particular members of the fourth and later generations, on the Short Term Policy, would not have been born at all if instead we had pursued the Long Term Policy. Given the effects of the policies on the details of people's lives, different marriages would increasingly be made. More simply, even in the same marriages, the children would increasingly be conceived at different times. As we argued, this would be enough to make them different children.
>
> We can now apply a "person-affecting" principle. The members of the later generations would be different on the different policies. So if we pursue the Short Term Policy there will never be anyone who is worse off than he would otherwise have been. The Short Term Policy harms no one. Since it benefits certain people (those who exist now), it is the policy chosen by our principle." (Parfit,1976a, p.102)

Parfit's NIP thought experiments have backed us into a corner. We have agreed with the argument that a difficult life worth living is better than no life. Using the very same reasoning it is also possible to argue that design decisions which lead to suboptimal, even genuinely awful, foreseeable futures do not technically affect anyone in those futures for the worse,

[3] Elsewhere, in variations of this thought experiment Parfit does explicitly frame the challenge in terms of natural resource depletion. See: (Parfit,1984,p.362).

because the people in these futures are different people whose lives, although difficult, are still worth living.

Our design decisions change the nature of the future population. Each choice in the present is a fork in the road with paths leading to alterative future states inhabited by distinct potential populations resulting from the diverging alternative conditions of existence set in motion by our decision. We are therefore not comparing relative levels of harm and benefit applied to the same set of individuals. In what sense do we harm a future person if we choose the policy which leads to their reality and brings them into existence under imperfect conditions, as opposed to not existing at all? A rational argument can be made that as their suboptimal life is a significant improvement on non-existence, then our design decision which contributes to the creation of this can only be judged a good thing for them.

It feels intuitively right that we have moral obligations to choose to do the thing which "harms people the least and benefits them the most" (Parfit,1976b, p.371). All other things being equal it would follow that we should seek to avoid harm and increase benefit for both present and future people. What Parfit's thought experiments demonstrate is that when we factor in the effect of present design decisions on the identity of future populations, all things are not equal. The Non-Identity Problem is a problem, because it appears to show that our obligations to future people who have not yet been born, are quite different to our obligations to future versions of people who already exist (you and I tomorrow: the humans around whom HCD is typically centred). If we can design the world in a way which concretely benefits some existing people, and apparently harms no future people (because our design interventions in the world have altered the identity of the future population), do we have a moral obligation to design with this short-term priority, even if this sets us on a foreseeable path towards evidently miserable long-term futures?

To be clear, Parfit (along with the majority of participants) is not satisfied that we should accept this reasoning, describing the outrageously counterintuitive conclusions which the NIP can lead us to as "repugnant" and "absurd" (1984,pp.381-418). The NIP challenges our moral intuitions, leaving thinkers bewildered, confused, disoriented and generally far less certain of what the right thing to do is than when we began. Why then should ethically concerned designers engage with this abstract mind-game exercise at all?

4. The value of thought experiments in design ethics

The NIP has very real implications for future-oriented design activity as it forces us to question received wisdom and carefully examine our intuitions about design's responsibilities to future people. Participants almost universally agree that while the premises of the argument seem reasonable, the conclusions are unacceptable. The implication that we effectively cannot 'harm' future people by our actions, and therefore wouldn't be doing wrong by knowingly acting in ways which would degrade the conditions of their future existence, just feels wrong. But why? If we maintain an intuition that we should strive to do good towards our future world, but we cannot rationally justify this, where do we stand? If we discover that

the foundations of our understandings of harm, benefit, obligation, responsibility, right and wrong in design ethics are shaky, where do we go from here?

Thought experiments are always filtered through the lens of the thinker's own interests and experience. Participants in our workshops bring a diverse range of future-oriented practical dilemmas and challenges, drawn from everyday practice, into their discussions around the NIP, for example: the role of alcohol marketing in a nation's economy and health; the paradox of resource extraction for green technologies through open cast mining; the balancing of cultural heritage against urban development. With the increasing accessibility of Generative Artificial Intelligence platforms in recent years, GenAI has become a prominent theme in recent workshops, with participants often discussing the implications of emergent technologies for design's task in shaping futures. For example, participants might deliberate on thinking along the lines of Marc Andreessen's Techno-Optimist Manifesto which argues that "any deceleration of AI will cost lives. Deaths that were preventable by the AI that was prevented from existing is a form of murder."(2023) Such arguments are simultaneously seductive to those excited by the possibility of new technologies, and difficult to challenge for those unpracticed in philosophical debate. Parfit's NIP provides thinkers both with the capability to reflect on the validity of such statements, and to frame responses that engage with the complexity they seek to obscure (is it possible, through design inaction, to murder someone who isn't yet alive?).

Designers are often called on to facilitate discussions and creative activities involving people from different demographic and cultural backgrounds, as well as those whose roles provide different perspectives on the importance of human, and more-than-human, centredness. As the designer's role becomes increasingly strategic and service-oriented, an appreciation of the philosophy of ethics grows in importance, as a method for helping people step outside their familiar framing of issues and to understand they are involved in 'bigger picture' problems without simple solutions.

It falls outwith the scope of this paper to provide a systematic account of the ways in which the employment of thought experiments can directly and indirectly impact design practice. This more systematic exposition is a task for future research. Our aim here is to introduce and illustrate the potential of the NIP as a topic of interest to the field of future oriented design ethics research.

The NIP provokes deep consideration of a range of foundational issues across the spectrum of ethical theory, yet does so in a remarkably accessible way. Yes, the puzzle of the thought experiment is tricky. It requires focus and brain-power, but it does not require any prior philosophical training or understanding of technical jargon. The self-contained challenge is enough to stimulate trains of thought leading organically to philosophically significant realisations. Without any knowledge of the technical language surrounding normative ethical theories such as consequentialism and deontology, participants can come to articulate fairly sophisticated critiques of these on their own terms and in their own everyday language. This is not an argument for anti-intellectualism. Quite the opposite: our experience has been that

thought experiments are a potent gateway drug luring designers to deeper intellectual engagement with ethical philosophy, in ways which remain relevant to their own practical design activity.

In over four decades of philosophical debate, many 'solutions' to the NIP have been presented, advancing interesting and valuable perspectives, yet none of these contributions have conclusively 'solved' the problem (Adams,1979; Kavka,1982; Mulgan,2006; Sikora & Barry,2012). When considering the value of thought experiments in the context of design ethics, our interest is not so much in the philosophical resolution of the problem as in the potential value of the process of struggling to address the challenge. It has consistently been our experience that participants find and express their appreciation of the use of thought experiments in design ethics, not in terms of direct application towards resolution of specific design ethics challenges, but rather in terms of opening up, exercising, and increasing confidence in employing these alternative ways of thinking as part of their holistic day-to-day design process. Thought experiments such as the NIP are accessible thinking devices providing engaging, dynamic structures through which we can thoughtfully and reflectively engage with the ethical challenges of wicked problems as encountered in the complexity of real-world design. The NIP represents the ability of thought experiments to defamiliarise and disorientate our existing ways of thinking, creating a space of creative disequilibrium in which to enable conversations in exciting and productive ways as we are challenged to reimagine our conception of future-oriented design ethics.

5. References

Adams, R. M. (1979). Existence, Self-Interest, and the Problem of Evil. *Noûs*, *13*(1), 53. https://doi.org/10.2307/2214795

Andreessen, M. (2023, October 16), *The Techno-Optimist Manifesto*. Andreeson Horowitz. https://a16z.com/the-techno-optimist-manifesto/

Boehnert, J., Sinclair, M. and Dewberry, E. (2022), Sustainable and responsible design education: Tensions in transitions. *Sustainability*, 14 (11), 6397

Bogdan, R., & Biklen, D. (2013). Handicappism. In M. Wappett & K. Arndt (Eds.), *Foundations of Disability Studies* (pp. 1–16). Palgrave Macmillan.

Buwert, P. (2016). *Ethical design: A foundation for visual communication.* [PhD]. Robert Gordon University.

Buwert, P. (2017). Potentiality: The ethical foundation of design. *The Design Journal*, *20*(sup1), S4459–S4467. https://doi.org/10.1080/14606925.2017.1352942

Buwert, P., Bianchin, M., & Heylighen, A. (2021). EXPLORING PHILOSOPHICAL DEVICES FOR THINKING ABOUT ETHICS IN. *SED2 Conference Proceedings*, *2*, 91–101.

Chan, J. K. H. (2023). The ethics of wicked problems: an exegesis. Socio-Ecological Practice Research, 5(1), 35–47. https://doi.org/10.1007/s42532-022-00137-3

Coyne, R. (2005). Wicked problems revisited. *Design Studies*, *26*(1), 5–17. https://doi.org/10.1016/j.destud.2004.06.005

Dennett, D. C. (2013). *Intuition Pumps and other Tools for Thinking*. Penguin.

Dilnot, C. (2011). Sustainability and Unsustainability in a World Become Artificial: Sustainability as a Project of History. Design Philosophy Papers, 9(2), 103–155. https://doi.org/10.1108/00070701111148441

Edmonds, D. (2014). *Would you kill the fat man?* Princeton University Press.

Fry, T. (2009). *Design Futuring: Sustainability, Ethics and New Practice*. Berg.

Fodor, J. A. (1964). On Knowing What We Would Say. *The Philosophical Review*, *73*(2), 198–212.

Foot, P. (1967). The Problem of Abortion and the Doctrine of the Double Effect. *The Oxford Review*, *5*, 5–15.

Harman, G. (1986). Moral Explanations of Natural Facts—Can Moral Claims be Tested Against Moral Reality? *The Southern Journal of Philosophy*, *24*(5), 57–68.

Huber, S. (2021, January 28), *What is planet-centric design?* samuelhuber.medium.com. https://medium.com/goodpatch-global/what-is-planet-centric-design-8d1754b52fba

Kavka, G. (1982). The Paradox of Future Individuals. *Philosophy & Public Affairs*, *11*(2), 93–112.

Königs, P. (2023). Of trolleys and self-driving cars: What machine ethicists can and cannot learn from trolleyology. *Utilitas*, *35*(1), 70–87. https://doi.org/10.1017/S0953820822000395

Miscevic, N. (2022). *Thought Experiments*. Springer International Publishing. https://doi.org/10.1007/978-3-030-81082-5

Mulgan, T. (2006). *Future People: A Moderate Consequentialist Account of our Obligations to Future Generations*. Oxford University Press.

Norman, D.A. (2005), Human-centered design considered harmful. *Interactions,* 12(4), pp.14-19.

Parfit, D. (1976a). On doing the best for our children. In M. Bayles (Ed.), *Ethics and Population* (pp. 100–115). Schenkman.

Parfit, D. (1976b). Rights, Interests, and Possible People. In S. Gorovitz (Ed.), *Moral Problems in Medicine* (pp. 369–375). Prentice-Hall.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Plunkett, D., & Greene, J. D. (2019). Overlooked Evidence and a Misunderstanding of What Trolley Dilemmas Do Best: Commentary on Bostyn, Sevenhant, and Roets (2018). *Psychological Science*, *30*(9), 1389–1391. https://doi.org/10.1177/0956797619827914

Rawls, J. (1999). *A theory of justice* (Rev. ed). Harvard University Press.

Schmeer, J. (2019), Xenodesignerly ways of knowing. *Journal of Design and Science*, 5, pp.1-21

Scruton, R. (2014). Parfit the Perfectionist. *Philosophy*, *89*(4), 621–634. https://doi.org/10.1017/S0031819114000266

Sikora, R. I., & Barry, B. (2012). *Obligations to Future Generations*. White Horse Press.

Singer, P. (1997). The Drowning Child and the Expanding Circle. New Internationalist, 289, 28–30.

Sorensen, R. A. (1999). Our Most Curious Device. In R. A. Sorensen, *Thought Experiments* (1st ed., pp. 7–20). Oxford University PressNew York. https://doi.org/10.1093/019512913X.003.0002

Tarcan, B., Pettersen, I.N. and Edwards, F. (2022), *Making-with the environment through more-than-human design*, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), DRS2022: Bilbao, 25 June - 3 July, Bilbao, Spain

Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, *59*(2), 204–217.

Umair, M., Ahmad, F., Bilal, M., Ahmad, W., & Alfadhel, M. (2018). Clinical Genetics of Polydactyly: An Updated Review. *Frontiers in Genetics*, *9*, 447. https://doi.org/10.3389/fgene.2018.00447

UAL. (2019), *Responsible Design Framework*. University of the Arts London. https://www.arts.ac.uk/__data/assets/pdf_file/0023/63716/ResponsibleDesignFramework.pdf

Walsh, A. (2011). A Moderate Defence of the Use of Thought Experiments in Applied Ethics. *Ethical Theory and Moral Practice*, *14*(4), 467–481. https://doi.org/10.1007/s10677-010-9254-7

Wood, J. (2008), Changing the change: a fractal framework for metadesign. In *Changing the Change Conference*, Institute of Biotechnology, Torino, Italy, 3-12th July 2008.

Willis, A.-M. (2006). Ontological Designing—Laying the ground. *Design Philosophy Papers*, *3*(January), 80–98. https://doi.org/10.2752/144871306X13966268131514

Yelavich, S., & Adams, B. (Eds.). (2014). Design as future-making. Bloomsbury.

About the Authors:

Peter Buwert is a design educator and researcher whose expertise lies in the ethics of design. His research explores the extent to which ethical foundations can be discovered and productively activated from within the activity of designing itself.

Matt Sinclair is Associate Professor and Head of Design, Photography & Advertising at Edinburgh Napier University. His research is situated within the field of Responsible Design Futures, generating sustainable scenarios to challenge the 'preferable' futures advertised by the technological status-quo.