



Delft University of Technology

Document Version

Final published version

Citation (APA)

Marchiori, S. (2026). *Conceptual engineering in the philosophy of technology*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:5b4318fd-c2b8-43bf-b8c4-d051a687dbfe>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

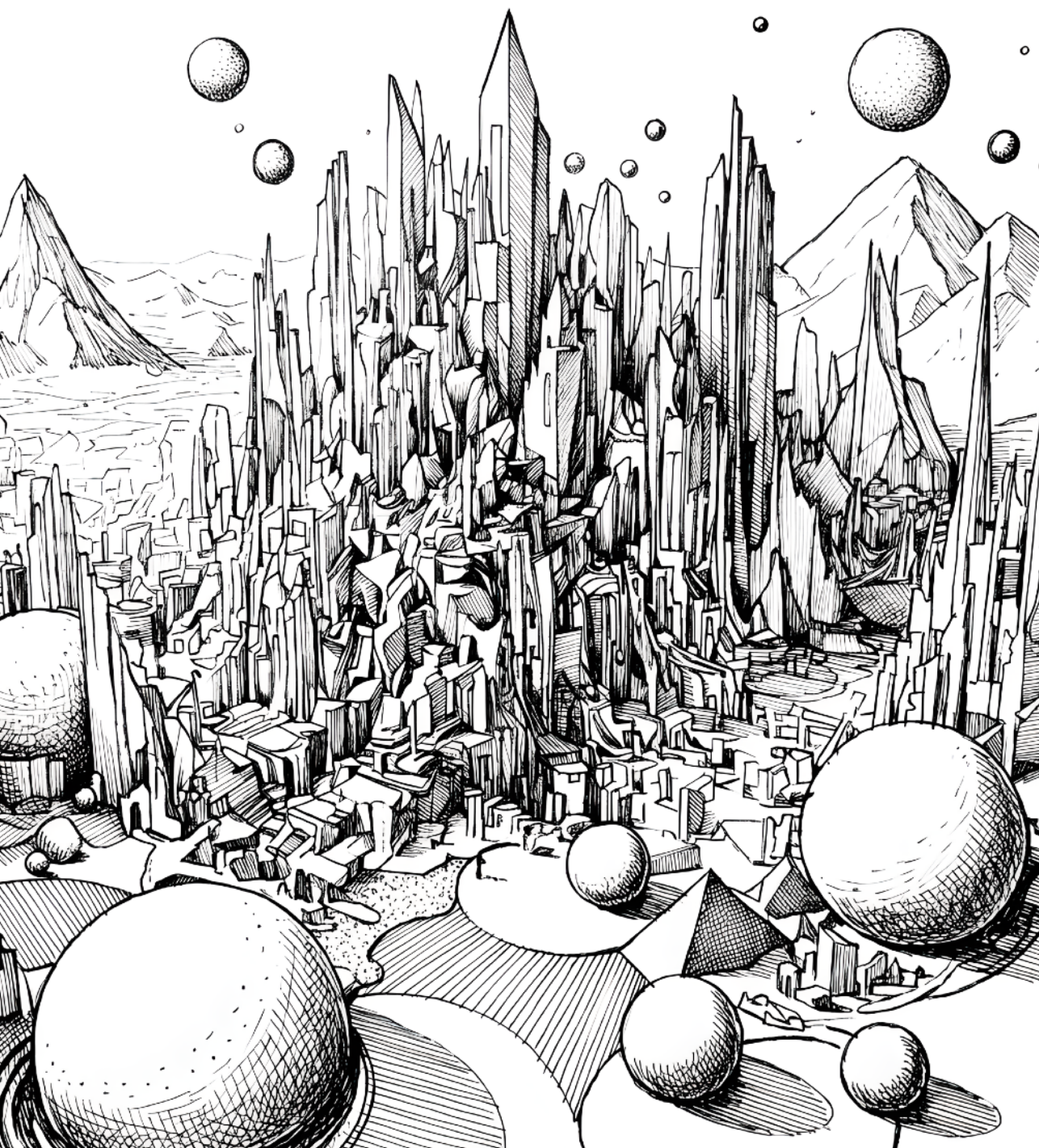
Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

CONCEPTUAL ENGINEERING *in the* PHILOSOPHY *of* TECHNOLOGY

Samuela Marchiori



This page intentionally contains only this sentence.

Conceptual engineering in the philosophy of technology

Conceptual engineering in the philosophy of technology

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus,
Prof.dr.ir. H. Bijl,
chair of the Board for Doctorates
to be defended publicly on
Thursday, August 27th at 13:00

by

Samuela MARCHIORI

This dissertation has been approved by the (co)promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. I. R. van de Poel	Delft University of Technology, <i>promotor</i>
Prof. dr. M. B. O. T. Klenk	University of Technology Nuremberg, <i>copromotor</i>

Independent members:

Prof. dr. M. Simion	University of Oxford
Dr. E. R. F. O'Neill	Technical University Eindhoven
Prof. dr. J. H. Anderson	Utrecht University
Prof. dr. mr. ir. N. Doorn	Delft University of Technology
Prof. dr. M. E. Warnier	Delft University of Technology, <i>reserve member</i>



This work is part of the research programme Ethics of Socially Disruptive Technologies, which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

Keywords: philosophy of technology; conceptual engineering; conceptual ethics; methods; ethics of technology; conceptual disruption; concepts; artefacts; functions; affordances

Printed by: Expresta (expresta.eu)

Design: S. Marchiori

ISBN: 978-94-6518-394-7 (ebook)

ISBN: 978-94-6384-997-5 (paperback)

© Samuela Marchiori, 2026

All rights reserved. No part of this publication may be used as input, training, fine-tuning, or evaluation data for any artificial intelligence system, machine learning system, large language model, generative AI systems, or analogous technology, without the prior written permission of the copyright holder. This includes, but is not limited to, uses such as processing for purposes of extraction, interpretation, summarisation, text generation, or translation. No part of this publication may be stored, distributed, or made available, whether in digital or physical form, behind any paywall, subscription service, or other access-restricted or fee-based means, without the prior written permission of the copyright holder.

An electronic version of this dissertation is available at: [http:// repository.tudelft.nl](http://repository.tudelft.nl)

Ai miei genitori
To my parents

*But I maintain that she would come if we worked for her,
and that so to work, even in poverty and obscurity, is worth while.*

Virginia Woolf, A Room of One's Own

Contents

Acknowledgements

Summary

Samenvatting

List of Figures

List of Tables

List and status of the papers used in the dissertation

Chapter 1. Introduction

1.1 Setting the stage

1.1.1 Ethics of Socially Disruptive Technologies (ESDiT)

1.1.2 Significance beyond ESDiT

1.1.3 A note to the Reader

1.2 Philosophy of conceptually disruptive technologies

1.3 From philosophy of technology to conceptual engineering and back

1.3.1 What concept of concept for conceptual engineering in the philosophy of technology?

1.4 Themes and guiding questions

1.5 Summary of chapters

PART I. FOUNDATIONS

Chapter 2. What is conceptual disruption?

2.1 Introduction

2.2 Background on conceptual disruption

2.3 Conceptual disruption and constitutive principles

2.4 Conceptual disruption. A framework

2.4.1 Levels of conceptual disruption

2.4.2 Types of conceptual disruption

2.4.2.1 Conceptual gaps

2.4.2.2	Conceptual conflicts
2.4.2.3	Conflicts, gaps, and conceptual change
2.4.3	Severity of conceptual disruption
2.4.3.1	Mild conceptual disruption
2.4.3.2	Moderate conceptual disruption
2.4.3.3	Severe conceptual disruption
2.4.4	Comparison
2.5	The role of conceptual engineering
2.6	Conclusion
 Chapter 3. Concepts as artefacts	
3.1	Introduction
3.2	Concepts as (socio-technical) artefacts
3.2.1	Concepts: broad and narrow accounts
3.2.2	Artefacts: broad and narrow accounts
3.3	Conceptual engineering meets the philosophy of technology. Two approaches, three insights
3.3.1	Insight I: Concepts bear affordances and have side effects
3.3.2	Insights II: Concepts are value-laden
3.3.3	Insight III: Conceptual engineering would benefit from greater systematicity
3.4	Conclusion
 Chapter 4. Conceptual affordances	
4.1	Introduction
4.2	From conceptual functions to conceptual affordances
4.2.1	Affordances
4.2.1.1	Types
4.2.1.2	Degrees
4.2.1.3	Dynamism and multi-stability
4.2.2	In which sense do concepts enable or constrict our actions and thoughts? Concepts, language, and society

4.3 Conceptual affordances-informed functionalism

4.3.1 Functions (still) matter. Why affordances should not replace functions

4.4 Conclusion

Chapter 5. Conceptual resilience

5.1 Introduction

5.2 From conceptual stability to conceptual resilience

5.3 A graded account of conceptual resilience

5.4 Modes of resilience. How do concepts exhibit resilience?

5.4.1 Conceptual Resilience as Immutability (CRI)

5.4.2 Conceptual Resilience as Adaptability (CRA)

5.4.3 Relation between CRI and CRA

5.5 (Re-)engineering for resilience. Conceptual engineering as prevention and cure

5.5.1 Thick and thin normative concepts

5.5.2 Levels of analysis. Individual concepts, conceptual clusters, conceptual scheme

5.5.2.1 Individual concepts

5.5.2.2 Conceptual clusters

5.5.2.3 Conceptual scheme

5.6 Conclusion

PART II. SYNTHESIS

Chapter 6. Navigating conceptual disruption with conceptual engineering

6.1 Introduction

6.2 Synthesis. Theoretical foundations

6.2.1 Conceptual adequacy. Functions and affordances

6.2.1.1 Concepts, conceptions, sub-concepts

6.2.2 Conceptual disruption

6.2.2.1 Types, levels, severity

6.2.3 Conceptual engineering interventions

6.2.3.1 Conceptual evaluation and conceptual preservation

6.2.3.2 De novo engineering, conceptual re-engineering, and conceptual abandonment

6.2.3.3 Conceptual implementation

6.3 Operationalisation

6.3.1 Case study. AUTHORSHIP and translation

6.3.2 Case study. AUTHORSHIP and AI(-assisted) translation

6.4 Conclusion

Chapter 7. Conclusion

7.1 Summary and key findings

7.2 Limitations

7.3 Prospects for future research

References

About the author

List of publications

Acknowledgements

To say that this PhD has been an utter delight would not do justice to the past four years. Many people are to blame for my genuine struggle to conceive how anything may live up to this, all of whom I encourage to try and prove me wrong.

First and foremost, my supervisors, Ibo van de Poel and Michael Klenk, who proved me wrong countless times and exhibited unending patience while I slowly figured it out. *Ibo*, thank you for leading by example, for sharing your seemingly boundless expertise with me, and for providing me with a safe and supportive environment during theoretical explorations and unplanned field work alike. *Michael*, thank you for challenging me, for reminding me to keep my eyes on the prize, and for rejoicing for every success as if it were your own. Thank you both for accompanying me on this journey, and for all the time and effort you have dedicated to me over these years.

My appreciation also goes out to the members of the Committee: Joel Anderson, Neelke Doorn, Elizabeth O'Neill, Mona Simion, and Martijn Warnier. Thank you for taking the time to read and engage with my work. I look forward to being demolished.

I am very thankful to Jeroen Hopster for the guidance, inspiration, and support as a mentor and co-author. Working alongside you has been an immense pleasure.

I am also indebted to my co-authors for allowing me to explore fascinating topics and questions that I would not have had the time and/or expertise to adequately tackle on my own. In particular, I owe a great deal to Kevin Scharp for his kindness and generosity.

It is not lost on me that being part of ESDiT and TU Delft gave me access to opportunities and a reach to my work that I would not have been able to aspire to otherwise. My appreciation goes out to both for providing me with a sense of security that allowed me to take more risks and more serenely weather the uncertainty that follows from working in an emerging research niche.

I am also deeply grateful to those who paved the way for me to be able to do the research I do and be taken seriously as a philosopher of technology, among which Filippo Santoni de Sio, Ibo van de Poel, Jeroen van den Hoven, Neelke Doorn, Peter-Paul Verbeek, Philip Brey, Pieter

Vermaas, Sabine Roeser, and Wybo Houkes, all of whom I was fortunate enough to cross paths with over the years. Your efforts, past and present, do not go unnoticed. Thank you.

My appreciation goes out to a great many people.

To my colleagues at the Ethics and Philosophy of Technology Section. To Lissy, Monika, Nathalie, Sophia, and particularly to Monica, for the constant support and kindness, and for always cheering me on. Walking past the secretariaat was always a joy: you turned chores into pleasures. To the TPM ICT magicians, for the frequent help over the years. To Nico, for the conversations over fonts and data, or lack thereof. To Morag, for the support towards the end of the journey. To 4TU.Ethics, OZSW, the Delft Design for Values Institute, and the Delft Digital Ethics Centre, particularly Helma Dokkum, Jeroen van den Hoven, and Stefan Buijsman. To the Klenk gang: Joost, Michael, Nicolien, and Ryan. To my buddies: Karen, Maria, and Thijs. To those who walked this path with me since day one, particularly Aarón, Bauke, Dmitry, Giorgia, Jordi, and Miao. To Aafke, Benedict, Eugen, Filippo, and Olya, for the inspiration. To Pieter, for being a delightful neighbour. To Aishwarya, Ben, Belén (and Jon), Llona, Maria, and Marc, for being wonderful office mates.

To Herman Veluwenkamp, for the unique thrill that comes from dangerously overlapping research interests. To Christian Nimtz, Delia Belleri, Irene Olivero, Manuel Gustavo Isaac, Ryan Nefdt, and Steffen Koch, for the warm welcome to conceptual engineering. To Grant Snider, for the art. To Gerdien de Vries, for the guidance on interdisciplinary research. To Martijn Warnier, for indulging my detour on concepts as socio-technical artefacts.

To Cristine Griffo, for believing in me before I did. To Dina Babushkina, for writing advice I still strive to follow. To Achille Varzi, for teaching the very first pure philosophy course I ever took. To Peter Railton, for the inspiration. To Paolo Sommaggio, for the unwavering support and encouragement since 2016, and for allowing me to keep exploring the legal ramifications of the philosophy of technology by inviting me to the Department of Private Law and Legal Theory at the University of Padua and the Faculty of Law at the University of Trento.

To Sven Nyholm and the AI Ethics research group, for hosting me during my stay at LMU Munich. To Giulio Mecacci and Vincent Müller, for inviting me to present my work at Radboud University and FAU Erlangen-Nürnberg, respectively, and to their research groups for the

warm welcome. To Enrico Galvagni (and Fabio Tollon), for inviting me to the University of Edinburgh, and to everyone at the Centre for Technomoral Futures, particularly Gina Helfrich, for their kindness and hospitality.

To Filippo Santoni de Sio, Jeroen Hopster, Juan Durán, Lavinia Marin, Martin Sand, Michael Klenk, Paolo Sommaggio, Steffen Steinert, and Udo Pesch, for giving me the chance to teach in their courses. To Andrea Gammon, for the endless support while doing so. To our students, for filling me with hope for the future.

To those who joined on the way, and those who were there all along. In particolare, ad Angela (e Leo), non saprei neanche da che parte iniziare: grazie. To Pietro, for sharing the peculiarity of having (the same) two homes. To Llona, for teaching me the art of small talk. To Ryan, for the first walk-and-talk and all those that followed. To Maria and Nick, for the blissful conversations and the future reading groups. To Sadjad, for the talks over (great) coffee. To Teun, for adopting me on my first day in Enschede and never letting me go. To Aliah, Anna, Emma, and Kritika, for the kindness and solidarity. To Fabio, for the camaraderie. To Aishwarya, for a friendship that defies continents. Ad Agata, Chiara, e Marialuisa, per il supporto e per ricordarmi che la salute viene prima di tutto. Ad Armanda, per l'affetto, la fiducia, e i consigli. To Robin, for being with me through thick and thin, conceptually and otherwise.

My gratitude goes out to more people than I am able to name.

To the anonymous peer reviewers who read my manuscripts, for improving them greatly with their feedback. *Un no se 'l dà.*

To my dealers, for the crucial support in the form of coffee, pastries, and oliebollen. *Doi no se i tol.*

To my family, for everything. *Trei i ven dal cor.*

What an absolute privilege and a joy it has been to share this journey with you all. My academic bucket list is much lighter than it was four years ago.

Thank you all so very much, I don't take any of this for granted.

Sam

Edinburgh, Spring 2026

Summary

While a rich scholarship has examined how technology impacts our practices, norms, values, and beliefs, considerably less explicit attention has been devoted to investigating how technology impacts our concepts, i.e., the notions through which we understand, reason about, and navigate our world. For example, the emergence of autonomous weapons systems has put pressure on the concept CONTROL adopted in international humanitarian law, which proved inadequate in establishing clear lines of responsibility for uses of force in contexts in which control over weapons was no longer exercised directly by a human agent, but distributed among human and non-human entities. Similarly, the advent of autonomous vehicles has raised pressing questions about (among others) liability regimes, safety standards, and the differential impact of this technology across different social groups. Many of such questions are tied to the concepts that underlie them.

This dissertation investigates how to navigate technology-induced conceptual disruption (i.e., the interruption in the functional adequacy of concepts) with conceptual engineering (i.e., the evaluation, design, and implementation of concepts). The dissertation is a sustained illustration of how conceptual engineering and the philosophy of technology—two disciplines which have so far proceeded on largely parallel tracks despite engaging with related problems, operating on similarly conceived objects of inquiry, and sharing broadly comparable normative aims—are better positioned to address their respective research in dialogue rather than in isolation. In particular, the intersection of conceptual engineering and the philosophy of technology represents a fertile but under-explored area of inquiry that this dissertation systematically opens up, advancing both the philosophy of technology and conceptual engineering in ways that neither discipline has thus far achieved while operating in isolation from the other. In the dissertation, this exchange proceeds through two moves.

On the one hand, the philosophy of technology literature is increasingly interested in exploring questions related to the impact of technology on the adequacy of our concepts. In particular, a newly recurring observation is that technologies can render our concepts inadequate (this is known as a *conceptual disruption*) and that, in such cases, those concepts may need to be changed. However, what it means for a concept to ought to be changed—e.g., on what grounds, towards what

end, according to which normative standards—remains underspecified in the philosophy of technology literature. This is where a first move from the philosophy of technology to conceptual engineering—i.e., normative conceptual work aimed at the improvement of concepts—becomes pertinent. Indeed, conceptual engineering has developed a vocabulary and a set of standards for evaluating, (re-)designing, and implementing concepts. However, such insights from conceptual engineering, while pertinent, are not yet present in the philosophy of technology. The first move involves bringing conceptual engineering to bear on conceptual work in the philosophy of technology, as it offers a more refined lexicon and success criteria for a practice that many in the philosophy of technology have been conducting implicitly.

On the other hand, questions such as when and how concepts ought to be changed—e.g., discussions surrounding adequacy standards for concepts, (how to identify) which functions concepts should serve, and how competing (re-)design proposals are to be evaluated against each other—are themselves debated within conceptual engineering. Here, the philosophy of technology turns out to offer unexpected resources. Indeed, the conceptual engineering scholarship routinely refers to concepts as entities that are analogous to artefacts. Yet, engagement with the rich philosophy of technology literature on (socio-technical) artefacts remains rare. This suggests that a second move is in order: from conceptual engineering to the philosophy of technology. This investigation explores the analogy between concepts and (socio-technical) artefacts further, and argues that the extensive philosophy of technology scholarship on artefacts can offer insights and resources to further ground and enrich conceptual engineering. The main innovations this dissertation brings in this respect are the introduction of *CONCEPTUAL AFFORDANCES*, to further refine the evaluation of conceptual adequacy, and *CONCEPTUAL RESILIENCE*, to evaluate the desirability of (engineering) concepts that resist conceptual disruption, and the push for more structured approaches to the practice of conceptual engineering.

Chapters 2 and 6, run primarily from the philosophy of technology to conceptual engineering, showing how the vocabulary and insights of conceptual engineering can equip philosophers of technology with sharper tools for conceptualising, identifying, evaluating, and responding to conceptual disruptions (chapter 2) in a systematic manner (chapter 6). Chapters 3, 4, and 5, run in the complementary direction: they show how insights and tools developed for the study of socio-technical artefacts can enrich the theoretical foundations and

operationalisation of conceptual engineering. They do so by drawing on the analogy between concepts and socio-technical artefacts to argue that engineered concepts are value-laden and conceptual engineering would benefit from greater systematicity (chapter 3), and by introducing CONCEPTUAL AFFORDANCES (chapter 4) and CONCEPTUAL RESILIENCE (chapter 5) in the conceptual engineering literature. As such, the investigation does not merely embed conceptual engineering in the philosophy of technology, but it also scrutinises and innovates conceptual engineering itself.

The dissertation unfolds across seven chapters.

Chapter 1 (*Introduction*) provides an overview of the philosophy of technology landscape in which this research is primarily situated, before offering a brief introduction to conceptual engineering and concepts, and illustrating an overview of the core themes and guiding questions addressed in this dissertation.

Chapter 2 (*What is conceptual disruption?*), co-authored with Kevin Scharp, provides the foundational groundwork for the dissertation. The chapter conceptualises conceptual disruption as an interruption in the functional adequacy of concepts, i.e., concepts are disrupted when they are no longer able to adequately fulfil their functions. It also offers a taxonomy of the phenomenon of conceptual disruption along three dimensions: (i) the level at which the conceptual disruption occurs (individual concept, conceptual cluster, or conceptual scheme); (ii) the type of disruption (conceptual gap and conceptual conflict, the latter including conceptual overlap and conceptual misalignment); and (iii) the severity of the conceptual disruption (from mild to severe). Beyond providing a diagnostic framework for identifying when a conceptual disruption is present and assessing how urgently it should be remedied, the chapter establishes that different types of disruption call for different responses—de novo engineering for conceptual gaps, re-engineering for conceptual conflicts—and situates conceptual engineering as the central tool for systematically identifying, assessing, and addressing conceptual disruptions.

Chapter 3 (*Concepts as artefacts*) is similarly foundational to the dissertation's broader contribution. The chapter tackles the analogy between concepts and artefacts—familiar in conceptual engineering but rarely developed in dialogue with the philosophy of technology scholarship on socio-technical artefacts—and examines what this analogy yields when pursued rigorously. After clarifying how the analysis of concepts as akin to socio-technical artefacts sets itself apart from existing

accounts of concepts as abstract cultural artefacts, the chapter develops three insights for conceptual engineering: (i) concepts, like artefacts, bear affordances and produce side effects, enabling uses beyond their intended purpose; (ii) re-engineered concepts are inherently value-laden, reflecting the priorities and assumptions of their designers and of the socio-technical contexts in which they operate; and (iii) conceptual engineering would benefit from greater methodological systematicity, drawing on the structured approaches that the philosophy of technology has developed for the evaluation, design, and implementation of artefacts. Together, these insights broaden and enrich the toolkit of conceptual engineering and set the agenda for the following chapters.

Chapter 4 (*Conceptual affordances*) argues that functionalism in conceptual engineering should be grounded in a theory of affordances. Conceptual engineering overwhelmingly adopts a functional approach to conceptual adequacy. Following functionalism, concepts have functions and are adequate to the extent that they fulfil their functions. Functionalism traditionally operates on the assumption that conceptual engineers ought to concern themselves with what concepts *should* do. This, the chapter argues, is too narrow, as it systematically neglects what concepts enable or constrain beyond the uses they should serve. By drawing on the notion of affordance as developed in ecological psychology and elaborated in the philosophy of technology, the chapter introduces the notion CONCEPTUAL AFFORDANCES, understood as the potential actions, uses, and thoughts that a concept enables or constrains relative to a given user and within a given context. This offers conceptual engineers a sharper, more holistic, and ethically attuned framework for conceptual evaluation, design, and implementation.

Chapter 5 (*Conceptual resilience*) broadens the scope of the investigation by asking whether conceptual engineering should function not only as a response to disruption, but also as a means of preventing it. The existing literature has primarily examined how conceptual engineering can mitigate or resolve disruptions after they have occurred; little attention has been given to how concepts might be designed to resist disruption in the first place. The chapter introduces the notion of conceptual resilience, understood as the capacity of concepts to sustain functional adequacy despite pressures or disturbances. The chapter also addresses the normative question of desirability, arguing that resilience is sometimes, but not always, desirable. By articulating conceptual resilience as a tool for preventative engineering, the chapter broadens the scope of conceptual engineering interventions from mere cure to

prevention, arguing that navigating conceptual disruption can (and sometimes, should) involve anticipating and preventing it, rather than merely remedying it *ex post*.

Chapter 6 (*Navigating conceptual disruption with conceptual engineering*) addresses the main research question by synthesising the contributions of the preceding chapters. Such insights are then operationalised into an iterative process illustrating how to navigate technology-induced conceptual disruption with conceptual engineering. The seven stages of the cycle are: (1) suspicion of disruption; (2) context and stakeholder specification; (3) functional and affordance mapping; (4) adequacy and disruption assessment; (5) conceptual engineering intervention; (6) conceptual implementation; and (7) evaluation and monitoring.

Chapter 7 (*Conclusion*) summarises the key findings of the investigation, highlights the limitations and unresolved challenges of this project, and identifies promising directions for future research building on the insights developed in the dissertation.

Ultimately, this investigation contributes to the advancement of both conceptual engineering—illustrating how insights and tools developed for the study of socio-technical artefacts can enrich the theoretical foundations and operationalisation of conceptual engineering—and the philosophy of technology—showing how the vocabulary of conceptual engineering can equip philosophers of technology with sharper tools for identifying, evaluating, and responding to conceptual disruptions. Moreover, the dissertation illustrates that bringing the philosophy of technology and conceptual engineering into sustained dialogue not only provides the necessary insights to adequately navigate technology-induced conceptual disruption with conceptual engineering, but also generates significant mutual benefits for both disciplines.

Samenvatting*

Hoewel er uitgebreid onderzoek is gedaan naar de manier waarop technologie onze praktijken, normen, waarden en overtuigingen beïnvloedt, is er veel minder aandacht besteed aan de vraag hoe technologie onze concepten beïnvloedt—de begrippen waarmee we onze wereld begrijpen, erover nadenken en ons daarin bewegen. Zo heeft de opkomst van autonome wapensystemen het concept CONTROLE in het internationaal humanitair recht onder druk gezet. Dit concept bleek ontoereikend om duidelijke verantwoordelijkheidslijnen vast te stellen voor het gebruik van geweld in situaties waarin de controle over wapens niet langer rechtstreeks door een mens werd uitgeoefend, maar verdeeld was over menselijke en niet-menselijke entiteiten. Ook de opkomst van autonome voertuigen heeft dringende vragen opgeroepen over onder meer aansprakelijkheid, veiligheidsnormen en de ongelijke impact van deze technologie op verschillende sociale groepen. Veel van zulke vragen hangen samen met de concepten die eraan ten grondslag liggen.

Dit proefschrift onderzoekt hoe technologie-geïnduceerde conceptuele ontwrichting—de onderbreking van de functionele toereikendheid van concepten—kan worden aangepakt met conceptual engineering: de evaluatie, het ontwerp en de implementatie van concepten. Het proefschrift laat zien dat conceptual engineering en de filosofie van technologie beter in staat zijn hun onderzoek te verrichten wanneer ze met elkaar in gesprek gaan dan wanneer ze apart opereren. Dat is opvallend, want het zijn twee disciplines die tot nu toe grotendeels parallel zijn verlopen, ondanks verwante probleemstellingen, vergelijkbare onderzoeksobjecten en gedeelde normatieve doelen. De kruising van conceptual engineering en filosofie van technologie is een vruchtbaar maar nauwelijks verkend terrein dat dit proefschrift systematisch betreedt, en dat beide disciplines vooruithelpt op manieren die geen van beide afzonderlijk heeft bereikt. In het proefschrift verloopt deze uitwisseling via twee bewegingen.

Aan de ene kant is er binnen de filosofie van technologie een groeiende interesse in de impact van technologie op de toereikendheid van beschikbare concepten. Steeds vaker klinkt de observatie dat technologieën concepten ontoereikend kunnen maken—wat conceptuele ontwrichting wordt genoemd—en dat die concepten dan wellicht herzien moeten worden. Maar wat het precies betekent dat een concept herzien moet worden, op welke gronden, met welk doel en volgens welke

normen, blijft in de filosofie van technologie onderbelicht. Hier komt een eerste toenadering van pas: van de filosofie van technologie naar conceptual engineering; het normatieve werk gericht op de verbetering van concepten. Conceptual engineering heeft namelijk een vocabulaire en een set maatstaven ontwikkeld voor het evalueren, (her)ontwerpen en implementeren van concepten—inzichten die in de filosofie van technologie nog nauwelijks zijn doorgedrongen. Deze eerste toenadering brengt conceptual engineering in gesprek met de filosofie van technologie en biedt een verfijnder lexicon en scherpere succescriteria voor een praktijk die velen in de filosofie van technologie al impliciet uitoefenen.

Aan de andere kant zijn vragen over wanneer en hoe concepten herzien moeten worden—zoals discussies over toereikendheidsnormen, de functies die concepten zouden moeten vervullen, en hoe concurrerende voorstellen voor (her)ontwerpen tegen elkaar afgewogen kunnen worden — zelf onderwerp van debat binnen conceptual engineering. Hier blijkt de filosofie van technologie onverwacht veel te bieden. In de literatuur over conceptual engineering worden concepten regelmatig beschreven als analoog aan artefacten, maar een serieuze verkenning van de filosofie-van-technologie-literatuur over sociotechnische artefacten blijft zeldzaam. Dat vraagt om een tweede toenadering: van conceptual engineering naar de filosofie van technologie. Dit onderzoek werkt de analogie tussen concepten en sociotechnische artefacten verder uit en betoogt dat de uitgebreide literatuur over artefacten inzichten kan bieden om conceptual engineering verder te onderbouwen en te verrijken. De belangrijkste innovaties die dit proefschrift op dit gebied introduceert zijn CONCEPTUELE AFFORDANCES, om de evaluatie van conceptuele toereikendheid verder te verfijnen, en CONCEPTUELE VEERKRACHT, om te beoordelen hoe wenselijk het is om concepten zo te ontwerpen dat ze bestand zijn tegen conceptuele ontwrichting. Daarnaast pleit het proefschrift voor meer gestructureerde benaderingen van conceptual engineering in de praktijk.

Hoofdstukken 2 en 6 bewegen zich primair van de filosofie van technologie naar conceptual engineering, en laten zien hoe het vocabulaire en de inzichten van conceptual engineering filosofen van technologie kunnen uitrusten met scherpere instrumenten voor het conceptualiseren, identificeren, evalueren en beantwoorden van conceptuele ontwrichtingen (hoofdstuk 2) op een systematische manier (hoofdstuk 6). Hoofdstukken 3, 4 en 5 gaan de andere kant op: ze laten zien hoe inzichten en instrumenten uit de studie van sociotechnische

artefacten de theoretische grondslagen en operationalisering van conceptual engineering kunnen verrijken. Ze doen dit door de analogie tussen concepten en sociotechnische artefacten te gebruiken om te betogen dat ontworpen concepten inherent waardegebonden zijn en dat conceptual engineering baat zou hebben bij meer methodologische systematiek (hoofdstuk 3), en door CONCEPTUELE AFFORDANCES (hoofdstuk 4) en CONCEPTUELE VEERKRACHT (hoofdstuk 5) te introduceren in de literatuur over conceptual engineering. Het onderzoek plaatst conceptual engineering dus niet alleen binnen de filosofie van technologie, maar bekritiseert en vernieuwt haar ook.

Het proefschrift is opgebouwd uit zeven hoofdstukken.

Hoofdstuk 1 (*Inleiding*) biedt een overzicht van het filosofie-van-technologie-landschap waarbinnen dit onderzoek primair is gesitueerd, gevolgd door een beknopte inleiding op conceptual engineering en concepten, en een overzicht van de kernthema's en leidende vragen die in dit proefschrift aan bod komen.

Hoofdstuk 2 (*Wat is conceptuele ontwricting?*), mede geschreven met Kevin Scharp, legt het fundament van het proefschrift. Het hoofdstuk definieert conceptuele ontwricting als een onderbreking van de functionele toereikendheid van concepten: concepten zijn verstoord wanneer ze niet langer in staat zijn hun functies adequaat te vervullen. Het biedt ook een taxonomie van het fenomeen langs drie dimensies: (i) het niveau waarop de ontwricting zich voordoet (individueel concept, conceptueel cluster of conceptueel schema); (ii) het type ontwricting (conceptuele lacune en conceptueel conflict, waarbij dit laatste conceptuele overlap en conceptuele misalignment omvat); en (iii) de ernst van de ontwricting (van mild tot ernstig). Naast een diagnostisch kader voor het vaststellen en beoordelen van de urgentie van conceptuele ontwrictingen, laat het hoofdstuk zien dat verschillende typen ontwricting om verschillende interventies vragen—*de novo* engineering voor conceptuele lacunes en herontwerpen voor conceptuele conflicten—en positioneert het conceptual engineering als het centrale instrument voor het systematisch identificeren, beoordelen en aanpakken van conceptuele ontwrictingen.

Hoofdstuk 3 (*Concepten als artefacten*) is eveneens fundamenteel voor de bredere bijdrage van het proefschrift. Het neemt de analogie tussen concepten en artefacten serieus—een analogie die in conceptual engineering bekend is maar zelden is uitgewerkt in dialoog met de filosofie-van-technologie-literatuur over sociotechnische artefacten—en onderzoekt wat die analogie oplevert wanneer ze rigoureuus wordt

doorgevoerd. Na een verduidelijking van hoe de analyse van concepten als verwant aan sociotechnische artefacten zich onderscheidt van bestaande opvattingen over concepten als abstracte culturele artefacten, worden drie inzichten voor conceptual engineering ontwikkeld: (i) concepten dragen, net als artefacten, affordances en brengen neveneffecten voort, waardoor gebruik dat verder gaat dan het beoogde doel mogelijk wordt; (ii) herontworpen concepten zijn inherent waardegebonden en weerspiegelen de prioriteiten en aannames van hun ontwerpers en de sociotechnische contexten waarin ze opereren; en (iii) conceptual engineering zou baat hebben bij meer methodologische systematiek, naar het voorbeeld van de gestructureerde benaderingen die de filosofie van technologie heeft ontwikkeld voor de evaluatie, het ontwerp en de implementatie van artefacten. Samen verbreden en verrijken deze inzichten het instrumentarium van conceptual engineering en stellen ze de agenda voor de volgende hoofdstukken.

Hoofdstuk 4 (*Conceptuele affordances*) betoogt dat het functionalisme in conceptual engineering gegrond moet worden in een theorie van affordances. Conceptual engineering hanteert overwegend een functionele benadering van conceptuele toereikendheid: concepten hebben functies en zijn toereikend voor zover ze die functies vervullen. Het functionalisme gaat daarbij traditioneel uit van de aanname dat conceptuele engineers zich moeten richten op wat concepten zouden moeten doen. Dat is te beperkt, zo betoogt het hoofdstuk, omdat het systematisch verwaarloost wat concepten mogelijk maken of beperken buiten de doeleinden die ze geacht worden te dienen. Door te putten uit de notie van affordance zoals ontwikkeld in de ecologische psychologie en uitgewerkt in de filosofie van technologie, introduceert het hoofdstuk de notie CONCEPTUELE AFFORDANCES: de potentiële handelingen, doeleinden en gedachten die een concept mogelijk maakt of beperkt voor een gegeven gebruiker en binnen een gegeven context. Dit biedt conceptuele engineers een scherper, meer holistisch en ethisch bewust kader voor de evaluatie, het ontwerp en de implementatie van concepten.

Hoofdstuk 5 (*Conceptuele veerkracht*) verbreedt de reikwijdte van het onderzoek door de vraag te stellen of conceptual engineering niet alleen als reactie op ontwrichting moet fungeren, maar ook als middel om die te voorkomen. De bestaande literatuur heeft zich vooral gericht op de vraag hoe conceptual engineering ontwrichtingen kan beperken of verhelpen nadat ze zich hebben voorgedaan. Aan de vraag hoe concepten zo ontworpen kunnen worden dat ze weerstand bieden aan ontwrichting is weinig aandacht besteed. Het hoofdstuk introduceert de notie van

conceptuele veerkracht: het vermogen van concepten om functioneel toereikend te blijven ondanks druk of ontwrichtingen. Het gaat ook in op de normatieve vraag naar wenselijkheid en betoogt dat veerkracht soms, maar niet altijd, nastrevenswaardig is. Door conceptuele veerkracht te formuleren als instrument voor preventieve engineering, verbreedt het hoofdstuk de reikwijdte van conceptual engineering van herstel naar preventie: het navigeren van conceptuele ontwrichting kan en moet soms het anticiperen op en voorkomen ervan omvatten, in plaats van haar alleen achteraf te verhelpen.

Hoofdstuk 6 (*Het navigeren van conceptuele ontwrichting met conceptual engineering*) beantwoordt de centrale onderzoeksvraag door de bijdragen van de voorgaande hoofdstukken samen te brengen. Deze inzichten worden vervolgens omgezet in een iteratief proces dat laat zien hoe technologie-geïnduceerde conceptuele ontwrichting met conceptual engineering kan worden aangepakt. De zeven fasen van de cyclus zijn: (1) het vermoeden van ontwrichting; (2) context- en stakeholderspecificatie; (3) functionele en affordancekartering; (4) toereikendheids- en ontwrichtingsbeoordeling; (5) conceptual-engineeringinterventie; (6) conceptuele implementatie; en (7) evaluatie en monitoring.

Hoofdstuk 7 (*Conclusie*) vat de belangrijkste bevindingen samen, bespreekt de beperkingen en openstaande vragen van het onderzoek, en wijst op veelbelovende richtingen voor toekomstig onderzoek dat voortbouwt op de inzichten uit dit proefschrift.

Dit onderzoek draagt bij aan de vooruitgang van conceptual engineering en filosofie van technologie. Voor conceptual engineering laat het zien hoe inzichten en instrumenten uit de studie van sociotechnische artefacten haar theoretische grondslagen en operationalisering kunnen verrijken. Voor de filosofie van technologie toont het hoe het vocabulaire van conceptual engineering filosofen kan uitrusten met scherpere instrumenten voor het identificeren, evalueren en beantwoorden van conceptuele ontwrichtingen. Bovendien laat het proefschrift zien dat de dialoog tussen beide disciplines niet alleen de inzichten oplevert die nodig zijn om technologie-geïnduceerde conceptuele ontwrichting goed aan te pakken, maar ook dat het voor beide disciplines aanzienlijke wederzijdse voordelen heeft.

* Summary written by Robin Hillenbrink.

List of Figures

Figure 5.1. Facets of conceptual resilience

Figure 6.1. Navigating conceptual disruption with conceptual engineering. Cycle

List of Tables

Table 2.1. Taxonomy of conceptual disruption

Table 2.2a. Old conceptual scheme

Table 2.2b. Conceptual conflict after ventilator technology

Table 2.2c. New conceptual scheme with conceptual gap

Table 2.2d. New conceptual scheme without conceptual gap

Table 2.3a. Old conceptual scheme

Table 2.3b. Conceptual conflict after Kuiper Belt discoveries

Table 2.3c. New conceptual scheme with conceptual gap

Table 2.3d. New conceptual scheme without conceptual gap

Table 6.1. Conceptual functions and affordances

Table 6.2. Conceptual disruption types

Table 6.3. Conceptual disruption. Levels, types, severity

Table 6.4. Conceptual engineering interventions

List and status of the papers used in the dissertation

Chapter 2

This chapter has been published as:

Marchiori, S., & Scharp, K. (2024). What is conceptual disruption?. *Ethics and Information Technology*, 26(1), 18.

Author contribution: SM took the lead on the taxonomy. KS took the lead on the constitutive principles. Both authors contributed to the writing and editing, and approved the final draft of the manuscript.

Chapter 3

This chapter has been published as:

Marchiori, S. (2026). Concepts as artefacts. Three insights that conceptual engineering can draw from the philosophy of technology. *Inquiry*, 1–25.

Chapter 4

This chapter has been published as:

Marchiori, S. (2025). Conceptual affordances: (How) should they inform conceptual engineering?. *Synthese*, 206(2), 88.

Chapter 5

Sections §5.2-§5.4.3 of this chapter have been published as:

Marchiori, S., & Sta. Maria, J. (2026). Conceptualising conceptual resilience. A comparative approach. *Philosophy & Technology*, 39(45).

Author contribution: SM took the lead and was responsible for the conceptualisation. SM wrote sections §5.2, §5.3, §5.4, and §5.4.3. Both authors jointly wrote sections §5.4.1 and §5.4.2.

*Melhor então seria não fazer nada, disse um dos filósofos otimistas,
os problemas do futuro, o futuro que os resolva,
O pior é que o futuro é já hoje, disse um dos pessimistas.*

*Better then to do nothing, said one of the optimistic philosophers,
the problems of the future, let the future solve them,
The worst thing is that the future is already today, said one of the pessimists.*

José Saramago, As intermitências da morte

Chapter 1

INTRODUCTION

1.1 Setting the stage — 1.1.1 Ethics of Socially Disruptive Technologies (ESDiT) — 1.1.2 Significance beyond ESDiT — 1.1.3 A note to the Reader — 1.2 Philosophy of conceptually disruptive technologies — 1.3 From philosophy of technology to conceptual engineering and back — 1.3.1 What concept of CONCEPT for conceptual engineering in the philosophy of technology? — 1.4 Themes and guiding questions — 1.5 Summary of chapters

1.1 Setting the stage

1.1.1 Ethics of Socially Disruptive Technologies (ESDiT)

This dissertation investigates how to navigate technology-induced conceptual disruption (i.e., the interruption in the functional adequacy of concepts) with conceptual engineering (i.e., the evaluation, design, and implementation of concepts). This project is situated in the Ethics of Socially Disruptive Technologies (ESDiT) program, whose focus lies on a specific set of technologies, termed *socially disruptive technologies*. Socially disruptive technologies are technologies that have “deep, important, ethically salient and wide-ranging impacts, that occur rapidly, provoke uncertainty and cannot be easily reversed” (Hopster, 2021, p. 7).

Importantly, this is not intended to point to a binary division between technologies that are socially disruptive and those that are not, as a given technology may qualify as socially disruptive even if it does not display all of these features. Rather, the presence and intensity of these dimensions of impact are indicators that the technology may be more or less socially disruptive. Moreover, the notion of socially disruptive technologies is itself the object of an ongoing process of (re-)conceptualisation and refinement (e.g., Hopster, 2021; Anderson et al, 2026). Accordingly, despite the increasing adoption of socially disruptive technologies as a discrete category of philosophical interest (e.g., Hermann, 2022; Hopster, 2026; Lane, 2025; Löhr, 2023b; Nyholm, 2026; van de Poel, 2022; Veluwenkamp et al, 2024), significant space for disagreement persists in terms of what should qualify as a socially disruptive technology and the extent to which this is a helpful category to adopt in the first place.

1.1.2 Significance beyond ESDiT

This investigation often draws on conceptual disruptions induced by socially disruptive technologies; nevertheless, its insights are not confined to the philosophy of socially disruptive technologies, but generalise to conceptual disruption in the philosophy of technology more broadly. In this dissertation, the emphasis will be specifically on the legal and moral philosophy of technology. However, the significance of this project extends further still, and in several directions worth making explicit. To begin with, while the primary focus of this dissertation lies on technology-induced conceptual disruption, this is not the only

phenomenon capable of prompting or contributing to conceptual disruption. As will be discussed in chapter 2, a range of other phenomena, not exclusively technological in character, may serve as catalysts for, or contributors to, conceptual disruption. Accordingly, the framework developed here is not limited in its applicability to cases where technology (alone) is the source of disruption. Rather, it is plausible that the insights from this project may be exported to conceptual disruptions more broadly understood, whether technology-induced or otherwise.¹

The approach to conceptual engineering developed in this dissertation also carries broader implications for philosophers of technology, irrespective of whether they are working on conceptual engineering. Concepts are foundational to technology governance: the ways in which we articulate, contest, and refine the concepts we use in discussions about technology have direct and consequential bearing on how technologies are understood and governed. Conceptual engineering offers a crisp vocabulary for addressing questions that have already emerged with considerable urgency in the philosophy of technology literature, such as in the technology governance and ethics, questions such as the moral and legal personhood of AI systems, the conditions for meaningful human control over autonomous systems, and the determination of appropriate liability regimes for AI. Adequate technology governance requires robust conceptual grounding, and this project aims to contribute to providing that grounding. Its findings are therefore of relevance not only to philosophers of technology, but also to those engaged in technology ethics, regulation, and governance more broadly.

Finally, this dissertation speaks to conceptual engineers working outside of the philosophy of technology altogether. Conceptual engineering is being conducted across numerous domains (often without bearing this name) and the insights developed here are generalisable beyond their immediate context. Insofar as this project offers guidance on how to identify, navigate, and respond to conceptual disruption, it may prove of use to all those engaged in conceptual engineering projects, irrespective of their disciplinary focus.

¹ Chapters 2 and 6 illustrate examples of uses of conceptual engineering to bridge instances of conceptual disruption in contexts other than the philosophy of technology and in response to disruption not prompted by technological innovation.

1.1.3 A note to the Reader

This research sits at the intersection of two disciplines—conceptual engineering and the philosophy of technology—that have so far had little dialogue with one another, and draws together assumptions, methodologies, and notions from both literatures. As a result, many of the concepts discussed here may be unfamiliar to scholars arriving from one side or the other of this discussion: philosophers of technology may have little exposure to conceptual engineering, and conceptual engineers may be equally unacquainted with the concerns and practice of philosophy of technology.

Matters are further complicated by this project belonging to a nascent research niche. This means that part of the very concepts and phenomena that this dissertation sets out to address had virtually no presence in the literature before this investigation began. For example, the study of the phenomenon of *conceptual disruption*, on which this research is grounded, began to take shape alongside the project itself, being first introduced by Löhr (2022) shortly after the start of the project and co-evolving with the inquiry rather than preceding it. In this sense, the ladder is being built as it is climbed. However, to preserve some narrative linearity, this introduction will make assumptions that are only later problematised, such as referring to the phenomenon of concepts being rendered inadequate by socially disruptive technology as a *conceptual disruption* in these terms from the outset, despite both the terminology and the legitimacy of that framing being part of what the dissertation contributes to examining and establishing.

The next sections provide an overview of the philosophy of technology landscape in which this research is primarily situated, before turning to an introduction to conceptual engineering and concepts. I will then provide an overview of the core themes and guiding questions addressed in this dissertation and a map of how each chapter contributes to addressing the main research question.

1.2 Philosophy of conceptually disruptive technologies

Technology has an impact on society.^{2,3} Consider the design of children's books. Thick pages, large fonts, wide line spacing, and illustrations are just some of the design choices that enable children to turn pages independently as well as to practice finger-tracking and text comprehension (Alaca, 2017; Chatzipanteli & Gorozidis, 2024; Lambert, 2017). These design features are crucial for literacy development (Kümmerling-Meibauer & Meibauer, 2013, 2017a, 2017b; Menegazzi & Nunes, 2015; Reese & Riordan, 2017). Were early readers presented with books made of thin paper pages, dense blocks of small, unspaced text, and without illustrations, the practices that children's book allow would be inhibited or precluded entirely (Martens, 2012; Uluğ & Bayraktar, 2014). It is plausible to expect that the result would not simply be a harder reading experience, but a fundamental narrowing of who gets to become a reader.

A comparable dynamic is a play in the design of phones intended for elderly users, which often feature enlarged buttons, simplified interfaces, and high-contrast displays (Al-Razgan et al, 2012; Gonçalves et al, 2017; Iancu & Iancu, 2020; Kurniawan, 2007; Lin et al, 2009; Petrovčič et al, 2018; Yusof et al, 2014). These design choices support the independence of elderly users and enable their digital inclusion (Pattinson & Stedmon, 2006). Were those same phones to be designed with small icons, complex interfaces, and low contrast, they would likely exclude many elderly users, making them dependent on others or disconnected from essential services. Thus, different design choices

² While hardly any positions (if any) deny outright that technology can impact its context of design, development, and use, some traditions do resist stronger or non-deflationary readings of technology's "impact." Instrumentalist accounts, for example, treat technology as an essentially neutral means whose social consequences are wholly determined by human intentions and use. On this view, technology does not itself exert any socially salient impact *as technology*; rather, such impact is traced back to how society impacts itself *through technology*. This is what Mitcham characterises as the "standard view" in much of 20th-century thought (Mitcham, 1994).

³ Philosophers of technology often refer to technologies as *socio-technical systems* and to the societal context in which technology is created and used as the *socio-technical landscape* to emphasise that society and technology influence and co-shape each other to such an extent that one cannot be neatly disentangled from the other.

produce different forms of inclusion and exclusion (Brey, 2016; Holmes, 2018; Reed & Monk, 2006; Umbrello, 2024; Verbeek, 2008). In short, technology contributes to shaping the socio-technical landscape in which it is embedded (Brey, 2010, 2018; Franssen et al, 2016; Ihde, 1995; Rosenberger & Verbeek, 2015; Verbeek, 2012; 2022).

While a rich scholarship has examined how technology impacts our practices, norms, values, and beliefs (Baier & Rescher, 1969; Crootof & Ard, 2020; Danaher & Sætra, 2022, 2023; Hopster & Maas, 2024; Jones, 2018; Lane, 2025; Nickel, 2025; Nickel et al, 2022; Sohn, 1973; Swierstra et al, 2019; van de Poel & Kudina, 2022; Waelbers & Swierstra, 2014), only recently has explicit attention been devoted to investigating how technology impacts our concepts (Hopster & Löhr, 2023; Löhr, 2023a, 2023b; Veluwenkamp et al, 2022), i.e., the notions through which we understand, reason about, and navigate our world (Margolis & Laurence, 2023).

For example, the emergence of autonomous weapons systems has put pressure on the concept of CONTROL adopted in international humanitarian law, which proved inadequate in establishing clear lines of responsibility for uses of force in contexts in which control over weapons was no longer exercised directly by a human agent, but distributed among human and non-human entities (e.g., Perri, 2001; Walzer, 2004; Sparrow, 2007; Chengeta, 2016).⁴ Similarly, the advent of autonomous vehicles, such as self-driving cars, has raised pressing questions about (among others) liability regimes, safety standards, and the differential impact of this technology across different social groups (Bissell et al, 2020; Boeglin, 2015; Coca-Vila, 2018; Liu et al, 2019; Lohmann, 2016; Rovira et al, 2019; Rowe et al, 2024; Shariff et al, 2021; Schellenkens, 2018; Stilgoe, 2021; Vellinga, 2016). Many of such questions are tied to the concepts that underlie them. For example, difficulties in harmonising existing mechanisms of liability attribution with regard to autonomous vehicles are often discussed in the legal and moral philosophy of technology as

⁴ The technology-induced inadequacy of the concept of CONTROL ultimately prompted the introduction of the concept of MEANINGFUL HUMAN CONTROL, which gained widespread traction among institutions and NGOs, as well as in the scholarly debate (Heyns & Kiai, 2016; Ekelhof, 2019; Rosert, 2019; Amoroso & Tamburrini, 2020; Riebe & Schmid, 2020; Saidi, 2022). The concept did not remain confined to the military domain, and has been adopted as a central notion for responsible innovation vis-à-vis AI more broadly, e.g., extending its application to autonomous vehicles, medical decision support systems, and AI governance (Santoni de Sio & van den Hoven, 2018; Siebert et al., 2022; Valéro et al., 2023).

related to this technology's impact on our concept of RESPONSIBILITY (Boeglin, 2015; Coeckelbergh, 2016; Gless et al, 2016; Hage, 2017), which is often deemed inadequate to tackle liability attribution in the context of AI systems, leading to the technology raising so-called *responsibility gaps* (Goetze, 2022; Matthias, 2004; Nyholm, 2018a, 2018b, 2023; Rowe et al, 2024; Santoni de Sio & Mecacci, 2021).

These examples illustrate a recurring pattern: technologies may introduce or reshape phenomena, practices, norms, or capabilities in ways that may strain existing conceptual frameworks, prompting the need for their evaluation and, if deemed appropriate, intervention, e.g., through (re-)design, implementation, or abandonment. The pattern appears across diverse domains. Digital surveillance technologies challenge concepts of PRIVACY by enabling monitoring at scales and in contexts previously impossible (Bennett, 2011; Coll, 2014; Friedewald & Pohoryles, 2016; Fuchs, 2011; Schlag, 2012). Gene editing technologies like CRISPR impact concepts of disease, therapy, and enhancement by enabling interventions that blur boundaries between treatment and enhancement (Chen, 2025; Condit, 2022; Deola et al, 2018; Insanguine Mingarro, 2023; Nordberg, 2021; Schleidgen et al, 2020; Waltz et al, 2024).

Medical technologies similarly challenge concepts of death. Historically, death was determined by the irreversible cessation of cardiopulmonary activity, criteria that aligned with available medical knowledge and technology (De Georgia, 2014). Advances in resuscitation and critical care transformed previously irreversible conditions, on which the determination of death had been grounded, into reversible ones. As new technologies were introduced, such as defibrillators and mechanical ventilators, which enabled patients to avoid cardiopulmonary failure, and to breathe artificially despite complete loss of brain function, the existing concept DEATH could no longer adequately guide medical decisions such as deciding when to discontinue treatment, determining when organ donation becomes permissible, as well as providing families with closure. These challenges to the concept DEATH as traditionally understood demanded a response, which led to the introduction of the

concept BRAIN DEATH as the irreversible cessation of brain function (Gillon, 1990; Lamb, 2020).⁵

In each of the cases illustrated above, technology serves as a catalyst for conceptual disruption, e.g., by exposing inadequacies in existing concepts or revealing that concepts designed for earlier socio-technical contexts no longer serve contemporary needs. As already illustrated, individual projects have been conducted in different domains tackling conceptual inadequacies prompted or enabled by scientific and technological progress. What remains largely missing is the move from *ad hoc* projects focusing on restoring the adequacy of individual concepts to a more structured approach to this type of technological impact. While this pattern appears across diverse domains, the focus of this dissertation is on the (legal and moral) philosophy of technology, which is often explicitly context-dependent, operating through the precise articulation of the concepts that structure its analysis.

The mismatch between technology's impact on (the adequacy of) our concepts and the absence of sustained and systematic philosophical investigation on such an impact highlights a gap in the literature that warrants being filled. This is where this dissertation is situated. This dissertation sets out to contribute to filling such a gap, by exploring how to navigate technology's disruptive impact on our concepts. It does so by connecting two fields of inquiry—philosophy of technology, on the one hand, investigating the study of socio-technical systems and their impact on the socio-technical landscape, and conceptual engineering, on the other, focusing on the evaluation, (re-)design, and implementation of concepts—which have so far proceeded on largely parallel tracks, and whose mutual communication is scarce, despite the rich affinities between them. Indeed, this dissertation argues that philosophers of technology and conceptual engineers engage in similar activities and are motivated by similar normative goals. Both disciplines are concerned with (questions surrounding) the evaluation, design, development, and

⁵ This conceptual intervention was not a merely stipulative redefinition but a complex normative project requiring careful evaluation of what functions the concept should serve, what medical, legal, and ethical considerations should guide its application, and how to implement revised criteria across medical institutions and legal systems. The question of whether the concept of BRAIN DEATH satisfactorily addresses the issue raised by scientific and technological advances remains the topic of discussion (e.g., Bernat, 1998; 2002; 2017; Boswell & Madder, 2011; Guerit, 2004; Korean & Machado, 2004; Machado, 2000; McMahan, 2006; Nair-Collins, 2010; Sarbey, 2016; Wasmuth Jr., 1969).

deployment of their respective objects of inquiry. Both operate from the broadly agreed upon understanding that technology and concepts serve specific functions. Moreover, contemporary discussions in both domains increasingly share a commitment to context-sensitivity, acknowledging that (the evaluation of) adequacy standards of technology and concepts—such as which functions technology and concepts *should* serve and the extent to which they do so adequately—should be sensitive to the specific users, contexts, and purposes for which technology or concepts are used. Crucially, both are animated by normative concerns. Conceptual engineers aim to not only analyse concepts, but improve them (when warranted and desirable). Similarly, philosophers of technology often engage in ethical research surrounding the ethical assessment, design, development, and deployment of technology.⁶

1.3 From the philosophy of technology to conceptual engineering and back

In this dissertation, I investigate how to navigate technology-induced conceptual disruption using conceptual engineering. Philosophers of technology have long recognised that technologies shape the socio-technical landscapes in which they are designed, developed, deployed, and embedded, influencing practices, social structures, power relations, and normative frameworks (Winner, 1980; Latour, 1992; Verbeek, 2005). However, little explicit attention has been devoted to how technologies specifically impact our concepts. A newly recurring observation is that technologies can render our concepts inadequate (this is known as a *conceptual disruption*) and that, in such cases, those concepts may need to be reconsidered or changed. However, what it means, for example, for a concept to ought to be changed—e.g., on what grounds, towards what end, according to which normative standards—remains underspecified in the philosophy of technology literature.

This is where a move from the philosophy of technology to conceptual engineering becomes pertinent. Conceptual engineering is the practice of evaluating, (re-)designing, and implementing concepts

⁶ This is particularly the case when considering contemporary Western philosophy of technology, in which this investigation is situated, and the philosophy of socially disruptive technologies specifically.

(Chalmers, 2020).⁷ In contrast to conceptual analysis, which focuses on the descriptive aspects of conceptual work, conceptual engineering asks whether concepts are adequate and, if not, how they can be improved. While the introduction of the term *conceptual engineering* is recent (originally, Blackburn, 1999; popularised in Burgess et al, 2020; Cappelen, 2018; Chalmers, 2020; Scharp, 2013), its practice predates the label. Carnap's (1950) work on explication, Quine's (1960) discussions of meaning revision, and Haslanger's (2000) ameliorative analysis all represent earlier manifestations of what is now grouped under the heading of conceptual engineering. Past work under headings such as *normative conceptual work*, *ameliorative analysis*, or *conceptual ethics* has pursued essentially the same (or a largely overlapping) aim: to intentionally shape our conceptual repertoire to better serve epistemic, practical, and normative ends. Moreover, the conceptual engineering literature acknowledges that conceptual engineering projects can, and do, regularly take place beyond the philosophical domain, albeit oftentimes without bearing this name (e.g., Cappelen, 2018; Cappelen & Plunkett, 2020).

For example, when members of the International Astronomical Union reconceptualised PLANET and DWARF PLANET in 2006 (IAU, 2006), and fellow scientists later advocated for ulterior reconceptualisations of PLANET (e.g., Runyon et al., 2017; Metzger et al., 2022; Hopster et al., 2023), they were performing conceptual engineering. When MIT chemical engineers developed 2DPA-1, a material “stronger than steel and as light as plastic” (Trafton, 2022), and introduced a new concept (2DPA-1) as a result, they were performing conceptual engineering (Zeng et al., 2022). When psychologists reclassified Asperger’s under Autism Spectrum Disorder in 2013 (APA, 2013), and when fellow psychologists later discussed and criticised that reconceptualisation (e.g., Regier et al., 2013; Giles, 2014), they were performing conceptual engineering.

⁷ While I follow much of the literature in treating concepts as the target of evaluation in conceptual engineering, it should be noted that this is not a universally shared assumption (e.g., Haslanger, 2000; Burgess, Cappelen, & Plunkett, 2020; Koch, 2021; Isaac, 2023). Indeed, some scholars have cast doubt on the coherence and utility of understanding concepts as the targets of conceptual engineering (e.g., Cappelen, 2018; Deutsch, 2020). I do not attempt to resolve these debates here, where I adopt a pragmatic approach to the issue: I assume, for the sake of argument and illustration, that there are such things as concepts and that concepts are the target of conceptual engineering.

Conceptual engineering proceeds through three interconnected activities: (i) *conceptual evaluation*, i.e., determining whether a concept is adequate; (ii) *conceptual (re)design*, i.e., creating a new concept, or revised an existing one; and (iii) *conceptual implementation*, i.e., ensuring uptake and sustained use (Chalmers, 2020). First, conceptual evaluation involves assessing whether a concept is adequate for its intended purposes. Conceptual engineering overwhelmingly operates from a functional perspective (e.g., Belleri, 2021; Chalmers, 2020; Haslanger, 2000; Nado, 2021; Plunkett & Sandell, 2013; Prinzing, 2018; Queloz, 2019; Simion & Kelp, 2020; Sundell, 2020; Thomasson, 2020). That is, concepts are taken to fulfil specific functions, and conceptual adequacy is tracked based on the extent to which concepts fulfil the functions they should serve. This requires determining what functions a given concept should serve and

determining whether it successfully performs those functions.⁸ Evaluation may reveal that a concept functions adequately, that it suffers from minor deficiencies amenable to revision, or that it is so deeply flawed that abandonment or replacement is warranted. Second, conceptual (re)design involves either revising existing concepts (re-engineering) or creating new ones (de novo engineering). A revised concept maintains some continuity with its predecessor while its content or structure is altered to improve functional adequacy. A newly created concept, by contrast, is introduced to capture phenomena that existing conceptual resources fail to articulate adequately. Third, conceptual implementation involves ensuring that adequate concepts achieve uptake and sustained use in relevant communities and contexts.

⁸ It is widely held that concepts can serve a plurality of functions based on the context of use and that question of how to determine which function(s) a given concept should serve is itself a substantive methodological one. Specific types of conceptual functions mentioned in the literature are *proper* functions (whereby concepts are assessed by the functions they were selected for over their evolutionary or historical development; Millikan, 1989), *etiological* functions (whereby the function of a concept is determined by its causal history and the role it has played in past successful deployments; Prinzing, 2018; Simion & Kelp, 2020), *system* functions (whereby concepts are assessed by the role they play within a broader system, irrespective of their history; Haslanger, 2020; Thomasson, 2020), *classificatory* functions (whereby concepts allow us to sort entities into kinds; Löhr, 2022; Hopster & Löhr, 2023), *practical* functions (whereby concepts produce useful effects in concept users' lives; Queloz, 2019), *normative* functions (whereby concepts produce effects that give concept users reasons to retain and deploy them; Köhler & Veluwenkamp, 2024), *evaluative* functions (whereby concepts express the evaluative viewpoint guiding their application; Queloz, 2019), *animating* functions (whereby concepts are governed by a conscious aim or goal; Queloz, 2019), *inferential* functions (whereby concepts license certain reasoning patterns; Queloz, 2019). One promising approach to determine which functions a given concept should fulfil proceeds by way of *reverse engineering*: by examining the purposes for which a concept is currently used (e.g., across disciplines and communities) one can reconstruct which function(s) the concept is expected to perform (Queloz, 2021; Veluwenkamp & van den Hoven, 2023). Haslanger's (2000) ameliorative analysis offers a version of this approach, asking what work a concept *is* doing and what work we *want* it to do. Similarly, Cappelen and Plunkett (2020), Riggs (2021), and Jorem (2022), emphasise the need to situate concepts within their contexts of use to assess their adequacy. Importantly, the functions identified through reverse engineering may diverge from any explicitly stated purpose, and different communities may rely on the same concept to serve different or even conflicting functions.

In short, the philosophy of technology literature is increasingly interested in exploring questions related to the impact of technology on the adequacy of our concepts, and conceptual engineering has developed a vocabulary and a set of standards for evaluating, (re-)designing, and implementing concepts. However, such insights from conceptual engineering, while pertinent, are not yet present in the philosophy of technology. The first innovative move in this dissertation involves bringing conceptual engineering to bear on conceptual work in the philosophy of technology, as it offers a more refined lexicon and success criteria for a practice that, as illustrated in the cases of MEANINGFUL HUMAN CONTROL and RESPONSIBILITY, many in the philosophy of technology have been conducting implicitly. The intersection of conceptual engineering and the philosophy of technology thus represents a fertile but under-explored area of inquiry that this dissertation systematically opens up, advancing both the philosophy of technology and conceptual engineering in ways that neither discipline has thus far achieved while operating in isolation from the other.

At the same time, questions such as when and how concepts ought to be changed—e.g., discussions surrounding adequacy standards for concepts, (how to identify) which functions concepts should serve, and how competing (re-)design proposals are to be evaluated against each other—are themselves debated within conceptual engineering (Cappelen & Plunkett, 2020). Interestingly, the conceptual engineering scholarship routinely refers to concepts as entities that are analogous to artefacts (Eklund, 2014; Haslanger, 2000; Kocurek, 2022; Miracchi, 2021; Thomasson, 2021). Yet, engagement with the rich philosophy of technology literature on (socio-technical) artefacts remains rare (with notable exceptions, e.g., Löhr & Veluwenkamp, 2025; Veluwenkamp, 2025).

This suggests that a second move is in order: from conceptual engineering to the philosophy of technology. In this dissertation, I argue that the analogy between concepts and artefacts should be explored further, and that the extensive philosophy of technology scholarship on artefacts can offer insights and resources to further ground and enrich conceptual engineering. Ultimately, this dissertation argues that bringing the philosophy of technology and conceptual engineering into dialogue not only provides the necessary insights to adequately navigate technology-induced conceptual disruption with conceptual engineering, but also generates significant mutual benefits for both disciplines.

1.3.1 What concept of CONCEPT for conceptual engineering in the philosophy of technology?

Conceptual engineering aims to improve our concepts (Chalmers, 2020; Isaac, 2021). Yet, any attempt to improve concepts presupposes some understanding of what concepts are. Without a clear understanding of the nature of concepts, it becomes difficult to determine what it means to evaluate, create, refine, replace, or implement them. For example, does conceptual engineering entail changing a mental representation, altering the linguistic practices through which the concept is expressed, or reshaping the inferential roles a concept plays in reasoning (among others)? Much depends on what a concept is taken to be.

Concepts are often described as the building blocks of thought: they allow us to think, to categorise, to reason, to communicate (Margolis & Laurence, 2023). Recent work on conceptual engineering in the philosophy of technology argues that technology can render our concepts inadequate or expose preexisting inadequacies with our concepts; this phenomenon is referred to as a *conceptual disruption* (Hopster et al, 2023; Hopster & Löhr, 2023; Löhr, 2022, 2023a, 2023b; Marchiori & Scharp, 2024). If concepts allow us to think, inadequate concepts may impair our ability to do so, or may hinder our ability to do it *well*. The analogy with tools helps to illustrate this point. Consider painters and their art equipment. The equipment allows painters to paint. While adequate tools are not indicative of the quality of the result of the activity for which one uses them—such that a skilled painter can produce a masterpiece with subpar equipment, and an amateur painter can produce a mediocre art piece with first-rate equipment—it seems intuitively reasonable to assume that, all other things being equal, adequate tools tend to enhance, and inadequate tools to hinder, the quality of one’s work. Concepts are like tools.⁹ It seems reasonable to assume that possessing adequate concepts will allow us to think well, or enhance the quality of our thinking. Conversely, possessing inadequate concepts will hinder our ability to think, or to think *well*.

⁹ The analogy of concepts as artefact-like is overwhelmingly accepted in the conceptual engineering literature. This is grounded on a functional approach to concepts, according to which, similarly to artefacts, concepts have functions and are adequate to the extent that they fulfil their functions (Belleri, 2021; Chalmers, 2020; Haslanger, 2000; Plunkett & Sandell, 2013; Prinzing, 2018; Queloz, 2019; Simion & Kelp, 2020; Sundell, 2020; Thomasson, 2020).

The scholarship on concepts is wide and varied, offering competing accounts of which concept of CONCEPT should be used in conceptual engineering,¹⁰ none of which has thus emerged as dominant (e.g., Koch, 2021; Isaac, 2023).¹¹ Owing to the lack of a unified concept CONCEPT, some scholars (most notably, Machery, 2009), have adopted an eliminativist stance towards concepts, arguing that the concept CONCEPT does not refer to a natural kind and should be eliminated from discourse (Machery, 2005, 2009; Pöyhönen, 2013). Indeed, many theories of concepts assume, either implicitly or explicitly, that concepts are a natural kind, i.e., a unified category grounded in some underlying essence or shared structure (Margolis & Laurence, 2023). Machery's (2009) argument challenges this assumption, contending that grouping different concepts of CONCEPT under the heading "concept" obscures important differences among them and is misleading for scientific inquiry, as CONCEPT is too heterogeneous to serve any meaningful explanatory role. As such, the

¹⁰ This assumes that there are any such things as concepts, and that concepts should be the target of conceptual engineering in the first place. However, this assumption is not uncontroversial in conceptual engineering (e.g., Machery, 2009; Cappelen, 2018; Burgess, Cappelen, & Plunkett, 2020).

¹¹ On the one hand, traditional philosophical accounts (*classical* and *neoclassical* accounts of concepts) seek to define concepts in terms of necessary and sufficient conditions (Margolis & Laurence, 1999, 2023; Raatikainen, 2009). Such approaches assume that concepts have a definitional structure, composed of necessary and sufficient conditions, such that their content is stable and can be exhaustively captured in precise terms. On the other hand, empirical psychological research has revealed that people do not seem to use rigid definitions in practice. This led to an array of psychological theories of concepts. *Prototype theory* (Rosch, 1978 [2024]; see also Geeraerts, 1989, 2006; Hampton, 1995; Margolis, 1994; Osherson & Smith, 1981) suggests that concepts often display a prototypical structure and are organised around central, typical exemplars, with some members of a conceptual category being judged to be more typical than others. For example, a robin is often judged to be a more prototypical bird than a penguin. A related but distinct view, *exemplar theory* (Medin & Schaffer, 1978; see also Capasso, 2005; Gradoville, 2023), conceives of concepts as collections of specific instances stored in memory, with categorisation occurring by comparing new stimuli to one's stored exemplars. Still another influential view is *theory-theory* (Cohen & Murphy, 1984; Gopnik & Meltzoff, 1997; Tuomela, 2012), which holds that concepts function as mini-theories (that is, miniature explanatory frameworks) guiding e.g., inference and prediction. In short, theories of concepts differ widely and research in psychology and philosophy suggests that concepts are heterogeneous, varying across contexts, cognitive tasks, and theoretical frameworks.

concept CONCEPT should be abandoned in favour of more precise terminology, e.g., referring specifically to prototypes, exemplars, theories, or other types of representations.

Lalumera (2009; 2010) offers an alternative by arguing that concepts should be understood as a *functional* kind rather than a *natural* kind. Unlike natural kinds, which are unified by shared intrinsic properties, functional kinds are unified by the roles that its members play. This idea is well established in philosophy of biology, where organs such as hearts are categorised by their function (pumping blood) rather than their structural similarity across species. Lalumera applies this reasoning to concepts, arguing that despite their differences, concepts are unified by the roles they play, e.g., in enabling categorisation, inference, and communication. On this view, concepts do not need to share a single essence or underlying structure to be meaningfully grouped together; instead, they are connected by the roles they play.¹²

This dissertation adopts an understanding of concepts as functional kinds (Lalumera, 2009, 2010). That is, concepts are individuated by the roles they play, including, centrally, in higher cognitive processes, such as the inferential relations they license, and the downstream cognitive, behavioural, and communicative outputs they

¹² A useful way to think about Lalumera's position is through an adaptation of Wittgenstein's notion of family resemblance (Wittgenstein, 1922, 1956). Wittgenstein argued that concepts like GAME do not have single defining features and thus do not admit of definitions in terms of necessary and sufficient conditions, but are connected through overlapping similarities. Lalumera's functional kind approach builds on this idea but shifts the emphasis from mere resemblance to shared function: rather than being unified by surface similarities, concepts are grouped together because they play a common role. To extend the analogy, we can think of concepts in the way we think of tools: a hammer, a nail gun, and a heavy rock can all be used to drive a nail, even though they differ in structure, design, and material. What unites them is their function, not their composition. Concepts, under Lalumera's view, function similarly: different cognitive mechanisms (e.g., prototypes, exemplars, theories) can serve the same higher-level function of categorisation and inference, making CONCEPT a coherent notion even if it lacks an essential nature. In this sense, Lalumera's functional kind approach provides an appealing and pragmatic alternative to both essentialism and eliminativism. Understanding concepts as functional kinds treats concepts as flexible, functionally defined entities rather than as rigid natural kinds. This account accommodates heterogeneity without abandoning the concept CONCEPT, making it a promising foundation for philosophical inquiry in the philosophy of technology.

enable, as well as in behavioural, communicative and social practices. The choice to work with a broad notion of concepts as functional kinds is both consistent with current practice in conceptual engineering, where commitment to substantive notions of concepts is often left open (Chalmers, 2020; Haslanger, 2020), and a theoretically careful response to the challenge of overcommitment in contested philosophical terrain. Indeed, competing views on the metaphysics of concepts are each supported by divergent explanatory goals and background commitments. Committing to any of such positions would embroil the dissertation in disputes that are orthogonal to its primary aim, as this dissertation's central argument does not depend on any particular theory of concepts, and the notion of concepts as a functional kind can be operationalised across a range of competing views. That said, chapter 4 illustrates through examples how the functional kind approach is compatible with competing views on concepts. This is meant to show that understanding concepts as a functional kind cuts across different theories, and to support the claim that such a notion is both broadly applicable and theoretically fruitful for conceptual engineering.

Understanding concepts as a functional kind also aligns with pragmatist approaches to conceptual engineering (Gascoigne, 2024; Henne, 2024; Henne & Huetter-Almerigi, 2024; Löhr, 2024, 2025; Thomasson, 2020a, 2020b, 2025). According to such approaches, concepts should not be assessed primarily by their conformity to abstract theoretical criteria, but rather evaluated based on how they operate within specific practices and serve the purposes for which they are employed. From a pragmatist perspective, concepts being a heterogeneous category is not a defect to be eliminated; rather, it is a feature to be accommodated. This also provides a framework for integrating insights from different theories of concepts, seeing them as complementary perspectives on how concepts fulfil their functional roles.

In sum, the functional kind approach to concepts, while potentially appearing less substantive than essentialist alternatives, is better suited to the demands of conceptual engineering in the philosophy of technology. It accommodates the heterogeneity of conceptual theories, aligns with pragmatist and practice-oriented methodologies, and resonates with the empirical, situated, and relational character of contemporary philosophy of technology (Achterhuis, 2001; Brey, 2010; Franssen & Koller, 2016; Franssen et al, 2016; Rip, 2000; Verbeek, 2022). The apparent lack of substantiveness is not a deficiency but a recognition that concepts, like technologies, are best understood through the roles

they play in practice, rather than in a vacuum. By embracing a bottom-up approach, we gain the flexibility and context-sensitivity needed to engage effectively with the diverse and evolving conceptual challenges posed by an ever-evolving socio-technical landscape.

1.4 Themes and guiding questions

The preceding sections have introduced the central question that motivates this dissertation: how should we navigate technology-induced conceptual disruption with conceptual engineering? The primary contribution of this project is at once methodological and substantive. The dissertation innovates both the practice of conceptual engineering and the philosophy of technology, illustrating how exploring the connection between the two disciplines—which have so far proceeded on largely parallel tracks despite engaging with related problems, operating on similarly conceived objects of inquiry, and sharing broadly comparable normative aims—generates significant mutual benefits. This is illustrated through two moves.

Chapters 2 and 6, run primarily from the philosophy of technology to conceptual engineering, showing how the vocabulary and insights of conceptual engineering can equip philosophers of technology with sharper tools for conceptualising, identifying, evaluating, and responding to conceptual disruptions (chapter 2) in a systematic manner (chapter 6). Chapters 3, 4, and 5, run in the complementary direction: they show how insights and tools developed for the study of socio-technical artefacts can enrich the theoretical foundations and operationalisation of conceptual engineering. They do so by drawing on the analogy between concepts and socio-technical artefacts to argue that engineered concepts are value-laden and conceptual engineering would benefit from greater systematicity (chapter 3), and by introducing CONCEPTUAL AFFORDANCES (chapter 4) and CONCEPTUAL RESILIENCE (chapter 5) in the conceptual engineering literature. As such, the investigation does not merely embed conceptual engineering in the philosophy of technology, but it also scrutinises and innovates conceptual engineering itself.

The dissertation as a whole is a sustained illustration of how conceptual engineering and the philosophy of technology are better positioned to address their respective research in dialogue rather than in isolation. In particular, two interconnected themes give shape to this investigation: conceptual disruption and concepts as artefacts.

Central to this dissertation is the notion of conceptual disruption, understood as an interruption in the functional adequacy of concepts (Marchiori & Scharp, 2024). When concepts are disrupted, they fail to adequately perform the functions we wish them to fulfil, leading to challenges for (among others) reasoning, communication, classification, and governance. Conceptual disruption serves as a diagnostic category providing crucial groundwork for this investigation, identifying when conceptual inadequacy warrants attention and potentially demands intervention through conceptual engineering. When this investigation began, such groundwork was not yet available, as the notion of conceptual disruption was nascent and underspecified. Soon after the beginning of this project, Löhr (2022) introduced the term to capture classificatory uncertainty arising from technological change, but a systematic account of the phenomenon (such as its types, levels, severity, and relationship to conceptual engineering responses) remained undeveloped. This created a methodological challenge for addressing the dissertation's central question. Indeed, identifying what conceptual disruption entails, e.g., how it occurs, which form it takes, and which responses it warrants, is a prerequisite for any systematic approach to navigating it. Consequently, part of this dissertation's contribution involves developing a more robust account of conceptual disruption itself, work that has been conducted in dialogue with parallel developments in the nascent literature niche at the intersection of conceptual engineering and the philosophy of technology (Hopster & Löhr, 2023; Hopster et al., 2023; Löhr, 2023a, 2023b). The discussion surrounding conceptual disruption raises questions such as: how is conceptual disruption best conceptualised? Does disruption manifest in different ways? To what extent is conceptual disruption desirable or undesirable? Is (re-)engineering concepts to resist disruption a worthwhile aim? Which are the considerations that should guide philosophers of technology when navigating conceptual disruption? These questions are addressed in chapters 2, 5, and 6.

The dissertation also builds heavily on the analogy between concepts and (socio-technical) artefacts. While the analogy of concepts as artefacts is a recurring theme in conceptual engineering, engagement with the extensive philosophy of technology scholarship on (socio-technical) artefacts has remained almost entirely missing from the conceptual engineering literature (with the notable exception of Veluwenkamp & van den Hoven, 2023, and Köhler & Veluwenkamp, 2024). This dissertation takes this analogy seriously and explores it

further, arguing that closer engagement with the philosophy of technology literature on socio-technical artefacts can enrich the conceptual engineering literature in ways that have not yet been systematically explored. Specifically, this analogy raises questions such as: to what extent can concepts be meaningfully compared to socio-technical artefacts? Can (and should) insights developed in the philosophy of technology for the study of artefacts be brought to bear on the evaluation, (re-)design, and implementation of concepts? If so, how? These questions are addressed in chapters 3 and 4, and operationalised in chapter 6.

1.5 Summary of chapters

In this section, I provide a summary of each chapter, outlining its main argument and showing how it contributes to answering the dissertation's guiding question, i.e., how to navigate technology-induced conceptual disruption with conceptual engineering. The chapters are organised so as to move from theoretical clarification to practical application, with each chapter providing the groundwork on which the subsequent ones build.

The investigation of the research questions begins in Chapter 2 (*What is conceptual disruption?*), co-authored with Kevin Scharp.¹³ This dissertation focuses on navigating technology-induced conceptual disruption. This presupposes an account of what conceptual disruption amounts to. This chapter draws on the resources advanced in the conceptual engineering literature to offer such an account. Specifically, the chapter proposes to understand conceptual disruption as an interruption in the functional adequacy of concepts: that is, concepts are disrupted when are no longer able to adequately fulfil their functions. The chapter also offers a taxonomy of the phenomenon of conceptual disruption structured along three dimensions, distinguishing between: (i) the level at which the disruption occurs (individual concept, conceptual cluster, conceptual scheme); (ii) the type of disruption (conceptual gaps and conceptual conflicts, the latter including conceptual overlaps and conceptual misalignments); and (iii) the severity of the disruption

¹³ This chapter appeared as an article: Marchiori, S., & Scharp, K. (2024). What is conceptual disruption? *Ethics and Information Technology*, 26(1), 18. Author contribution: SM took the lead on the taxonomy. KS took the lead on the constitutive principles. Both authors contributed to the writing and editing, and approved the final version of the manuscript.

(extending from mild to severe). Beyond providing a diagnostic framework for identifying when a conceptual disruption is present and assessing how urgently it should be remedied (if at all), the chapter presents conceptual engineering as the appropriate approach for addressing conceptual disruptions. Moreover, it suggests that conceptual disruption may occur in sequence, with conceptual conflicts often leading to conceptual gaps, and that different types of disruption may reliably call for different conceptual engineering responses: conceptual gaps and conflicts should be addressed through *de novo* engineering and re-engineering, respectively. Ultimately, the chapter provides the foundations to determine what it means for technologies to disrupt concepts and situates conceptual engineering as a central tool to systematically identify, assess, and address conceptual disruptions.

Chapter 3 (*Concepts as artefacts*)¹⁴ is similarly foundational to the dissertation's broader contribution and introduces the second innovative move of this investigation: from conceptual engineering to the philosophy of technology. Indeed, while bringing conceptual engineering to bear on the philosophy of technology goes some way towards understanding how to address conceptual disruption, many questions remain open, such as which adequacy standards concepts should meet and how competing conceptual proposals are to be evaluated. Here, the philosophy of technology turns out to offer unexpected resources. In conceptual engineering, concepts are often compared to artefacts, yet engagement with the extensive philosophy of technology scholarship on socio-technical artefacts remains rare. This analogy serves as the linchpin of this chapter. After clarifying how the analysis of concepts as akin to socio-technical artefacts sets itself apart from existing scholarship on concepts as abstract cultural artefacts (Thomasson, 1999, 2007, 2021), the chapter argues that the analogy between concepts and socio-technical artefacts yields important and novel insights for conceptual engineering when pursued in dialogue with the extensive philosophy of technology scholarship on socio-technical artefacts. In particular, the chapter develops three insights that can enhance the practice of conceptual engineering. First, alongside functions, concepts, like artefacts, bear affordances and produce side effects, enabling uses beyond their intended purpose. Second, (re-)engineered concepts are inherently value-

¹⁴ This chapter appeared as an article: Marchiori, S. (2026). Concepts as artefacts. Three insights that conceptual engineering can draw from the philosophy of technology. *Inquiry*, 1–25.

laden, reflecting the priorities and biases of their designers and the socio-technical contexts in which they operate. Third, conceptual engineering would benefit from greater methodological systematicity, drawing on the structured approaches employed in the philosophy of technology for the evaluation, design, and implementation of concepts. Together, these three insights suggest that exploring a closer alignment with the philosophy of technology scholarship on socio-technical artefacts can both refine our theoretical understanding of concepts as artefact-like and offer practical directions for future conceptual engineering work, by broadening and enriching the toolkit of conceptual engineering and illuminating new and promising directions for research and practice. Two of such insights will be tapped into in chapter 4 (affordances) and 6 (operationalisation).

Chapter 4 (*Conceptual affordances*)¹⁵ develops the first insight advanced in Chapter 3, by introducing the notion of affordances in the conceptual domain. The starting point of the chapter is the understanding that much of conceptual engineering adopts a functional approach to conceptual adequacy. That is, conceptual engineering is a normative approach to conceptual work aimed at the improvement of concepts through evaluation, design, and implementation. To this end, conceptual engineers need to have a measure of what an adequate concept amounts to. Functionalism offers such a standard. Following the functional approach, concepts have functions and are adequate to the extent that they fulfil their functions. Functional approaches have traditionally operated on the assumption that conceptual engineers ought to concern themselves with what concepts *should* do. Recent contributions have advocated for a more fine-grained and context-sensitive understanding of conceptual functions, extending beyond proper functions to encompass normative and possible functions. Building on these developments, and drawing on the philosophy of technology literature on the affordances of socio-technical artefacts, this chapter introduces the notion of *conceptual affordances*, understood as the potential actions, uses, and thoughts that a concept enables or constrains relative to a given user and within a given context. It is argued that an affordances-informed approach can and should supplement and enrich functionalism. Indeed, by attending to conceptual affordances, conceptual engineers can better capture what concepts *enable* us to do,

¹⁵ This chapter appeared as an article: Marchiori, S. (2025). Conceptual affordances: (How) should they inform conceptual engineering?. *Synthese*, 206(2), 88.

thus offering a sharper, more holistic, and ethically attuned framework for conceptual evaluation, design, and implementation.

Chapter 5 (*Conceptual resilience*)¹⁶ expands the scope of conceptual engineering beyond remedial interventions, by exploring how concepts can be made resistant to disruption, and the extent to which they should be (re-)engineered to do so. While the literature on conceptual disruption has primarily examined how conceptual engineering can be used to mitigate or solve instances of conceptual disruptions after they have occurred, little attention has been given to how concepts can avoid disruption, and how conceptual engineers can (re-)engineer concepts for resilience. Building on Lundgren's (2024) suggestion to use conceptual engineering as a tool to prevent future instances of conceptual disruption, this chapter introduces the notion of *conceptual resilience*, understood as the capacity of concepts to sustain functional adequacy despite pressures or disturbances. Unlike accounts that reduce resilience to stability (like Lundgren, 2024), the chapter proposes two complementary understandings of resilience: Conceptual Resilience as Immutability (CRI), where resilience consists in preserving a stable core, and Conceptual Resilience as Adaptation (CRA), where resilience consists in flexibly accommodating an evolving socio-technical landscape. These are not mutually exclusive but represent frames through which to illuminate different dimensions of resilience. The chapter then addresses the normative question of desirability, arguing that resilience should not always be a goal for conceptual engineers. Specifically, resilience is more commonly desirable in thin concepts at the individual level, while at the level of the conceptual scheme resilience may be counterproductive. By articulating conceptual resilience as a tool for preventative engineering, the chapter broadens the scope of conceptual engineering interventions from mere cure to prevention, arguing that navigating conceptual disruption can (and sometimes, should) involve anticipating and preventing it, rather than merely remedying it *ex post*.

Chapter 6 (*Navigating conceptual disruption with conceptual engineering*) synthesises the contributions of the preceding chapters and

¹⁶ Sections §5.2-§5.4.3 of this chapter have appeared as an article: Marchiori, S, & Sta. Maria, J. (2026). Conceptualising conceptual resilience. A comparative approach. *Philosophy & Technology*, 39(1), 45. Author contribution: SM took the lead and was responsible for the conceptualisation. SM wrote sections §5.2, §5.3, §5.4, and §5.4.3. SM and JS jointly wrote sections §5.4.1 and §5.4.2. Both authors approved the final version of the manuscript.

situates them within the broader literature at the intersection of conceptual disruption and conceptual engineering in the philosophy of technology, illustrating how the dissertation's findings, taken together, enable to answer the main research question. The chapter advances a cyclical, iterative 7-stage process, intended to guide philosophers of technology from the suspicion of conceptual disruption through conceptual engineering intervention. The seven stages of the cycle are: (1) suspicion of disruption; (2) context and stakeholder specification; (3) functional and affordance mapping; (4) adequacy and disruption assessment; (5) conceptual engineering intervention; (6) conceptual implementation; and (7) evaluation and monitoring. By synthesising and operationalising theoretical insights into a structured process, this chapter provides a practical toolkit for navigating conceptual disruption in rapidly evolving socio-technical landscapes and makes conceptual engineering more accessible for philosophers of technology.

Chapter 7 (*Conclusion*) summarises the key findings of the investigation, highlights the limitations and unresolved challenges of this project, and identifies promising directions for future research building on the insights developed in the dissertation.

PART I:
FOUNDATIONS

*ALEXANDER HAMILTON: If you stand for nothing, Burr,
what'll you fall for?*

Lin Manuel Miranda, Hamilton

Chapter 2

WHAT IS CONCEPTUAL DISRUPTION?¹⁷

2.1 Introduction — 2.2 Background on conceptual disruption — 2.3 Conceptual disruptions and constitutive principles — 2.4 Conceptual disruption. A framework — 2.4.1 Levels of conceptual disruption — 2.4.2 Types of conceptual disruption — 2.4.2.1 Conceptual gaps — 2.4.2.2 Conceptual conflicts — 2.4.2.3 Conflicts, gaps, and conceptual change — 2.4.3 Severity of conceptual disruption — 2.4.3.1 Mild conceptual disruption — 2.4.3.2 Moderate conceptual disruption — 2.4.3.3 Severe conceptual disruption — 2.4.4 Comparison — 2.5 The role of conceptual engineering — 2.6 Conclusion

¹⁷ This chapter appeared as an article in *Ethics and Information Technology*. Marchiori, S., & Scharp, K. (2024). What is conceptual disruption? *Ethics and Information Technology*, 26(1), 18. SM took the lead on the taxonomy. KS took the lead on the constitutive principles. Both authors contributed to the writing and editing, and approved the final draft of the manuscript.

2.1 Introduction

Recent work on philosophy of technology emphasises not only the ways in which technology can disrupt our social structures, e.g., by making certain occupations obsolete, but also the ways in which it can disrupt our concepts (Hopster, 2021). The newly emerging literature on the matter considers *conceptual disruptions* as challenges or interruptions to the normal course of conceptual activity (Hopster, 2021; Löhr, 2022, 2023a, 2023b; Hopster & Löhr, 2023; Hopster et al., 2023). The same literature recognises three kinds of conceptual disruption: *gaps* (i.e., no concepts apply to some new phenomenon), *overlaps* (i.e., multiple incompatible concepts apply to some item), and *misalignments* (i.e., concept activity is not aligned with norms and values). Here we focus our attention on gaps and overlaps, while leaving the investigation of conceptual misalignments for future work.

We present a new tripartite taxonomy for conceptual disruption, which divides this phenomenon across types of disruption, levels of disruptions, and degrees of severity of the disruption. First, we introduce a new type of conceptual disruption, namely, a conceptual conflict. We suggest that the category of overlaps is better understood as conceptual conflicts, and we propose several benefits of making this change. Second, we emphasise the diversity in the category of conceptual disruptions by tracing out three distinct levels at which they occur: individual concepts, clusters of concepts, and conceptual scheme. Third, we distinguish among mild, moderate, and severe disruptions, based on the varying degrees of urgency for bridging instances of conceptual disruption.

We ground the taxonomy in a novel and substantive account of conceptual disruption and its types by appealing to the idea that concepts have constitutive principles that partly determine their identity. These accounts illuminate the nature of conceptual disruption in general and the specific kinds of conceptual disruptions, but also the complex relationships between conceptual gaps, conceptual conflicts, and the conceptual changes that occur when technologies lead to changes in our concepts. Although we focus on technology-induced instances of conceptual disruption, the phenomena we consider can occur in a wider class of domains beyond the technological one.

The paper is structured as follows. Section §2.2 introduces and engages with the literature on conceptual disruption and sets the stage for the taxonomy to follow. In section §2.3, we present both the view that concepts have constitutive principles and the idea that concepts have

certain functions. Such insights are applied in section §2.4, which proposes a tripartite framework for conceptual disruption. Here, we distinguish among conceptual disruptions occurring at three levels (individual concepts, clusters of concepts, conceptual schemes), taking on two forms (conceptual gaps, conceptual conflicts), and leading to three different levels of severity (mild, moderate, severe). Furthermore, we avail ourselves of two case studies to illustrate how conceptual gaps and conceptual conflicts can occur at different times in the very same process of conceptual disruption, and how they lead to different kinds of conceptual changes. Section §2.5 focuses on the relationship between conceptual disruption and conceptual engineering, which is the process of evaluating our concepts and improving our concepts and conceptual schemes. Section §2.6 contains our conclusions.

2.2 Background on conceptual disruption

The philosophical literature on conceptual disruption is currently an emerging research niche, as the phenomenon only recently gained attention from philosophers of technology, in particular scholars interested in the ethics of socially disruptive technologies such as artificial intelligence, blockchain, and gene editing (Hopster, 2021; Löhr & Michel, 2023; Löhr, 2022, 2023a, 2023b; Hopster & Löhr, 2023; Hopster et al., 2023). Indeed, conceptual disruptions occur because new technologies can change the way our concepts function or the features of our conceptual schemes.

For the purpose of this paper, we will adopt Machery's (2009) definition of concepts, whereby "*a concept of x is a body of [information]¹⁸ about x that is stored in long-term memory and that is used by default in the processes underlying most, if not all, higher cognitive competences when these processes result in judgments about x*" (Machery, 2009, p. 11).¹⁹ We do not distinguish between concepts and linguistic meanings for the purposes of this paper.

¹⁸ In the original formulation, "body of knowledge" (Machery, 2009, p. 11).

¹⁹ It should be noted that Carey's (2009) work on concepts aligns with this definition and she provides a vast amount of detail for each of these aspects and embeds them within a much larger and empirically-supported framework for thinking about the relationship between concepts and other kinds of mental phenomena.

In recent work representative of the nascent conceptual disruption literature as a whole, conceptual disruptions are defined as “a challenge to the meaning of concept that prompts a possible future revision of it” (Hopster et al., 2023, p. 143).²⁰ In addition, conceptual disruptions are taxonomised in the following way:

We argued that conceptual disruptions can be interpreted in three ways, and that technology typically plays a prominent role in each of them (Hopster and Löhr, 2023). First, we may be faced with a ‘conceptual gap’. That is, we lack the concepts needed to describe a novel technological artifact, or to normatively evaluate the new impacts and affordances to which it gives rise. Second, we may be faced with a ‘conceptual overlap’. That is, more than one of our existing concepts may be appropriate to describe and evaluate a novel technology, but there is uncertainty as to which concept is most suitable. Third, there may be cases of ‘conceptual misalignment’. In such cases, existing concepts do seem applicable to conceptualize a new technology and its impacts and affordances. However, this apparent good fit actually masks an underlying value misalignment: the concept and its use do not express the values that a community of concept-users, upon ethical reflection, would like it to express. (Hopster et al., 2023, p. 144)

In what follows, we focus on only the first two kinds of disruptions, gaps and overlaps; we leave misalignments for future work. However, the ways in which each of these three terms has been defined thus far do not fully capture some features that we deem significant.

First, conceptual disruptions are not best thought of as consisting in the challenging of the meaning of a concept. For example, when there is a conceptual gap, which is a kind of conceptual disruption, there is some phenomenon to which our existing concepts cannot properly be applied. It might be that the phenomenon in question does not qualify as belonging to the extension of any of our existing concepts. In this case, no concept has its meaning challenged. Rather, the conceptual scheme as a whole is deemed inadequate for the purposes at hand. Thus, the above definition of “conceptual disruption” does not need to adequately capture the phenomenon under investigation.

²⁰ It should be noted that, while we choose to focus on this contribution, this is not the first publication on the matter, as evidenced by the papers cited above.

Löhr's work on conceptual disruption emphasises the *conceptual uncertainty* associated with disruptions to our concepts. For example, Löhr writes:

The term *conceptual disruption* in this paper means: Any intentional or unintentional challenge or interruption of the ways in which the individual or group has intuitively classified individuals, properties, actions, situations, or events, leading to *classificatory uncertainty*, i.e., uncertainty about the application conditions of a word or concept," (Löhr, 2022, p. 838).

Admittedly, Löhr is not proposing a definition that holds in general, but merely making a stipulation to capture the phenomenon for the purpose of the contribution under consideration. Nevertheless, since this is one of the few papers on the subject, we are interested in understanding whether this may hold, were this to be treated as a more general proposal for a definition of conceptual disruption. However, it should be noted that the following criticism might not apply to Löhr.

We have two worries about this stipulation. First, it applies only to classification, which is a function performed by concepts expressed by *predicates* in language. However, conceptual disruption is more widespread. For example, the development of quantum mechanics and its counterintuitive consequences caused some to doubt the distribution law in classical logic, i.e., $A \text{ or } (B \text{ and } C) \text{ iff } (A \text{ or } B) \text{ and } (A \text{ or } C)$. This doubt about distribution led to the development of quantum logics that violate the distribution law (Putnam, 1965). This is clearly a conceptual disruption, but it pertains to logical connectives, not predicates used for classification.²¹

The second worry about the definition under consideration is that, although we agree that uncertainty about whether certain concepts properly apply to a case does often accompany conceptual disruptions, it does not seem to occur in every case. For example, the concept EXOPLANET and the term "exoplanet" go back to at least 1992 when astronomers hypothesised the existence of objects that are similar to planets but orbit other stars. Planets have been defined as orbiting the sun, so these new objects were clearly not planets and no other astronomical concept seemed to apply to them. Thus, a new concept and

²¹ An anonymous referee pointed out that logical connectives can be used for classification of sentences, which is true. However, they also have non-classificatory uses (e.g., uttering a conjunction), and that is all we need for the counterexample.

new term were invented to classify this phenomenon. In this example, it does not seem like there was any uncertainty about whether the objects in question were planets. Everyone agreed that they were not planets. Nor was there uncertainty about whether these objects would be moons or some other kind of familiar celestial object. Instead, it was clear that a new category was needed to describe these objects properly (Burke, 1992). Therefore, while instances of conceptual disruption may oftentimes involve a certain degree of classificatory uncertainty, such uncertainty is not a necessary feature of conceptual disruption as a whole.

Moreover, it seems clear that there are cases of conceptual uncertainty that are not conceptual disruptions. For example, in 2011, some physicists working on the Oscillation Project with Emulsion-tRacking Apparatus (OPERA) project measured the velocity of some neutrinos as being greater than the speed of light. This caused considerable uncertainty about whether the concept of superluminal motion applied to these neutrinos. It turned out that a fibre optic cable was not attached properly to a piece of equipment which resulted in the mistaken measurement (Cartlidge, 2012). Therefore, while instances of conceptual disruption may manifest as instances of classificatory uncertainty, conceptual uncertainty is neither necessary nor sufficient for conceptual disruption. Instead, we propose that an interruption in the normal functioning of a concept, cluster of concepts, or conceptual scheme, is a necessary condition for disruption.

Finally, we disagree with how one kind of conceptual disruption, conceptual overlaps, have been characterised. Hopster et al. (2024) offer the following definition: “A conceptual overlap emerges if there is more than one concept that fits to describe a new type of artefact, action or event. In turn, this may prompt us to make a decision as to which concept to apply” (p. 14).

One worry is that the term “overlap” does not fully capture the problem with this kind of conceptual disruption in a perspicuous way. Indeed, there is an assumption that conceptual overlaps involve the application of *incompatible* concepts (e.g., natural and artificial, or alive and dead) (Hopster & Löhr, 2023). However, referring to “overlaps” may lead to misunderstandings. The fact that two or more concepts overlap is not a problem by itself. For example, the concepts BROWN and DOG overlap on brown dogs in the sense that brown dogs are in the extension of the predicate “brown” and in the extension of the predicate “dog”. Of course, this situation does *not* involve a conceptual disruption. Thus, the

term “overlap” may not be the best choice to capture such a kind of conceptual disruption.

Moreover, as we will see, this kind of conceptual disruption can occur with a single concept, so the term “overlap” is apt to be misleading. We prefer the term “conflict” instead to capture the sense of incompatibility at play in these examples.²² In addition, the kind of disruption in question involves some sort of conflict among concepts or between concepts and facts in the world. Thus, our category of conceptual conflicts incorporates what theorists have called conceptual overlaps, but is somewhat more expansive. We expand on this point in the next section.

2.3 Conceptual disruptions and constitutive principles

In this section, we outline a theory of conceptual disruption in general and what distinguishes particular kinds of conceptual disruptions like conflicts and gaps. The key element in the theory we offer is that concepts are determined by constitutive principles which help guide the application of the concept and provide identity conditions for concepts.

A constitutive principle for a concept is part of the content of that concept—they are “built in” to concepts. Constitutive principles are descendants of analytic principles. However, constitutive principles are designed to be different from analytic principles in at least two ways. First, constitutive principles can be false. This aspect marks an appreciation of the fact that our concepts and the meanings of our words are not pristine. They are historically situated and potentially messy items that have worked well enough in past situations for our ancestors to have passed them down to us. Concepts with false constitutive principles might work well in most situations, like our concepts of absolute space and absolute time. We know these concepts incorporate false claims about the world (e.g., simultaneity is observer-independent), but they do not cause problems in everyday situations, so it is fine to use these concepts in those contexts. Second, constitutive principles are not mandatory beliefs. That is, one need not believe that all bachelors are unmarried in order for one’s word ‘bachelor’ to mean *bachelor*. Instead,

²² It should be noted that, within the growing body of literature on conceptual disruption, different authors may attribute different meanings to the same terms. Therefore, the focus of attention should be placed not on the terms themselves, but on the way in which such terms are conceptualised.

the relationship between the possessors of a concept and that concept's constitutive principles is more complex (Eklund, 2002).

The theory of constitutive principles outlined here is based on previous work by Scharp (2013). When one has a conversation and one's interlocutor denies a constitutive principle, that is an "interpretive red flag"—this event ought to give one pause about going on interpreting the interlocutor in the usual way. A philosophically substantive way of explaining this metaphor is to say:

X is a constitutive principle for concept C iff when an interlocutor denies X, one has a pro tanto reason to think that one does not mean the same thing as one's interlocutor by the word that expresses C.

Because the sentence "dogs are animals" is constitutive for the concept DOGS, if a person speaking to you sincerely asserted that dogs are not animals, that would give you a defeasible reason to believe that this person does not mean what you mean by DOG. The reason is defeasible, so it can be overcome. That is, one might discover that this person does mean what you mean by DOG but simply has very strange or unusual beliefs about dogs. Moreover, one who possesses concept *C* is *entitled* to the constitutive principles for *C* in the sense that one would be justified in believing them unless one has strong evidence to the contrary. So anyone who possesses the concept of dog is entitled to believe that dogs are animals, even though (presumably) some people who possess this concept do not have this belief.

One might reasonably wonder how a constitutive principle can be false if it implicitly defines the term in question. This aspect can be confusing, especially because of the dominance of analyticity in thinking about meaning and concepts. A constitutive principle is just any old proposition with the extra status that we who possess and use the concept in question are going to presuppose that the principle in question is true and we are going to expect others to presuppose this as well. However, what we presuppose to be true can turn out to be false. Moreover, it might seem like stipulative definitions could never be false, because they give the word in question whatever meaning makes them true. However, that cannot be right since some stipulative definitions are impossible to make true (e.g., !tonk"#like definitions). Instead, a stipulative definition gives the term in question whatever meaning makes the definition constitutive, which is about how one thinks and communicates.

With this understanding of constitutive principles, we can provide substantive accounts of the kinds of conceptual disruptions and how they relate to conceptual change and conceptual engineering in particular. A disruption is usually understood as an interruption in the normal course of a process or activity. As such, a *conceptual disruption* is an interruption in the normal functioning of a concept, cluster of concepts,²³ or conceptual scheme (in this sense, also Löhr, 2023a).

In this definition, we have appealed to the normal functioning of a concept or concepts. By this, we just mean the pattern of using the concept for certain purposes in certain situations. It is common in the philosophical literature on conceptual engineering (i.e., the philosophical study of how to assess and improve our conceptual resources) to appeal to the functions of concepts in order to explain why a new concept is a good replacement for an existing concept (for a survey, see Riggs, 2021). However, this appeal has generated considerable controversy. On the one hand, this appeal to conceptual functions is popular, but on the other hand there is also strong opposition to it (Cappelen, 2018). The debate has generated a host of new substantive accounts of what conceptual functions are and some philosophers have advocated minimalist views on the matter. For example, Jorem argues that advocates of conceptual engineering ought to say no more than that a function of a concept is simply something the concept is used for (Jorem, 2022). We do not take a stand on this issue here, but appealing to conceptual functions allows us to draw important distinctions between conceptual disruption and conceptual change.

Conceptual disruption can be identified as the precursor to a conceptual change, whereby the interruption brought about by the

²³ What we mean by conceptual cluster is a set of concepts that are interrelated. For example, ALIVE and DEAD are part of a cluster of concepts because they are part of a network of interlinked concepts. This is relevant as changes to the former may force changes in the latter due to the concepts' interconnectedness. Conversely, ALIVE and NEUTRINO are not part of the same cluster because they have very little to do with one another. Conceptual clusters can vary in size, but are always smaller than a conceptual scheme.

conceptual disruption may then be followed by conceptual change.²⁴ In this sense, while conceptual disruption itself is a kind of change, it is not a kind of *conceptual* change. A conceptual change is the addition or elimination of a concept or a change in one or more constitutive principles for one or more concepts. However, a conceptual disruption is a change in the normal functioning of a concept, cluster of concepts, or conceptual scheme. Although conceptual changes are often prompted by conceptual disruptions, the two phenomena are distinct (Löhr, 2023a).

2.4 Conceptual disruption: A framework

It should be noted that conceptual disruption is not a neatly ordered group of phenomena, rather a fairly heterogeneous set of phenomena that ought to be explained to some extent independently from each other.

It is important to emphasise that the phenomenon of conceptual disruption should better be understood as occurring at different levels of granularity. In some instances, conceptual disruption primarily affects one or more concepts. In others, it primarily affects the conceptual scheme as a whole. Glossing over this different way of framing instances of conceptual disruption can make it difficult to understand the phenomenon itself, and may risk attributing features to conceptual disruption in general that apply only to certain kinds of them.

Indeed, conceptual disruption is a multi-layered phenomenon occurring on one of three main levels (conceptual schemes, conceptual clusters, individual concepts), which manifests in a number of distinct forms (conceptual gaps and conceptual conflicts), and presents varying degrees of severity (extending from mild to severe) (Table 2.1).

²⁴ In this sense, conceptual disruption and conceptual change could be considered as two components of a (micro) conceptual revolution, as described in Thagard (1993). However, the analogy is not perfect, as conceptual disruption is not a rare occurrence, nor does it require the equivalent of a paradigm shift at the macro level. Rather, it occurs regularly and, as illustrated in section §2.5, it can be addressed and bridged within the confines of “normal science” i.e., conceptual engineering.

Level of disruption	Type of disruption	Severity of disruption
Level 1: Conceptual scheme	Conceptual gap	Mild disruption
Level 2: Conceptual cluster	Conceptual gap	Moderate disruption
	Conceptual conflict	Severe disruption
Level 3: Individual concept	Conceptual gap	Moderate disruption
	Conceptual conflict	Severe disruption

Table 2.1. Taxonomy of conceptual disruption

2.4.1 Levels of conceptual disruption

The first level of conceptual disruption affects the conceptual scheme. At this level, the conceptual disruption is a change in a property of a conceptual scheme, such that the conceptual scheme that was complete before the instance of conceptual disruption occurred is no longer exhaustive. That is to say, a conceptual scheme that previously applied to almost any experienced phenomenon and could adequately classify them, is no longer able to do so.²⁵

For example, consider physicists inventing a new material,²⁶ biologists discovering a new species, engineers developing a new technology, or astronomers discovering a new celestial object. In any of these scenarios, neither our concepts nor our conceptual scheme change in terms of something being added or subtracted. What does change is the property of the scheme as being comprehensive. This is a change in the function of the conceptual scheme, not a change in any constitutive principles for the concepts in question.

The second level of conceptual disruption affects conceptual clusters. In the second case, the instance of conceptual disruption affects multiple concepts at once, which may trigger additional disruptions in related concepts. An example is the disruption of the concepts LIFE and DEATH following the introduction of the mechanical ventilator (Baker, 2013; Nickel, 2020; Nickel et al., 2022).

²⁵ It should be noted that, as an anonymous referee pointed out, conceptual schemes are hardly ever comprehensive. However, non-exhaustive schemes may appear to be exhaustive until a conceptual disruption uncovers their lack of completeness.

²⁶ For instance, see 2DPA-1, the material created in 2022 by MIT engineers (Zeng et al., 2022).

The third level of conceptual disruption affects individual concepts. In this case, the instance of conceptual disruption affects one concept, which then oftentimes triggers a disruption in related concepts. An example is the disruption of the concept of moral agency following the introduction of artificial intelligence (e.g., Sullins, 2009; Brožek & Janik, 2019). However, the disruption is not limited to one concept, but leads to more concepts linked to the first one also being disrupted (e.g., moral responsibility, as in Babushkina, 2020; Santoni de Sio & Mecacci, 2021; Sebastián, 2021), somewhat like a domino effect. In certain instances, the latter set of concepts would not necessarily have been disrupted if the former concept had not been disrupted to begin with. This echoes the notion of “inferential risk” in Löhr & Michel (2023).

In the second and third level, we can further distinguish between first-order disruptions and second-order disruptions. Indeed, our concepts are so interconnected with one another that conceptual disruption affecting one or more concepts is likely to ramify and propagate through the conceptual scheme. As a result, instances of conceptual disruption can exert direct effects on one or more concepts, and indirect effects on nearby concepts. For example, by blurring the boundaries between the human and the non-human, social robots can be said to disrupt the concept PERSON, which can prompt the disruption of AGENCY, which in turn may prompt the disruption of RESPONSIBILITY, and so on (van de Poel et al, 2023). We adopt the expressions *first-order disruption* and *second-order disruption* to account for the difference between concepts being directly and indirectly affected by the conceptual disruption, respectively.

In addition to characterising conceptual disruption, we focus on two specific types: gaps (in which the disruption affects the feature of exhaustiveness of concepts and conceptual schemes) and conflicts (in which the disruption manifests itself as an incompatibility among different concepts or constitutive principles). Conceptual gaps can take place at the level of the conceptual scheme, at the level of the conceptual cluster, and at the level of the individual concept. Conceptual conflicts can occur at the level of individual concepts and conceptual clusters. Let us illustrate each type of conceptual disruption in more detail.

2.4.2 Types of conceptual disruption

2.4.2.1 *Conceptual gaps*

A *conceptual gap* occurs if and only if there is some phenomenon that ought to be classified at a certain level of precision but no relevant concepts properly apply to the phenomenon in question. The phrase “at a certain level of precision” is crucial in this definition since many phenomena fit into categories at multiple levels of description.

For example, in the case of exoplanets described above, scientists were not completely without ways of classifying exoplanets before this conceptual innovation. Everyone would agree that something like a planet that is orbiting a star other than the sun would be a physical object, and it would be an object roughly in hydrostatic equilibrium from its own gravity, and it would be an object in orbit, and so on. Hence, the concept of physical object and the others listed would properly apply to this phenomenon. Thus, it is not the case that we have *no way whatsoever* of classifying something that is a conceptual gap. Rather, it is that we do not have any way of classifying it *at the relevant level of precision*. Moreover, the “ought” in the definition of “conceptual gap” pertains to the relevant projects, goals, and aims of the people involved. In the case of exoplanets, there was a conceptual gap because scientists wanted a way to classify these celestial objects that would differentiate them from other kinds of celestial objects like asteroids, comets, and moons.

With regard to the first level of conceptual disruption, conceptual gaps highlight a feature of the conceptual scheme, such that the scheme was complete before the instance of disruption, and ceases to be so following the conceptual disruption. A conceptual scheme exhibits the feature of non-exhaustiveness whenever a new socio-technical phenomenon emerges, which none of the concepts within it are able to express. This type of disruption prompts the introduction of a new concept to fill such a gap and restore the exhaustiveness of the conceptual scheme. For example, the development of artificial intelligence technologies disrupted our conceptual scheme and prompted the subsequent introduction of the concept ARTIFICIAL INTELLIGENCE, which effectively appeased such a disruption.

As for the second level of conceptual disruption, non-exhaustiveness affects conceptual clusters in the sense that, while one or more existing concepts may allow to partially express a phenomenon, such concepts either do not always allow to do so to a satisfactory degree,

or can convey the phenomenon satisfactorily, but highlight the need for a more straightforward way to do so, e.g., through the introduction of a new concept.

An example of non-exhaustiveness as the need to satisfactorily convey a phenomenon is the case of PRIVACY. Despite PRIVACY being a somewhat recent concept (Warren & Brandeis, 1890), especially when compared to some of its neighbouring concepts (such as AUTONOMY, DIGNITY, FREEDOM), one could argue that a series of phenomena largely overlapping with our understanding of the phenomenon currently identified with PRIVACY occurred well before the introduction of the related concept (DeCew, 2018). Moreover, one could argue that privacy can be expressed to a considerable extent as a combination of different concepts. An example would be expressing PRIVACY as the intersection of “moral autonomy, human dignity, identity, personhood, liberty, anonymity and confidentiality” (van de Poel, 2020, p. 55). We argue that what prompted the introduction of privacy was an instance of conceptual disruption in terms of non-exhaustiveness. That is, despite the combination of other concepts being somewhat able to express the phenomenon that PRIVACY would later convey, the advent of photography and its uses in sensationalist journalism highlighted that something was still missing: a new concept that would convey all the nuance of that particular phenomenon that needed to be expressed (Warren & Brandeis, 1890). In this scenario, the instance of conceptual disruption manifested itself as uncertainty in relation to both the nature of the conceptual gap and how such a gap should be bridged.

An example of non-exhaustiveness understood as the need for more straightforward ways to convey a phenomenon is the case of INFORMATION PRIVACY.²⁷ One could argue that INFORMATION PRIVACY can be adequately expressed as a combination of different concepts, e.g., PRIVACY, INFORMATION, DATA, INFORMATION TECHNOLOGY. Therefore, what prompted the introduction of the concept INFORMATION PRIVACY was an instance of conceptual gap such that, while the phenomenon under consideration could have been conveyed by only relying on existing concepts, its salience justified the development of a new concept to facilitate its expression.

²⁷ For the purpose of this paper, following Chalmers (2020), we consider INFORMATION PRIVACY to be a separate (while deeply interconnected) concept instead of a conception of PRIVACY.

What makes instances of conceptual disruption as non-exhaustiveness on the conceptual scheme level different from those occurring on the conceptual cluster level is ultimately a matter of degree. In the former case, the gap is such that it is extremely hard, if not impossible, to meaningfully express the new concept using the existing conceptual repertoire. In the latter case, uncertainty emerges over the presence, extension, and severity of the conceptual gap, as well as over the application conditions of the existing concepts.

As for the third level of conceptual disruption, conceptual gaps affect individual concepts whenever they do not seem able to adequately convey the phenomenon that they should express. While this does not always occur as a result of technological advancements, technology can play a role in undermining existing concepts, e.g., by uncovering potential issues with relation to assumptions or constitutive principles related to such concepts. In this case, the non-exhaustiveness manifests itself as a defect in a concept that makes it inadequate to express certain nuances that it should be able to encompass.

2.4.2.2 *Conceptual conflicts*

A *conceptual conflict* occurs if and only if there is some known fact that is inconsistent with some constitutive principles for some concepts.²⁸ For example, in the late 1990s and early 2000s, scientists using new astronomical technologies were able to identify great numbers of objects in the Kuiper belt that were very similar to Pluto. These objects seemed to satisfy the definition of PLANET that was used at the time, which goes back to antiquity.

Nevertheless, astronomers were reluctant to call these objects planets for at least two reasons. First, planets are supposed to be special

²⁸ A clarification is in order. Conceptual conflicts can occur in three main ways: (a) as inconsistencies among constitutive principles; (b) as inconsistencies between constitutive principles and facts about the world; (c) as inconsistencies between constitutive principles and values or social norms. Here, we address the first two types, and leave the third for future work on conceptual misalignment. In this paper, we focus specifically on the second type of conflict, and characterise conflicts as inconsistencies between constitutive principles and facts. This is because it is extremely difficult to generate a conflict among constitutive principles alone, as there seems to always be some wiggle room. The liar paradox can exemplify this challenge: in order to get the contradiction, one first needs to have the empirical claim that the liar's sentence exists.

in that there are very few of them. Second, there seemed to be a difference in kind between Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune on the one and Pluto and the other Kuiper belt objects on the other hand. Lumping all these objects together under a single category obscured an important natural difference among them.

To remedy the situation, professional astronomers voted to redefine the word “planet” so that all planets have cleared the neighbourhood around their orbit of debris (IAU, 2006). This redefined concept of planet applied to all the familiar objects except Pluto. Hence, on the new definition, Pluto is not a planet. Moreover, the astronomers introduced a new concept and word—dwarf planet—that applies to Pluto and the other Kuiper belt objects in question. In this case, the conceptual conflict was an inconsistency between the constitutive principles for the concept of planet (i.e., there are few planets) and the empirical discovery that there are many objects orbiting the sun in roughly hydrostatic equilibrium from their own gravity (i.e., there are very many planets). There cannot be both only a few planets and not only a few planets. This is the heart of the conceptual conflict. The conflict in this case was resolved by adding a new constitutive principle in a redefinition of PLANET so that planets have cleared the debris from the neighbourhoods of their orbits.

Conceptual conflicts take place at the second and third level of conceptual disruption, affecting conceptual clusters and individual concepts. When they affect individual concepts, they manifest themselves as instances of incompatibility among a concept’s constitutive principles. An example of this is the disruption of the concept of TRUTH (Scharp, 2013). When they affect conceptual clusters, they manifest themselves as instances in which two or more concepts are only partially able to convey what we want to express, but none of them is able to do so to a satisfactory degree, or to a distinctly better degree compared to the other(s), which leads to a phenomenon being expressed through a combination of mutually incompatible concepts.

For example, in the case of a person whose cognitive functions are irrevocably lost, but who is able to breathe through mechanical ventilation, we might find it difficult or even impossible to determine whether either the concept ALIVE or DEAD can adequately convey what we want to express about the condition of the person. Not only does neither concept taken by itself seem good enough to express such a phenomenon, but the combination of the two concepts does not seem to meaningfully help solve the issue. That is to say, describing the condition

of the person in the example as them being alive, as well as them being dead, does not seem to fully convey the salient features of the phenomenon at hand. At the same time, describing the condition of the person as them being both alive and dead at the same time does not seem particularly helpful either, given that the state of being alive is incompatible with the state of being dead.

Whenever instances of conceptual disruption affect conceptual clusters, we are confronted with uncertainty as to how to determine the conditions according to which existing concepts can or should be deemed suitable to fill the gap of a seemingly missing concept, and apply such conditions to the case at hand. In such scenarios, it may be challenging to discern instances of conceptual gaps from instances of conceptual conflicts.

For example, in the case of the mechanical ventilator, one may argue that the conceptual disruption at hand is an instance of non-exhaustiveness, to the extent that neither the concept ALIVE nor the concept DEAD can adequately express what we want to convey about the status of a person who is able to breathe through a ventilator, but has no cognitive functions and cannot regain them. However, one may also argue that this is an instance of conceptual conflict, as both concepts are partially able to convey what we want to convey, and the difficulty lies not so much in determining *whether* one of those concepts or the combination of those concepts is adequate enough to describe the phenomenon at hand, but rather in us not being able to determine *which* of the two concepts, equally suitable for conveying what the seemingly missing concept should convey, should be preferred.

We argue that whether an instance of conceptual disruption should be described in terms of conceptual gap or conceptual conflict ultimately depends on the constitutive principles of the concepts involved. That is, understanding whether (the constitutive principles of) two (or more) concepts stand in an A vs B or A vs -A relation allows to qualify the conceptual disruption at hand as an instance of conceptual gap or conceptual conflict, respectively (in similar terms, Sommaggio & Marchiori, 2018).

Let us illustrate this with an example. In both the cases of PRIVACY and BRAIN DEATH, the instance of conceptual disruption emerged in such a way that a phenomenon could not be adequately addressed by any existing individual concept, and could only partially be addressed by a combination of existing concepts. However, the way in which such concepts relate to each other and come together to express the seemingly

missing concept (PRIVACY and BRAIN DEATH, respectively) varies significantly.

In the case of privacy, the individual concepts considered in the combination stand against each other in an A vs B type of relation. That is, neither the concepts as a whole nor their constitutive principles are incompatible with each other, thus identifying a conceptual gap. Conversely, in the case of the ventilator, the individual concepts considered to jointly express the phenomenon at hand stand against each other in an A vs -A relation. This is because some of their constitutive principles are mutually incompatible. Indeed, DEATH is usually expressed in terms of the absence of life, and LIFE as a condition opposite to death. This means that, on the one hand, the intersection of the two concepts is unable to satisfactorily express the phenomenon later conceptualised as BRAIN DEATH. On the other hand, the concepts that can be used to approach (albeit only approximately) the phenomenon at hand cannot be meaningfully combined, because combining them would result in a logical inconsistency. This raises a conceptual conflict.

2.4.2.3 Conflicts, gaps, and conceptual change

We want to point out that conceptual conflicts and conceptual disruptions can occur at different times in the very same process, and this has led to some confusion in the literature. For example, Hopster (2021, p. 4) suggests that the advent of ventilator technology in the mid 20th century caused a conceptual disruption in the form of an overlap for the concepts LIFE and DEATH, whereas Löhr (2022, p. 845) claims that the ventilator case is an example of a conceptual gap because it led to the introduction of the new concept BRAIN DEATH.

To pursue this matter, consider in more detail the invention of the mechanical ventilator and the conceptual disruption it caused, which is a popular example in the ethics of technology literature (Nickel, 2020; Nickel et al., 2022). Ventilators were invented in the 1950s and doctors realised that they could keep the basic processes of life like breathing, heart beating, cellular respiration going in patients even if those patients had no brain activity and would never again wake up or perform any of the higher functions of human life like talking or being aware of one's surroundings. This development led to a conceptual disruption in our concepts of life and death because these concepts at the time were based in part on whether a person was breathing on their own.

Here we focus on the concept DEATH, which had at the time as a constitutive principle that a person is dead if and only if that person is not breathing on their own. However, it was also standard to treat anyone whose heart was obviously beating as not dead. One problem was that after ventilators became available, the two criteria came apart. This is a conceptual conflict since some patients now seemed to be both dead (not breathing on one's own) and not dead (heart beating).

Although the history is complex (because of developments like organ transplantation, cardiopulmonary resuscitation, and the electroencephalogram) and we are idealising to some extent to highlight the features that illuminate our account of conceptual disruption, it makes sense to break down the development into the following stages. At first, we have the old conceptual scheme where DEATH is tied to *not* breathing on one's own. The new technology of ventilators then causes the *conceptual conflict* just described pertaining to the concept DEATH and the new facts about some patients on ventilators. We explained this conflict in terms of inconsistency between empirical facts (i.e., patients with no possibility of living on their own can be kept breathing and their hearts kept beating by a new technology).

Next, experts came to think that heartbeat and breathing were not the best criteria for understanding DEATH, which instead ought to be defined in terms of lack of brain activity. Thus, after this conceptual change, some patients whose hearts are beating and are on ventilators are actually dead. However, this change in the concepts LIFE and DEATH away from the lungs and heart and toward the brain left a *conceptual gap* – how do we label patients who are on ventilators and so are breathing and their hearts are beating but have no brain activity? This question is especially pressing because these are the best candidates for organ transplantation since organs quickly deteriorate once the heart stops beating.

A second conceptual change then occurred when the concept of *brain death* was introduced to classify patients who fit into this category. Therefore, in the development of a new conceptual scheme where LIFE and DEATH are defined in terms of brain activity, we see *both* a conceptual conflict *and* a conceptual gap, but these occur at different times in the development. The conceptual conflict occurs first and spurs one kind of conceptual change—redefining the concepts of life and death. This first conceptual change leaves a conceptual gap and so spurs another kind of conceptual change—introducing a new concept to use on those patients that fall into the gap.

The following four tables illustrate the development just described and highlight the two different kinds of conceptual disruption (conflict and gap) which occur at different times and the two kinds of conceptual change which also occur at different times in the development. The labels on the left display the shifting constitutive principles for being alive, being dead, and being brain dead (Tables 2.2a, 2.2b, 2.2c, 2.2d).

Breathing on one's own?	Heart beating?	Concept applied
Yes	Yes	ALIVE
No	No	DEAD

Table 2.2a. Old conceptual scheme

Breathing on one's own?	Heart beating?	Concept applied
Yes	Yes	ALIVE
No	Yes	? (conflict)
No	No	DEAD

Table 2.2b. Conceptual conflict after ventilator technology

Breathing on one's own?	Heart beating?	Brain activity?	Concept applied
Yes	Yes	Yes	ALIVE
No	Yes	Yes	ALIVE
No	Yes	No	DEAD (gap)
No	No	No	DEAD

Table 2.2c. New conceptual scheme with conceptual gap

Breathing on one's own?	Heart beating?	Brain activity?	Concept applied
Yes	Yes	Yes	ALIVE
No	Yes	Yes	ALIVE
No	Yes	No	DEAD (BRAIN DEAD)
No	No	No	DEAD

Table 2.2d. New conceptual scheme without conceptual gap

We want to emphasise that conceptual gaps and conceptual conflicts do not always display this pattern, but one can find this sort of development in a range of examples. Consider again the redefinition of PLANET discussed above.

The following tables illustrate the development described above and highlight the two different kinds of conceptual disruption (conflict and gap) which occur at different times in the development and the two kinds of conceptual change which also occur at different times in the development. The labels on the left display the shifting constitutive principles for being a planet, not being a planet, and being a dwarf planet. Again, these principles pertain to whether the object orbits the sun, whether the object is in hydrostatic equilibrium from its own gravity so as to take on roughly spherical shape, whether the object is in a class with few members, and whether the object has cleared debris from the neighbourhood of its orbit (Tables 2.3a, 2.3b, 2.3c, 2.3d).

Orbits sun?	Sphere?	Are there only a few?	Concept applied
Yes	Yes	Yes	PLANET
Yes	No	No	Not PLANET

Table 2.3a. Old conceptual scheme

Orbits sun?	Sphere?	Are there only a few?	Concept applied
Yes	Yes	Yes	PLANET
Yes	Yes	No	? (conflict)
Yes	No	No	Not PLANET

Table 2.3b. Conceptual conflict after Kuiper Belt discoveries

Orbits sun?	Sphere?	Are there only a few?	Concept applied
Yes	Yes	Yes	PLANET
Yes	Yes	No	Not PLANET (gap)
Yes	No	No	Not PLANET

Table 2.3c. New conceptual scheme with conceptual gap

Orbits sun?	Sphere?	Are there only a few?	Concept applied
Yes	Yes	Yes	PLANET
Yes	Yes	No	Not PLANET (DWARF PLANET)
Yes	No	No	Not PLANET

Table 2.3d. New conceptual scheme without conceptual gap

In the case of the development of PLANET and DWARF PLANET, we see roughly the same pattern as in the example with LIFE, DEATH, and BRAIN DEATH. The old conceptual scheme is disrupted first by a conflict that occurs after the introduction of a new technology. Then a new conceptual scheme is introduced that eliminates the conflict but contains a conceptual gap. Finally, a new concept is introduced to fill the conceptual gap.

2.4.3 Severity of conceptual disruption

We assign different degrees of severity to different instances of conceptual disruption, based on a combination of the type of conceptual disruption involved and the level at which such disruption occurs.²⁹ We consider severity not merely in terms of a higher or lower degree of uncertainty, but in terms of urgency for a conceptual change.³⁰

2.4.3.1 Mild conceptual disruption

We classify an instance of conceptual disruption as mild when it does not give rise to conceptual uncertainty. An example of this is a conceptual

²⁹ It should be noted that the severity attributions to follow are merely indicative and may vary depending on the concepts under consideration in a given case. Specifically, the degree of severity may vary depending (among others) on the range, depth, and valence of impacts of an instance of conceptual disruption (Hopster 2021). For example, it seems reasonable to expect that the “expected” severity may increase when foundational concepts are being disrupted, e.g., logical concepts and moral concepts.

³⁰ As such, despite the use of similar terminology, this should not be confused with the discussion regarding deep and shallow accounts of conceptual disruption as covered in Löhr (2023a).

disruption occurring at the level of the conceptual scheme that manifests itself as a conceptual gap understood in terms of the non-exhaustiveness of the conceptual scheme under consideration. This type of conceptual disruption does not give rise to conceptual uncertainty, but merely uncovers the absence of a needed concept. Indeed, when discussing conceptual gaps, there is a salient difference between not knowing which concept(s) to apply and knowing that no concept applies. Just like conceptual disruption is a type of change without being a type of *conceptual* change, conceptual disruption entails uncertainty, but it does not necessarily entail *conceptual* uncertainty.³¹ For example, a conceptual gap may result from a new technology being developed for which we do not yet have a concept. In such a scenario, we know that what we want to convey cannot be expressed using any of the concepts in our current conceptual repertoire. Therefore, it is possible for a conceptual scheme to not be exhaustive without such a conceptual gap to raise questions related to the application conditions of each of the concepts included in that scheme.

³¹ Let us consider the following example. I am getting ice cream and I can choose among several different flavours. There is a salient difference in terms of the nature of my uncertainty if I cannot choose a flavour because I enjoy several flavours equally or I do not particularly enjoy any flavour, and if I cannot choose a flavour because every flavour contains milk and I am lactose intolerant. In the first case, I am comparing flavours that I deem to be (at least) somewhat suitable choices, and my uncertainty lies in trying to understand which flavour I would prefer in that specific instance. I do not know which works best. In the second case, I know that no flavour of ice cream would be a suitable choice. I know that nothing works. Let us now illustrate this through an example more closely related to conceptual disruption.

2.4.3.2 Moderate conceptual disruption

We classify an instance of conceptual disruption as moderate when it gives rise to conceptual uncertainty, which can be mitigated provisionally, while a more adequate and permanent solution is being developed. An example of this is conceptual disruption occurring in the form of conceptual gaps intended as non-exhaustiveness at the level of conceptual clusters and individual concepts. For example, while the combination of “moral autonomy, human dignity, identity, personhood, liberty, anonymity and confidentiality” (van de Poel, 2020, p. 55) cannot fully express the phenomenon that can be conveyed by *PRIVACY*, it can provide an approximation of such a phenomenon that can serve as a meaningful placeholder, while the conceptual disruption is addressed and the conceptual gap bridged.

2.4.3.3 Severe conceptual disruption

Lastly, we classify an instance of conceptual disruption as severe when it gives rise to conceptual uncertainty, which cannot be meaningfully mitigated provisionally, i.e., before it is addressed and solved. An example of this is conceptual disruption occurring in the form of conceptual conflicts at the level of conceptual clusters. For example, in the case of the ventilator, *ALIVE* and *DEAD* cannot be combined to meaningfully serve as a placeholder to express the phenomenon later conceptualised as *BRAIN DEATH*, due to such concepts being mutually contradicting. This means that there is no effective temporary solution to the conceptual disruption, which signals the urgency of addressing the instance of conceptual disruption. Indeed, the literature seems to confirm this hypothesis, as the disruption of *ALIVE* and *DEAD* in the case of the ventilator has been addressed urgently after the introduction of the technology, and has prompted an urgent assessment and reconceptualisation of *DEATH*, which ultimately led to the introduction of *BRAIN DEATH* (Baker, 2013).

2.4.4 Comparison

Overall, we have offered many suggestions for how to develop the idea of a conceptual disruption, and many of these build on the existing literature. *First*, we suggest that conceptual disruptions be characterised in terms of interruptions in the normal functioning of a concept, cluster of concepts, or conceptual scheme. The definition we propose differs from

previous suggestions in ways outlined in Section §2.2. *Second*, we offer a new terminology for a kind of conceptual disruption. Instead of calling them ‘overlaps’, we use ‘conflicts’, which brings out the sense in which the concepts involved indicate conflicting or incompatible information. *Third*, we explain conceptual conflicts in terms of the constitutive principles of the concepts involved. This novel explanation fits well with our account of conceptual disruption, while also representing a significant step forward in the literature investigating the foundations of the phenomenon of conceptual disruption. *Fourth*, we offer a substantive account of the difference between conceptual disruption and conceptual change. Conceptual disruption is an interruption in the functioning of concepts, whereas conceptual change occurs when there is a change in the constitutive principles for the concepts involved. A difference in constitutive principles is a difference in concepts. Although it can, conceptual disruption need not force any conceptual changes. This account builds on the way these terms have been used in the literature on conceptual disruption and the literature on conceptual engineering. *Fifth*, we emphasise the distinction between disruptions involving a single concept, multiple concepts, and entire conceptual schemes. This distinction is essential for correctly characterising conceptual gaps, which are the other main kind of conceptual disruption we consider. This emphasis differs from what is found in the literature on conceptual disruptions. *Sixth*, we incorporate this distinction with the distinction in severity of conceptual disruptions, intended in terms of urgency for a conceptual change. Together, these distinctions allow for a fine-grained classification system of conceptual disruptions that builds on those found in the literature. *Seventh*, we explained how distinct kinds of conceptual disruptions can occur together in a single event. We provided two case studies, one involving the artificial respirator and one involving the definition of PLANET. In each case, there is a conceptual conflict caused by some new technology, which is then followed by a conceptual gap caused by how people reacted to the conceptual conflict. This dynamic relationship between kinds of conceptual disruptions has not been noticed before, and our account consistently explains why some theorists have labelled these cases as gaps and others have labelled them as conflicts. Moreover, we find the same pattern in two disparate examples. *Eighth*, our account of conceptual disruption fits well with existing accounts of kinds of conceptual engineering (explained in the next section). Explaining how our conceptual disruption in general, and various kinds in particular, interact with the process of conceptual

engineering is a major benefit of our view. Although the connection between the two areas of research is self-evident and has been noted in the literature on conceptual disruptions, our insight about how certain kinds of conceptual engineering are appropriate for certain kinds of conceptual disruptions is novel and significant.

2.5 The role of conceptual engineering

We have discussed the literature on conceptual disruption, provided new accounts of conceptual conflict and conceptual gaps, and applied these accounts to some examples of conceptual disruption. In each of the examples, it is clear that conceptual engineering can be a response to conceptual disruptions. *Conceptual engineering* is the topic of how to assess our concepts and how to improve concepts that turn out to be problematic (Cappelen & Plunkett, 2020). In each of the cases discussed in the previous section, there were multiple cases of conceptual engineering corresponding to the multiple cases of conceptual disruption. In each case, a new technology led to a conceptual conflict, and this conceptual conflict was eliminated by redefining the old concepts in a way that allowed them to be applied to new situations brought about by new technologies. In the ventilator case, the concepts of life and death were redefined, while in the planet case, the concept of planet was redefined. The redefinition eliminated the conceptual conflict in each case but led to a new conceptual disruption – a conceptual gap. In the ventilator case, the gap pertained to how to label patients on ventilators with no brain activity. In the planet case, the gap pertained to how to label celestial objects that had not cleared their orbits of debris. Then, in each case, a new concept was introduced to fill the gap. In the ventilator case, the new concept is BRAIN DEATH, and in the planet case, the new concept is DWARF PLANET.

In his recent influential discussion of conceptual engineering, Chalmers (2020) distinguishes two kinds of conceptual engineering: *re-engineering*, where existing concepts are updated or replaced and *de novo* engineering, where wholly new concepts are introduced. We want to draw attention to how this distinction shows up in our discussion of the two examples. In each case, the development begins with a new technology that causes a conceptual conflict, and then that conceptual conflict is removed by re-engineering. In these cases, the old concepts that display a problematic conflict are re-engineered with new concepts that are meant to do roughly the same job but without giving rise to the

problematic conflict. Moreover, in each case, the re-engineering to avoid conceptual conflict causes a new conceptual disruption in the form of a conceptual gap, which is then removed by de novo engineering. In these cases, an entirely new concept is introduced to classify phenomena that fell into the conceptual gap. Thus, it seems like re-engineering is the response to a conceptual conflict, whereas de novo engineering is the response to a conceptual gap. Moreover, the conceptual gap is caused by the re-engineering effort in light of the next technology in question. We want to emphasise that this pattern need not hold for all examples, but it does seem reasonable to see this distinction between kinds of conceptual engineering as a result of the distinction between kinds of conceptual disruption.

Some conceptual disruptions do not call for re-engineering or de novo engineering; instead, eliminating the concepts in question is the appropriate course of action. These conceptual disruptions are in the form of conflicts, and take place at the second and third level of conceptual disruption, affecting conceptual clusters and individual concepts. Such instances of conceptual disruption do not raise uncertainty about the constitutive principles or application conditions of the concept(s) under consideration, but uncover an incompatibility between one or more concepts and some facts about the world. Indeed, there is a salient difference between *not knowing* the constitutive principles of a concept or *not knowing* when and how a concept can be applied, and *disagreeing* with the application conditions of a concept.³² In these cases, such conceptual disruptions highlight grounds for the discontinuation or abandonment of one or multiple concepts. An example of this type of disruption can be drawn by considering how technology can cause some of our concepts to no longer apply, thus prompting the abandonment of such concepts.

2.6 Conclusion

³² For example, when an instance of conceptual opposition concerns a derogatory concept (e.g., a concept imbued with racial, sexist, homophobic, or ableist undertones), this does not mean that we do not understand the constitutive principles of such a concept, or that we do not understand how and when the concept can be applied. Rather, the instance of conceptual opposition uncovers a rupture such that we do not think the concept *should* be applied.

New technologies have the power to disrupt our social structures and even the very concepts we use to represent and understand the world. In this paper, we have critically engaged with recent work on conceptual disruption and offered an alternative account of the phenomenon of conceptual disruption, not only in terms of a new way to divide the kinds of conceptual disruption, but also as substantive accounts of the general category of conceptual disruption and two of its types: gaps and conflicts. We appeal to the idea that concepts have functions in order to explain *conceptual disruption* in general – interruptions in the normal functioning of a concept, cluster of concepts, or conceptual scheme. We distinguish conceptual disruption from *conceptual change* because the latter involves the introduction or elimination of a concept or the changing of a concept's constitutive principles. A *conceptual gap* occurs when there is some phenomenon that ought to be classified at a certain level of precision but no relevant concepts properly apply to the phenomenon in question. A *conceptual conflict* occurs just in case there is some known fact that is inconsistent with some constitutive principles for some concepts. We argue that it makes more sense to classify certain conceptual disruptions as conceptual conflicts rather than the established category of conceptual overlaps.

In addition, we explain how conflicts and gaps occur at different times in processes by which new technologies lead to changes in our concepts. In particular, we explained how gaps and conflicts show up in both the case of the mechanical ventilator (which spurred conceptual changes leading to the new concept of brain death) and astronomical technologies (which spurred conceptual changes leading to the new concept of dwarf planet). In each case, we saw that the new technology lead to a conceptual conflict, that was followed by a conceptual change. This conceptual change introduced a conceptual gap, which was then addressed by introducing a new concept to classify the phenomena in question. Throughout our study of conceptual disruption, we emphasised its heterogeneous nature and the fact that it can show up at different levels, including individual concepts, clusters of concepts, and conceptual schemes. Moreover, conceptual disruptions differ in severity, which we explain by appealing to how urgent it is to address them. Finally, we defended the idea that certain kinds of conceptual engineering are appropriate for addressing different sorts of conceptual disruptions. In particular, conceptual conflicts often induce either reengineering or abandonment, whereas conceptual gaps often spur *de novo* engineering where entirely new concepts are introduced.

Chapter 3

CONCEPTS AS ARTEFACTS. THREE INSIGHTS THAT CONCEPTUAL ENGINEERING CAN DRAW FROM THE PHILOSOPHY OF TECHNOLOGY³³

3.1 Introduction — 3.2 Thomasson's account of concepts as artefacts: strengths and limitations — 3.2.1 Concepts: broad and narrow accounts — 3.2.2 Artefacts: broad and narrow accounts — 3.3 Conceptual engineering meets the philosophy of technology. Two approaches, three insights — 3.3.1 Insight I: Concepts bear affordances and have side effects — 3.3.2 Insight II: Concepts are value-laden — 3.3.3 Insight III: Conceptual engineering would benefit from greater systematicity — 3.4 Conclusion

³³ This chapter appeared as an article: Marchiori, S. (2026). Concepts as artefacts. Three insights that conceptual engineering can draw from the philosophy of technology. *Inquiry*, 1–25.

3.1 Introduction

The conceptual engineering scholarship commonly discusses concepts in terms of something akin to artefacts. Concepts have been—seemingly uncontroversially—compared to tools, objects, artefacts, instruments, building blocks of thought (Margolis & Laurence, 2023). While many of such parallels remain tacit, numerous scholars explicitly draw similarities between concepts and artefacts, broadly understood.³⁴

Most notably, when discussing the practice of conceptual engineering, Thomasson (1999, 2007, 2021) argues that there are good reasons to '[think] first and foremost of [engineering *concepts* in terms of] engineering words, considering words as a certain kind of *abstract cultural artifact*' (Thomasson, 2021, p. 1, italics mine).³⁵ Moreover, while investigating conceptual functions and criteria for determining the adequacy of (existing) concepts, Haslanger (2000) wonders if '[such *concepts* are] effective *tools* to accomplish our (legitimate) purposes' (Haslanger, 2000, p. 33, italics mine). Similarly, while discussing the importance and desirability of philosophers engaging in conceptual engineering, Eklund (2014) argues that 'philosophers should also be asking themselves whether [our actual *concepts* or the properties of relations they stand for] really are the best *tools* for understanding the relevant aspects of reality' (Eklund, 2014, p. 295, italics mine).

Furthermore, Miracchi (2021) states that '*concepts* are the *tools* with which we reason' (Miracchi, 2021, 12028, italics mine), Klenk (2024 [2021]) discusses the Concepts-As-Tools hypothesis, arguing that '*concepts* are *like tools*' (Klenk, 2024 [2021], p. 2884, italics mine), Kocurek (2022)

³⁴ This extends even beyond the 'core' conceptual engineering literature, provided that any such thing exists. Indeed, broadening the picture to include conceptual engineering-adjacent literature, Gallie (1955) talks of '[understanding] a *concept* or theory or other tool of thought' (Gallie, 1955, p. 197, italics mine). Similarly, when talking about concepts in law, Ginsburg & Stephanopoulos (2017) write that '[c]oncepts are the *building blocks* of legal doctrine' (Ginsburg & Stephanopoulos, 2017, p. 147, italics mine).

³⁵ Thomasson's account will be addressed in the following section. For now, it suffices to know that, while Thomasson focuses on *abstract cultural* artefacts, the focus on this paper is on *socio-technical* artefacts. In this paper, I do not wish to argue that concepts are themselves socio-technical artefacts. Rather, I intend to highlight the similarity between concepts and socio-technical artefacts and how merely establishing such an analogy between the two can lead to fruitful insights for conceptual engineers.

writes that '*[c]oncepts are like tools that we can employ for various purposes*' (Kocurek, 2022, p. 4, italics mine), and Hopster & Löhr (2023) write that '*concepts frequently serve as tools*' (Hopster & Löhr, 2023, p. 12, italics mine). Again, when discussing the adoption of a process of reverse engineering of concepts to determine their functions, Veluwenkamp & van den Hoven (2023) state that '[t]he idea behind reverse engineering [of *concepts*] is that we discover the function of an *object* by disassembly and observation of its behavior in different contexts. This is *not unlike* what archaeologists do when they find puzzling *objects* or *artefacts*' (Veluwenkamp & van den Hoven, 2023, p. 6, italics mine).

Although several scholars have gestured at parallels between concepts and artefacts, there remains little sustained engagement with the extensive theorisation of artefacts in the philosophy of technology literature. With few exceptions (e.g., Veluwenkamp & van den Hoven, 2023; Veluwenkamp & Köhler, 2024; Marchiori, 2025)³⁶, references to concepts as artefacts remain often metaphorical in nature, rather than attempting to draw from, or contribute to, the philosophy of technology literature.³⁷ As such, little to no explicit connection has been established between the conceptual engineering literature on concepts and the philosophy of technology literature on (socio-technical) artefacts. Consequently, the relation between concepts and artefacts—as understood in the philosophy of technology—remains broadly unexplored terrain.

This is a missed opportunity, as much is to be gained from a deeper exploration of the convergence of the two bodies of scholarship. My starting point is the intuition, stemming from the conceptual engineering literature, that concepts resemble artefacts in some salient respects. That is, that, at a minimum, there are good reasons to believe that concepts are *like* artefacts. Here, I explore this intuition further and zoom in to consider the relation between concepts and socio-technical

³⁶ Veluwenkamp & Köhler (2024) draw analogies between concepts and artefacts, insofar as their (normative) functions are concerned, while Marchiori (2025) draws from the philosophy of technology scholarship in on conceptual functions and affordances to align conceptual functions (and affordances) with artefactual functions (and affordances).

³⁷ While Chalmers (2020) also invokes the artefact analogy in framing normative conceptual work through the engineering metaphor, the parallel between concepts and artefacts is illustrative rather than systematic, and does not engage with the philosophy of technology literature.

artefacts specifically. In so doing, I begin by advancing an account of concepts as akin to socio-technical artefacts. I ultimately argue that the Western philosophy of technology literature on socio-technical artefacts is not only extremely relevant for conceptual engineering, but can spotlight insights and approaches that are un(der)explored in conceptual engineering, but would greatly benefit from further investigation. In short, I present three ways in which conceptual engineers should tap into the philosophy of technology literature.

To make this case, I start by clarifying in Section §3.2 how this paper relates to existing work in conceptual engineering on the relation between concepts and artefacts. Here, I introduce an account of concepts as akin to socio-technical artefacts and illustrate salient differences and commonalities between Thomasson's work on concepts as abstract *cultural* artefacts (1999, 2021) and this paper's focus on concepts as (abstract) *socio-technical* artefacts. In Section §3.3, I argue that much is to be learnt from insights and approaches developed in the philosophy of technology on socio-technical artefacts, and present three pointers for future research in conceptual engineering. First, in Section §3.3.1, I argue that conceptual engineering should not only focus on conceptual functions, but integrate such an approach with the investigation of conceptual affordances and side effects. Second, in Section §3.3.2, I claim that engineered concepts are value-laden, and that this insight should guide approaches to the evaluation, design, and implementation of concepts. Third and last, in Section §3.3.3, I argue that conceptual engineering would benefit considerably from a greater focus on methodological work. Section §3.4 contains the conclusion.

3.2 Concepts as (socio-technical) artefacts

Perhaps the most notable example of concepts being compared to artefacts in the conceptual engineering literature is the one offered by Thomasson (1999, 2007, 2021), who has explored at length the connection between concepts and artefacts, ultimately arguing that there are good reasons to '[think] first and foremost of [engineering *concepts* in terms of] engineering words, considering words as a certain kind of *abstract cultural artefact*' (Thomasson, 2021, p. 1, italics mine). Given the focus of this paper on artefacts from a philosophy of technology perspective, it is not immediately clear that Thomasson's work on concepts as artefacts will be a relevant starting point for this investigation. Therefore, two clarifications are in order with regard to what one takes concepts and

artefacts to be, respectively. This will illuminate the commonalities and differences between the approach of concepts as *abstract cultural* artefacts proposed by Thomasson and the one of concepts as *socio-technical artefacts* advanced in this paper, which I argue better suits conceptual engineering in the philosophy of technology, while also being broadly applicable to conceptual engineering at large.

3.2.1 Concepts: broad and narrow accounts

To contextualise Thomasson's account of concepts as (abstract) cultural artefacts, it is important to note that her approach to conceptual engineering heavily draws from, and is grounded in, the scholarship on the ontology of words (e.g., Thomasson, 1999, 2021; Kaplan, 1990; Irma, 2019). Specifically, Thomasson proposes that the object of conceptual engineering should be words. Indeed, she argues that 'if we think first and foremost of [conceptual engineering as the practice of] engineering words, and think of words as a certain kind of cultural artifact, governed (like other artifacts) by norms of use, then we can begin to understand better how words do change naturally, and use that as a way into seeing how they (like other cultural artifacts and normative practices) can also be changed artificially' (Thomasson, 2021, p. 2). In so arguing, Thomasson employs a narrow notion of concept which, if not completely, substantially overlaps with that of word, at least for the purpose of conceptual engineering.

While on Thomasson's account concepts are words and words are abstract cultural artefacts, the move from concepts to artefacts is not a straightforward one outside of Thomasson's paradigm. To follow this line of reasoning, one would first need to subscribe to Thomasson's account of concepts as words, at least within the confines and for the purposes of conceptual engineering. This is not an uncontroversial position in the conceptual engineering literature (e.g., Lalumera, 2009; Machery, 2009). Furthermore, and perhaps most crucially, Thomasson's account of concepts as words may be overly restrictive, and there may be good reasons to explore how concepts are artefact-like regardless of whether they are equated with words.

In contrast to Thomasson, in this paper, I do not commit to a narrow understanding of concepts as words, but rather employ a broader account of concepts, which I understand as a function kind (i.e., what makes something a concept is its role in the cognitive system; as in Lalumera, 2009). In doing so, I remain uncommitted as to which

structure, if any in particular, concepts (do or should) have. Therefore, I am not bound to the assumption of uniformity, and admit a plurality of conceptual structures. This makes the notion of concept I will employ in this paper significantly broader than Thomasson's.³⁸

3.2.2 Artefacts: broad and narrow accounts

Second, it should be noted that, when Thomasson establishes a parallel between concepts (read: words) and artefacts, she primarily focuses on artefacts understood as *abstract cultural* artefacts. Instead, the focus on this paper is on artefacts as understood from a philosophy of technology perspective, hence *socio-technical* artefacts. This is relevant in that, as will become clearer in the following sections, it is not straightforward that either abstract or cultural artefacts qualify as socio-technical artefacts as understood in the philosophy of technology. Once again, one can identify two main conceptualisations of *artefact*, distinguishing between a narrow and a broad account.

On the one hand, a longstanding tradition in the philosophy of technology maintains that technology is anything human made (Hickman, 2001). Correspondingly, based on this rather broad view of

³⁸ The notion of concepts as a functional kind adopted in this paper is intentionally broad. The analogy between concepts and artefacts is often articulated at the functional level: what makes the comparison productive is precisely that both concepts and artefacts serve specific purposes. Investigating concepts as a functional kind keeps this shared dimension in focus without foreclosing questions about the extent to which concepts and artefacts may differ. Moreover, understanding concepts as a functional kind does not commit one to any particular view of their structure and is compatible with diverse positions currently defended in the literature, among which word-focused, definitional, and prototypical views of concepts (Margolis & Laurence, 2023). Remaining uncommitted on conceptual structure allows this investigation to proceed at a level of generality that is appropriate for an analogy-based argument, while remaining open to further specifications that any particular conceptual engineering project may require.

technology, anything human made is a socio-technical artefact.³⁹ Moreover, a well-established tradition in philosophy of language and linguistics specifically considers language to be a technology or artefact (Mufwene, 2013). That is, language is a tool that we have created to think and to communicate. Therefore, to the extent that concepts can be understood as (at least partially) overlapping with language, as they are in Thomasson's account, it seems intuitively reasonable to think of (at least some) concepts as socio-technical artefacts.

On the other hand, the broad conceptualisation of *artefact* is not uncontroversial. Indeed, a considerable strand within the philosophy of technology literature prefers a narrower account of both *technology* and *artefact*, often motivated by practical considerations. The intuition here is that, if it is the case that anything human made is an artefact, then being able to classify something as an artefact would cease to be helpful in any meaningful way.⁴⁰ This intuition can be substantiated by approaching the study of artefacts departing from a narrower understanding of technology as engineering. According to this view, artefacts are generally understood as intentionally modified material objects (Houkes &

³⁹ For the purpose of this paper, the distinction between *technology* and *artefact* is not salient. However, it should be noted that the philosophy of technology literature—in particular the portion of the scholarship subscribing to a narrow understanding of both technology and artefact—has investigated the relation between these two notions at length and formulated numerous hypotheses and frameworks to understand how they relate to each other (e.g., Brey, 2012; Soltanzadeh, 2021).

⁴⁰ Consider the concept 'planet'. Advances in observational technologies between the 1990s and 2000s revealed a proliferation of entities that satisfied the prevailing criteria for planethood, thereby threatening to expand the category beyond practical or theoretical utility by rendering the concept insufficiently selective. As a result, the International Astronomical Union (IAU) redefined 'planet' by introducing a new constitutive condition aimed at restoring the concept's discriminatory usefulness (IAU General Assembly, 2006). This example illustrates that, at least in some contexts, the adequacy of a concept turns on its capacity to meaningfully classify phenomena, which may be compromised when the extension of the category becomes too broad. A similar concern may arise in the philosophy of technology: one may argue that, if the category of socio-technical artefact is expanded to include all human-made entities, its classificatory and explanatory utility risks being diluted to the point of vacuity.

Vermaas, 2010).⁴¹ Such a narrow understanding of *artefact* is often investigated by dual theories of artefacts, which focus on artefacts as a blend of structure and function.

Two points are of particular interest here. First, it seems to be the case that, when conceptual engineers establish a connection between concepts and artefacts, they are adopting a broad characterisation of artefacts. Hence, Thomasson's account of concepts as abstract cultural artefacts, which employs a similarly broad view of artefacts, seems to be *prima facie* consistent with, and representative of, the way in which artefacts are conceptualised by conceptual engineers when they are compared to concepts. Second, if the previous point holds, it is not immediately clear that such a broad account of artefacts would be compatible with the understanding of artefacts that is prevalent in the philosophy of technology literature. Indeed, from a philosophy of technology perspective, the relation between cultural and socio-technical artefacts is not a neat one: based on different accounts of both concepts and artefacts, abstract cultural artefacts can be considered as partially or fully overlapping with socio-technical artefacts, as a subcategory of socio-technical artefacts, or as distinctly separate entities from socio-technical artefacts, with no salient points of commonality.

In brief, while conceptual engineers often compare concepts to artefacts, few scholars have so far expanded on the relation between the two, among which, perhaps most notably, Thomasson (1999, 2007, 2021). However, Thomasson's account of concepts as artefacts is not uncontroversial, as it employs too narrow a notion of concepts from a conceptual engineering perspective and potentially too broad a notion of

⁴¹ It should be noted that substantive objections could be, and have been, moved to the intentionality, modification, and physicality requirements. First, the intentionality requirement could be undermined when understood in terms of intentional design, either by highlighting how, historically, not all artefacts are the result of a design process, or by noting how, if artefacts are indeed intentionally designed, this would make their existence mind-dependent. Second, one could put pressure on the modification requirement from an ontological or mereological perspective, e.g., by questioning whether the process of modification of already existing objects can indeed lead to the creation of a new (material) entity or being. Third and last, the physicality requirement can be undermined by the advent of digital artefacts, such as computer programs and databases, which are rather uncontroversially understood in artefactual terms, but can also be considered immaterial. In this case, emphasis is being placed on the fact that, despite requiring a physical medium to exist, such artefacts are not understood in traditionally material terms (e.g., see Shep, 2015, about digital materiality).

artefacts from a philosophy of technology perspective. As such, Thomasson's account may not be satisfactorily moved across disciplinary boundaries without losing (at least) part of its cogency, and its impact on the investigation at hand is limited. Therefore, the account of concepts advanced here is not a refinement or extension of Thomasson's, but a distinct account of concepts as artefacts. Ultimately, in this paper, I adopt a broad notion of concepts as functional kinds, individuated by their function(s), and a narrower account of artefacts as socio-technical artefacts, i.e., objects that have been created or modified by humans and fulfil a function. For reasons of conciseness, I will proceed to use 'artefact' to mean socio-technical artefact for the remainder of the paper, unless otherwise specified.

3.3 Conceptual engineering meets the philosophy of technology. Two approaches, three insights

In the following sections, I discuss three main ways in which the philosophy of technology scholarship can provide helpful insights for conceptual engineering, particularly in relation to the evaluation and (re-)design of concepts. I refer to such perspectives in terms of a *function-first* approach—focussing on what concepts and artefacts *can* do and on what they *should* do—and a *design-first* approach—focussing on how to decide what concepts and artefacts should do, and how to (re-)design them in such a way that they are more likely to do what we wish them to do.⁴² To clarify the distinction, consider a hammer. A function-first approach would analyse what the hammer allows us to do (e.g., driving nails) and what it should allow us to do (e.g., driving nails efficiently and safely). Meanwhile, a design-first approach would focus on how to design a hammer to achieve these goals (e.g., decisions about materials, handle ergonomics, and head weight distribution).

3.3.1 Insight I: Concepts bear affordances and have side effects

The first parallel between concepts and technical artefacts that warrants further examination concerns the centrality of the role of functions in the respective bodies of literature. Building on functionalism as an entry

⁴² The function-first approach and design-first approach are not mutually exclusive.

point for the exploration of the similarities between the study of concepts and that of artefacts, this section illustrates how concepts, like artefacts, do not merely possess functions, but also give rise to affordances and side effects.

Functionalism, i.e., the view that something (in this case, a concept or an artefact) is at least partially determined by its function(s), is discussed at length both in conceptual engineering and in the philosophy of technology. Both bodies of scholarship present a broad spectrum of positions with regard to functionalism, ranging from functionalism opponents, to functionalism supporters, to functionalism advocates. Some scholars deny the benefits of a functional account of concepts and artefacts, mostly on the basis on its triviality, i.e., it is trivially true that concepts and artefacts have functions, and functionalism adds nothing meaningful to their conceptualisation or study (e.g., Cappelen, 2018). The opposite view is also present, with both literatures featuring a version of functional essentialism, according to which concepts and artefacts are identified based *solely* on their function (e.g., Lalumera, 2009).

However, many (perhaps, most) scholars fall in a middle position between such extremes. This is true for both conceptual engineers and philosophers of technology. In conceptual engineering, this means subscribing to the view that concepts have functions, without necessarily committing to functions being the essential or constitutive feature of concepts. Numerous conceptual engineers are aligned with this view (e.g., Cappelen & Plunkett 2020; Chalmers 2020; Haslanger 2020; Plunkett & Sandell, 2013; Prinzing, 2018; Queloz 2019; Simion & Kelp 2019; Sundell 2020; Thomasson 2020; to a lesser extent: Jorem, 2022; Riggs, 2019). Similarly, numerous approaches within the philosophy of technology widely endorse moderate forms of functionalism, which typically take the form of hybrid accounts, wherein artefacts are individuated through a combination of their functional roles and additional criteria (e.g., structure, intention). One of such accounts, the dual nature theory of technical artefacts illustrates the intuition that artefacts combine structure and function: they are described 'structurally, in terms of their physicochemical capacities; [and] teleologically in terms of goals and actions' (Houkes & Vermass, 2010, 138). Both bodies of scholarship also draw on the (philosophy of) biology literature, most notably Millikan's work on biological, teleological, and proper functions (Millikan, 1984, 2000). This lends further credence to the intuition that both literatures share remarkably similar intuitions, reinforcing the parallel between concepts and artefacts.

In short, the functional perspective offers a salient and productive lens through which to examine both concepts and artefacts, and has been the subject of extensive investigation in their respective domains. This suggests that an adequate understanding of concepts and artefacts benefits from attending to the roles or functions they are designed or expected to fulfil—i.e., what concepts and artefacts *should* do. This intuition has been further explored by the philosophy of technology scholarship through two main moves: the first move, from functions to (inter)actions; the second move, from functions, through (inter)actions, to affordances. Let us consider both in turn.

The intuition that artefacts are, at least in part, identified based on their functions has been investigated in depth within the philosophy of technology literature (e.g., Hansson, 2006; Vermaas & Houkes, 2003, 2006; Houkes & Vermaas, 2009). The starting point of such investigations is the following intuition: if artefacts (are meant to) serve specific functions, studying artefacts in isolation would be limiting, in that it would keep considerations which should be relevant at arm's length. Consequently, the study of artefacts should not adopt an *entity*-first perspective, but an (*inter*)*action*-first perspective. That is, artefacts should be studied as part of the (inter)actions in which they are (expected to be) used, where their functions are put to use. Here, I will sketch two ways in which this approach has been developed in the philosophy of technology scholarship: mediation theory and activity realism.

Mediation theory (also known as the theory of technological mediation) is an approach in philosophy of technology that emphasises the interconnectedness between humans and technology. It draws from post-phenomenology, i.e., an approach exploring the ways in which artefacts mediate our actions, perceptions, and experiences of the world (Ihde, 1990), and research in science and technology studies such as actor-network theory (Latour, 1987; 2007) and the social construction of technology (Pinch & Bijker, 1984), arguing that artefacts are not passive tools, but active agents that contribute to shaping social relations. Mediation theory highlights how society and technology influence and shape each other, and provides a framework for understanding the complex relationships between humans and artefacts, shedding light on the active roles that artefacts play in mediating our understanding of, and relationships with, the world (e.g., how artefacts mediate our perceptions, interactions, and experiences), as well as its ethical implications (Verbeek, 2005, 2015).

Alongside mediation theory, the shift towards relational analyses of artefacts has been advanced through activity theoretical approaches, which share the intuition that the conceptualisation and study of artefacts should be practically significant. According to such approaches, when artefacts are studied from a metaphysical standpoint that is divorced from practical considerations, the artefact/non-artefact distinction is rendered inert. Consequently, the question of whether something qualifies as an artefact becomes practically useless (Houkes & Vermaas, 2009). First introduced by Soltanzadeh (2015, 2016, 2019, 2021), activity realism exemplifies the broader family of activity theoretical approach.⁴³

Activity realism classifies entities as artefacts in relational terms, i.e., based on the relation between such entities and 'purposeful activities of reflective beings' (Soltanzadeh, 2021, p. 33). This can be contrasted with *entity realist* accounts, which classify entities as artefacts based on features that they possess in isolation.⁴⁴ According to the activity realist perspective, artefacts 'do not become practically significant because they are artefacts; rather, because they play roles as means to achieve ends' (Soltanzadeh, 2021, p. 28). Thus, activity realism suggests that a material object is an artefact when it is used in problem-solving activities (Soltanzadeh, 2021). In this sense, the categorisation of an object as an artefact is not static, but dynamic, and objects can acquire and lose their

⁴³ In the words of Soltanzadeh (2021),

This is not to say that artefacts are not useful—we use artefacts on a daily basis. Rather, *the metaphysical distinction between artefacts and non-artefacts is not useful*. To repeat, a categorisation is useful if our practical decisions are sensitive to whether we are dealing with objects that belong to the category or those which do not. The artefact/non-artefact distinction does not require us to use different decision-making or evaluators frameworks, and is thus not impactful. It is of little to no use in guiding practical actions (Soltanzadeh, 2021, p. 28, emphasis added).

⁴⁴ A parallel can be drawn here with debates regarding the ontological status of concepts as natural kinds versus functional kinds. Machery (2009) argues for an eliminativist stance, owing to the absence of a uniform set of intrinsic properties shared across all concepts. Analogously, entity realist approaches classify artefacts based on internal, self-standing traits. In contrast, activity realist perspectives ground the status of artefacts in relational terms—namely, their embeddedness within purposeful human (or cognitive) activities and contexts of use. This is reminiscent of functional kind theories of concepts (Lalumera, 2009).

status of artefact over time, depending on whether they are being used in problem-solving activities.

While similar intuitions have only been explored to a limited extent within normative conceptual work (e.g., Brun, 2022), the conceptual engineering literature does feature positions compatible with, or at least adjacent to, the (inter)action-first approach. For example, one may argue that the case has been made in conceptual engineering in favour of individuation criteria for concepts that are not primarily based on the properties of concepts as entities taken in isolation, but are rather attributed on the basis of the conditions associated with the actor who interacts with the concept. Based on such accounts, to be a concept is to be grasped and used for specific purposes, e.g., to think, to draw inferences, to judge (e.g., as in Peacocke's possession conditions; Peacocke, 2005; Lalumera, 2009).

The intuition that artefacts have functions has been further explored in a second move, one that synthesises functional and (inter)activity-based perspectives. Whereas the first move reframes artefacts as elements within a process of interaction, the second expands this to consider the broader context of use. In other words, the focus shifts from what artefacts (should) do to what they enable someone to do within a given context and with respect to given users. Therefore, attention should not be placed on artefactual functions alone (i.e., what artefacts do or should do), but should also extend to their affordances (i.e., what artefacts could enable us to do).

The notion of affordances describes the possibilities for action that an object or environment offers to an individual (Gibson, 1977a, 1977b). In the context of artefacts, affordances refer to the potential actions that are available to users when interacting with a device or system. This includes both intended and unintended uses of technology, as well as the constraints that may limit certain actions. Consider an e-reader. Its buttons suggest specific possibilities for interaction, such as switching the device on or off, turning pages, and adjusting font size. These affordances guide user behaviour and facilitate intuitive navigation. Additionally, the physical characteristics of e-readers enable uses that extend beyond their intended functions. For example, they can serve as paperweights, trays, and even light sources. Conversely, what e-readers do not (currently) enable are activities such as watering plants, making coffee, or flying. In short, affordances shape how users approach, interpret, and engage with technology in everyday life.

The scholarship on artefactual affordances emphasises the relational properties of artefacts: what an artefact enables does not emerge solely from the artefact itself, but from its interaction with specific users in a specific context. Supplementing functionalism by incorporating the notion of affordances thus yields a more nuanced understanding of how artefacts shape human behaviour and vice versa (Bærentsen & Trettvik, 2002; Yaneva, 2015). From a design standpoint, mapping the affordances of a given artefact allows designers to anticipate patterns of interaction and to design for intuitive and effective user experiences (Gaver, 1991; Tollon, 2022; Vyas et al, 2006).

Within the (moral) philosophy of technology, affordances have been investigated primarily through two interrelated lenses. The first concerns the ethical implications of artefacts (e.g., Swierstra & Waelbers, 2012). In both constraining and enabling particular actions, artefacts can give rise to (unintended) consequences (and side effects) affecting individuals, the environment, or society at large. By attending to both the intended and unintended affordances of artefacts, philosophers are better equipped to assess their ethical implications, as well as to contribute to their responsible design, development, and deployment.

The second lens through which affordances have been examined concerns on the values embedded in artefacts (Klenk, 2020). Indeed, affordances can reflect and reinforce specific values (such as efficiency, transparency, or control), influencing how users perceive and interact with an artefact. Therefore, analysing the affordances of artefacts can reveal the underlying assumptions, values, normative priorities, and goals, which are inscribed into artefacts' design and which inform their function(s). Such analysis, in turn, can shed light on the evaluation of the ethical implications of artefacts, providing a starting point for the evaluation of value alignment, and offering guidance for value-sensitive design practices.

This leads to a first key insight for conceptual engineering. Having acknowledged the analogy between concepts and artefacts, it becomes reasonable to expect that, similarly to artefacts, concepts themselves may also bear affordances and have side effects. That is, even when designed to fulfil specific functions, concepts will invariably enable additional (foreseen or unforeseen) usages and have a broader range of impacts beyond those intended by their designers. This points to a source of largely untapped potential for conceptual engineering, in that it suggests that conceptual engineering should attend to both the functions and the affordances of concepts (Marchiori 2025). This extends the

current functional debate in conceptual engineering (e.g., Cappelen, 2018; Nado, 2021; Riggs, 2021; Jorem, 2022) in several directions.

Attending to the affordances and side effects of concepts yields a more comprehensive picture of the full impact of both existing concepts and conceptual proposals prior to their deployment and adoption. This enhances conceptual evaluation and broadens the diagnostic toolkit of conceptual engineering, improving the identification of inadequacies warranting a conceptual engineering intervention. Moreover, by mapping the functional and affordances profile of a given concept, conceptual engineers may more aptly identify risks and impacts before they materialise, and extend conceptual engineering approaches from (predominantly) reactive to proactive interventions (e.g., Brey, 2012; Floridi & Strait, 2021; Lundgren, 2024; Marchiori & Sta. Maria, 2026). This invites anticipatory approaches to conceptual engineering, which may be particularly valuable whenever post-hoc conceptual revisions may be costly or face implementation issues. Ultimately, while functionalism has its place in conceptual engineering, and it is fruitful to investigate whether and to what extent concepts fulfil the functions we wish them to fulfil, more attention should be placed on investigating what concepts enable and constrain.

3.3.2 Insight II: Concepts are value-laden

A second salient similarity between conceptual engineering and the philosophy of technology lies in their shared focus on design. This section adopts a design-first lens to explore how the philosophy of technology treats artefacts as value-laden, and to argue that recognising the value-ladenness of concepts should be central to their evaluation, design, and implementation. In both literatures, design is a central object of analysis. In conceptual engineering, which is often described in terms of the evaluation, design, and implementation of concepts, design encompasses both *de novo engineering*—i.e., creating new concepts from scratch—and *re-engineering*—i.e., the modification of existing concepts (Chalmers 2020). Likewise, in the philosophy of technology, artefacts are examined through the lenses of the interconnected processes of design,

development, and use—where design is understood as the purposeful (re)structuring of a material object towards an end.⁴⁵

This shared concern with design also invites reflection in both literatures concerning what should qualify as design. For example, in conceptual engineering, some concepts (such as scientific and expert concepts) may be more intuitively recognised as the outcome of a design process, whereas others (like many lay concepts) may rather seem to emerge organically and be adopted through a function-driven selection process. Comparable debates arise in the philosophy of technology with regard to artefacts. Indeed, it may be the case that, even when an entity rather uncontroversially qualifies as an artefact, it may not be clear whether it was the product of a design process. For example, naturally modified branches used for fishing are often considered artefacts, even when left entirely unaltered by humans.⁴⁶

From here, both literatures extend their analysis to broader meta-theoretical questions about design (e.g., Feng & Feenberg, 2008; Grant & Frank, 1992; Potter, 1980): Who should qualify as a designer? Is design a specialised activity or can laypeople also be designers? The philosophy of technology scholarship further addresses in great depth questions which seem directly relevant for conceptual engineering, such as: Should design be open to a wide audience because, at least in part, of the wider impact of the implementation of an artefact on society at large? Should design be a participative activity bringing together experts and laypeople? Such questions have helped foreground one assumption that has greatly shaped the contemporary Western philosophy of technology scholarship on artefacts, and which offer fertile ground for further exploration within conceptual engineering: artefacts are value-laden rather than value-neutral.

⁴⁵ While such an understanding of what the design of concepts and artefacts amounts to is sufficient for the purpose of the discussion at hand, it should be noted that this greatly simplifies the debates surrounding the notion and process of design, which are the object of rich debates in the respective bodies of scholarship.

⁴⁶ This complexity deepens when considering how design unfolds across domains. For example, architectural design, software engineering design, and the design of everyday products, each follow markedly different processes. Such diversity may suggest that a broad notion of what design entails might be preferable to account for all such variations under the umbrella term of design (Archer, 1979; Ralph & Wand, 2009; Sun, 2021; Vissers et al, 2016).

The intuition that artefacts are inherently value-laden underpins much of contemporary Western philosophy of technology. The value-leadenness thesis rests on the understanding of artefacts as inevitably carrying the imprint of their designers. That is, designers hold preconceptions, biases, and values, that inevitably make their way into the artefacts that they design, whether intentionally or unintentionally. This parallels the intuition that concepts have affordances and enable actions beyond designer intent. The value-leadenness position emerged as a corrective to the neutrality thesis, according to which artefacts are value-neutral. By recognising that technology is not neutral, but actively shapes and reflects societal values and power dynamics, philosophers of technology transitioned towards a value-laden view of artefacts as imbued with social, cultural, and moral values (Winner 1980).

Acknowledging the value-laden nature of artefacts has prompted philosophers of technology to place greater emphasis on the investigation of the ethical implications of technological innovation and the role of values in the design of artefacts. Indeed, recognising the values embedded in artefacts allows for a more critical analysis of technology and promotes a more responsible approach to technological development. This also led to discussions on the role that society at large should play in the creation of artefacts, leading to participatory and value-sensitive design approaches aimed at ensuring that artefacts reflect diverse values and stakeholder perspectives. Moreover, the recognition of artefacts as value-laden sheds light on the reason why philosophers of technology overwhelmingly refer to artefacts in terms of *socio-technical* artefacts, to underscore the interconnectedness of technology and society (Mumford, 1987; Verbeek, 2005).

This leads us to the second source of untapped potential that conceptual engineering can derive from the philosophy of technology: concepts, like artefacts, are not value-neutral, but value-laden. (Re-)designed concepts are shaped by, and in turn reflect, the values of their designers, as well as those embedded in the broader socio-technical landscape in which conceptual engineering takes place. This carries significant implications for conceptual engineering.

On an abstract level, recognising that all (re-)engineered concepts are value-laden can seem, at first glance, to blur the familiar distinction between normative and descriptive concepts—the former distinguished by the presence of an evaluative element in their content, the latter by its absence. Intuitively, a concept cannot be purposely designed without normative considerations playing a role in its creation. The worry is that

if every act of conceptual design is informed by values, then no (re-)engineered concept is truly value-neutral. Indeed, considerations such as accuracy, efficiency, safety, and usefulness are values no less than fairness, responsibility, freedom, or sustainability. From this perspective, it might appear that all (re-)engineered concepts retain an evaluative component in virtue of their value-laden genesis. Yet this inference risks conflating two different senses in which a concept can be value-laden, and thereby misclassifying concepts whose content is not itself evaluative as 'normative' in the technical sense. First, a concept can be *intrinsically evaluative* in its content, such that grasping and competently applying it requires making a normative judgment. Such concepts qualify as normative concepts. Second, a concept can be *extrinsically value-laden*, even if its content is purely descriptive, by virtue of its place within a value-laden framework, practice, or institutional setting.

Recognising this distinction is crucial for conceptual engineering. Without it, there is a risk of conflating all (re-)engineered concepts with normative concepts, simply because all are, in some sense, value-laden. Yet this would obscure important differences in how concepts should be evaluated, designed, and implemented. Concepts that are normatively evaluative in content require scrutiny of their evaluative dimensions and their justificatory grounds as well as of their broader impact on the socio-technical landscape; descriptive concepts embedded in value-laden frameworks require scrutiny of their socio-technical embedding, anticipated impacts, and alignment with broader value commitments, even if their content remains non-evaluative. Ultimately, the implication is not that the distinction between normative and descriptive concepts collapses following the recognition of value-ladenness, but rather that it must be supplemented by a second dimension: the source and mode of that value-ladenness. A solution may be to distinguish between content-level evaluativity and context-level evaluativity, and adopt different evaluation and design strategies for each.

A further implication of the value-ladenness of (re-)engineered concepts concerns its implications for the evaluation, design, and implementation of concepts. That is, if (re-)engineered concepts are the bearers of values, more (and more systematic) attention should be paid to investigating the anticipation, identification, evaluation, and (if necessary) mitigation of the satisfaction of such values, especially during the evaluation and design stage of the conceptual engineering process. This has been the object of recent work in conceptual engineering aiming to develop substantive accounts of the (problematic) mismatch between

the values portrayed and perpetuated by our concepts and societal values (i.e., conceptual misalignment, Hopster & Löhr, 2023; Hopster et al, 2024; Marchiori & Scharp, 2024).

One last observation is in order. Here, we have focused primarily on the design process, arguing that value-ladenness is inescapable since the purposes a concept is intended to serve, the trade-offs accepted in its structure, and the affordances it is meant to enable are shaped by explicit or implicit values. It is reasonable to expect the value-ladenness of concepts to pervade any conceptual engineering project, beyond those focussing specifically on conceptual (re-)design. Indeed, assessing whether a concept is adequate, defective, or in need of revision is itself an evaluative act: the criteria employed, the weight assigned to different functions or affordances, and the context in which the assessment is conducted all reflect particular value commitments. Similarly, deploying a concept assumes that it has been judged fit for purpose, which entails a prior evaluative decision about its adequacy and desirability relative to alternatives. In this sense, any concept that has undergone conceptual engineering will bear the imprint of the values of the conceptual engineers and of the broader socio-technical landscape in which conceptual engineers operate. This underscores how value-ladenness is not an incidental feature of (re-)engineered concepts, but a structural property of the conceptual engineering process itself.

3.3.3 Insight III: Conceptual engineering would benefit from greater systematicity

The third area of untapped potential for conceptual engineering is methodologically grounded: both function-first and design-first approaches in the philosophy of technology are frequently operationalised through structured methodologies. This section argues that more attention should be placed on developing systematic frameworks in conceptual engineering, akin to those emerging in the philosophy of technology. Indeed, one notable feature of much of the philosophy of technology scholarship is its commitment to developing structured methodologies for analysing, assessing, and guiding the evaluation, design, development, and deployment of artefacts. The philosophy of technology literature offers a sustained body of work demonstrating how systematic approaches can be fruitfully integrated into philosophical inquiry and used to support decision-making in contexts where technological change has significant practical and societal

consequences. In short, philosophers of technology have been, and continue to be, prolific in their methodological output.

These methodologies are not monolithic: they differ in their theoretical commitments, intended scope, and level of prescriptiveness. Some methodological frameworks are diagnostic, such as models for identifying ethical risks or mapping stakeholder perspectives. For example, consider approaches to ethical foresight analysis, which are pursued to forecast and investigate ethical concerns at all stages of technology development through anticipatory techniques (Boenink et al, 2010; Brey, 2012; Grunwald, 2000; Johnson, 2007; Lucivero, 2016; Turoff & Linstone, 1975). Such approaches, which abound in the philosophy of technology literature, are useful in at least two respects: '(1) to analyse the ethical, political, or societal effects of a current or emerging technological artefact; and (2) to identify actionable ways to mitigate any risks of harm an artefact poses (e.g. through changes to product features, research design, or the cancellation of a project altogether)' (Floridi & Strait, 2021, p. 80).

Other methodologies are prospective and incorporate value considerations early in the design process to better align artefacts with societal priorities. Examples of such approaches are value-sensitive design (Friedman, 1996; Friedman & Kahn Jr, 2007)—which focuses on incorporating ethical values and considerations into the design and development of artefacts, emphasising the need to engage stakeholders in the design process to identify and address ethical concerns—and anticipatory technology ethics (Brey, 2012; Johnson, 2007), which uses 'the design phase to reflect upon how a system or technology's affordances will impact their use and potential consequences' (Shilton, 2015, p. 1).

Still other approaches are reflexive, encouraging iterative reassessment of assumptions, design choices, and contextual fit over time. Examples include approaches in post-phenomenology such as technological mediation (Ihde, 1990; Tromp et al, 2011; Verbeek, 2007), as well as actor-network theory (Latour, 1987, 2007) and the social construction of technology (Pinch & Bijker, 1984), which examine the dynamic interplay between artefacts, users, and the socio-technical landscape, encouraging ongoing study and (re-)assessment of choices related to the design, development, and deployment of artefacts. While different, all such approaches are united in their effort to make the design, development, and deployment of artefacts more transparent and attentive to multiple dimensions of evaluation.

The relevance of this for conceptual engineering is straightforward. Much of the work in conceptual engineering is conducted either at the foundational level, i.e., clarifying the aims, scope, or feasibility of conceptual engineering (Belleri, 2024; Isaac, 2020, 2023; Koch, 2021; Prinzing, 2018), or at the level of individual projects focussed on specific concepts (Haslanger, 2000; Jenkins, 2016; Scharp, 2013; Simion, 2018). What is less common is the development of meso-level, systematic tools that can guide practitioners, e.g., in identifying when a concept might be in need of intervention, diagnosing the type of intervention required, or evaluating the adequacy of proposed revisions (Queloz, 2019; Thomasson, 2020). Drawing on the philosophy of technology's emphasis on systematicity could help fill this gap, not by importing any one methodology wholesale, but by adapting the underlying commitment to structured approaches to the evaluation, design, and implementation of concepts.

Some promising moves in this direction already exist. For example, Veluwenkamp & van den Hoven (2023) make the connection explicit between philosophy of technology methodologies and conceptual engineering, and propose to mutually integrate insights from responsible innovation frameworks and conceptual engineering. Moreover, there seems to be a tendency in the scholarship at the intersection of philosophy of technology and conceptual engineering to engage in research that is more systematic in nature.⁴⁷ The literature on conceptual disruption exemplifies how greater systematicity can be embedded in conceptual engineering.⁴⁸ The notion and phenomenon of conceptual disruption, first introduced by Löhr (2022), has been investigated and developed through a combination of case studies and frameworks aimed at articulating typologies of conceptual disruption—distinguishing, for example, between conceptual gaps, conflicts, overlaps, misalignments, and appropriations (Löhr, 2022, 2023a, 2023b, 2023c; Hopster & Löhr, 2023; Hopster et al, 2023a, 2023b; Hopster, 2024; Marchiori & Scharp, 2024)—and then connecting such types to suitable conceptual

⁴⁷ Here, it should be noted that conceptual engineering in the philosophy of technology is a nascent niche at the intersection of the two disciplines and, as such, any trends and generalisations inferred from such a limited sample are to be taken as highly speculative.

⁴⁸ Conceptual disruption refers to the phenomenon of concepts no longer being able to adequately fulfil their functions, often (but not limited to) in response to technological innovation.

engineering responses (Hopster & Löhr, 2023; Hopster et al, 2023a; Marchiori & Scharp, 2024).

With regard to conceptual disruption, a further promising export from the philosophy of technology scholarship to conceptual engineering has been the move from reactive to proactive approaches to conceptual engineering. This can take on two dimensions. On the one hand, as discussed in the previous sections, concepts are best understood as embedded in their context of application rather than in a vacuum. The shift from reactive to anticipatory approaches is well-documented in the philosophy of technology. The key insight is that, by the time an artefact is deployed and becomes increasingly more entrenched, the window for meaningful revision often reduces substantially. A similar logic applies to concepts. Concepts, like artefacts, can become entrenched. Consequently, it is reasonable to expect that, the more a concept is entrenched in a given context—e.g., policy, scientific and academic practice, or everyday discourse—the more the costs and efficacy of revising it or abandoning it might progressively increase. On the other hand, approaches are increasingly being developed to anticipate and (if deemed desirable) mitigate or prevent conceptual disruptions with conceptual engineering, e.g., by focusing on how different concepts may be more or less prone to disruption (Marchiori & Sta. Maria, 2026) and designing concepts to resist disruption (Lundgren, 2024).

Taken together, the three insights developed in this paper point towards the need for the development of a comprehensive approach to responsible conceptual engineering (an *ethics of conceptual engineering*) modelled on analogous approaches in technology ethics and responsible innovation.⁴⁹ Furthermore, they suggest that comprehensive mapping of concepts' functional and affordance profile, the specification of the context of use as well as of the users and other stakeholders involved, and the mapping of the values embedded in concepts, should feature among the dimensions that should guide responsible conceptual engineering projects. Such intuitions may be productively operationalised into a structured approach guiding conceptual engineering practitioners from the suspicion of a conceptual disruption

⁴⁹ An ethics of conceptual engineering could draw on approaches including ethical technology assessment (Palm & Hansson, 2006), responsible research and innovation (von Schomberg, 2013), value-sensitive design (Friedman, 1996; Friedman et al, 2002), design ethics (Verbeek, 2006, 2008, 2011), and ethics by design (Dignum et al, 2018).

through to evaluation and, if needed, design, implementation, and maintenance.⁵⁰

These examples illustrate the potential benefits of more systematic approaches: they help practitioners see patterns across cases, identify recurring challenges, and more easily operationalise theoretical insights. Importantly, embracing systematicity need not mean imposing rigid procedures or narrowing the scope of conceptual engineering to fit a single model. Rather, it recognises that it is advantageous to have common reference points, diagnostic tools, and evaluative criteria to perform conceptual work, perhaps especially when embedded in value-laden socio-technical contexts.

3.4 Conclusion

In conceptual engineering, concepts are often *de facto* considered artefact-like. However, few scholars have explored this intuition further by engaging directly with the philosophy of technology literature on (socio-technical) artefacts. In this paper, I argued that this is a missed opportunity. Indeed, by establishing a closer connection between its study of concepts and the study of socio-technical artefacts in the philosophy of technology, conceptual engineering can be enriched by integrating valuable insights from the philosophy of technology scholarship. After introducing an account of concepts as akin to socio-technical artefacts and clarifying how such an approach sets itself apart from Thomasson's work on concepts as abstract cultural artefacts (Thomasson, 1999, 2007, 2021), I illustrated three insights from the philosophy of technology that merit further investigation within conceptual engineering.

First, conceptual engineering should not limit itself to the study of the functions of concepts (i.e., what concepts do or should do), but would benefit greatly from the investigation of the affordances and side effects of concepts (i.e., what concepts enable or constrain), which would enhanced conceptual evaluation and foster anticipatory conceptual engineering approaches. Second, conceptual engineering should consider

⁵⁰ Developing and operationalising such an approach will be the object of future work. Such an approach could draw on methodological resources developed in ethical foresight analysis and anticipatory technology ethics (Brey, 2012; Floridi & Strait, 2021), stakeholder analysis and constructive technology assessment (Schot & Rip, 1997), and value-sensitive design (Friedman & Kahn Jr, 2007).

(re-)engineered concepts as value-laden. That is, concepts reflect (some of) the values of their designers and of the broader socio-techno-cultural landscape in which the conceptual engineering process takes place. Third, acknowledging that the investigation of artefacts in the philosophy of technology is often operationalised through structured methodologies, I argued that conceptual engineering would benefit from exploring similar systematic approaches to the evaluation, design, and implementation of concepts.

In doing so, the paper moved from an initial analogy between concepts and socio-technical artefacts to an illustration of what that analogy yields for conceptual engineering. Together, these three insights suggest that exploring a closer alignment with the philosophy of technology scholarship on socio-technical artefacts can both refine our theoretical understanding of concepts as artefact-like and offer practical directions for future conceptual engineering work, by broadening and enriching the toolkit of conceptual engineering and illuminating new and promising directions for research and practice. Here, I make no claim to exhaustiveness. It is reasonable to expect that additional sources of untapped potential will arise from the further exploration of the connection between the conceptual engineering and the philosophy of technology literatures.

Chapter 4

CONCEPTUAL AFFORDANCES. (HOW) SHOULD THEY INFORM CONCEPTUAL ENGINEERING?⁵¹

4.1 Introduction — 4.2 From conceptual functions to conceptual affordances — 4.2.1 Affordances — 4.2.1.1 Types — 4.2.1.2 Degrees — 4.2.1.3 Dynamism and multi-stability — 4.2.2 In which sense do concepts enable or constrict our actions and thoughts? Concepts, language, and society — 4.3 Conceptual affordances-informed functionalism — 4.3.1 Functions (still) matter. Why affordances should not replace functions — 4.4 Conclusion

⁵¹ This chapter has been published as Marchiori, S. (2025). Conceptual affordances: (How) should they inform conceptual engineering?. *Synthese*, 206(2), 88.

4.1 Introduction

Conceptual engineering is a normative approach to conceptual work. The purpose of conceptual engineering is the improvement of our concepts⁵² by means of conceptual evaluation, design, and implementation

⁵² Here, I do not adopt a substantive notion of concept. There is significant disagreement in the literature about what concepts are—provided there are any such things as concepts, and that concepts should be the target of conceptual engineering in the first place (e.g., Machery, 2009; Cappelen, 2018; Burgess, Cappelen, & Plunkett, 2020; Koch, 2021; Isaac, 2023). Rather than attempting to settle this dispute here, this paper adopts a methodologically ecumenical approach (I take this apt description from Kölbel, 2023) and maintains a broad understanding of concepts as functional kinds (Lalumera, 2009). I acknowledge that this is an incredibly broad notion. The choice to work with a broad notion of concepts as functional kinds is deliberate: I take this to be both consistent with current practice in conceptual engineering, where commitment to substantive notions of concepts is often left open (Chalmers, 2020; Haslanger, 2020), and a theoretically careful response to the challenge of overcommitment in contested philosophical terrain. On the one hand, competing views on the metaphysics of concepts are each supported by divergent explanatory goals and background commitments. Committing to any of such positions would embroil the paper in disputes that are orthogonal to its primary aim: to argue for the value of conceptual affordances in informing and guiding conceptual engineering projects. On the other hand, this paper's central argument does not depend on any particular theory of concepts, as the notion of conceptual affordance can be operationalised across a range of competing views. That said, Sections 2.1.1, 2.1.2, and 2.1.3 illustrate through examples how the affordance-theoretical approach can be embedded within competing views on concepts. This is meant to show that the notion of conceptual affordance applies across different theories, and to support the claim that affordances are a broadly applicable and theoretically fruitful tool for conceptual engineering.

(Cappelen & Plunkett, 2020; Chalmers, 2020).⁵³ Improving concepts presupposes a standard of what an *adequate* concept amounts to. Functionalism has been proposed as an approach that can act as a normative compass in this respect. According to functionalism, concepts have functions, and the *point* of concepts, what concepts are *for*, is the fulfilment of such functions. Relatedly, concepts are adequate to the extent that they fulfil their functions.⁵⁴ With some notable exceptions (Cappelen, 2018), conceptual engineers have often adopted some version

⁵³ While I follow much of the literature in treating concepts as the target of evaluation in conceptual engineering, I acknowledge that this is not a universally shared assumption (e.g., Haslanger, 2000; Burgess, Cappelen, & Plunkett, 2020; Koch, 2021; Isaac, 2023). Indeed, some scholars have cast doubt on the coherence and utility of understanding concepts as the targets of conceptual engineering (e.g., Cappelen, 2018; Deutsch, 2020). I do not attempt to resolve these debates here, where I adopt a pragmatic approach to the issue: I assume, for the sake of argument and illustration, that there are such things as concepts and that concepts are the target of conceptual engineering. It should be noted that this is a methodological choice, not a metaphysical commitment—indeed, the affordance-theoretical approach advanced in this paper can in principle be extended to other targets of conceptual engineering (such as words, e.g., Thomasson, 2021), to the extent that they enable or constrain possibilities for action in a specific context and with regard to specific users.

⁵⁴ It should be noted that, to be able to meaningfully guide conceptual engineering projects, it is insufficient to argue that concepts have functions and that fulfilling its function(s) makes a concept adequate. Rather, functionalism must be able to address two additional questions. That is, when conceptual engineers claim that concepts have functions and are adequate to the extent that they fulfil such functions, which functions are they referring to, and what does it mean for a concept to fulfil its function(s)? For the purpose of this paper, I will leave such issues aside. It suffices to know that the research on functionalism lies in what can be described as a pre-paradigmatic state, but some tentative hypotheses have been advanced with regard to both questions. On the one hand, the conceptual engineering scholarship has recently started to move towards a deflationary account of conceptual functions (Nado, 2021; Riggs, 2021; Jorem, 2022; Hopster & Löhr, 2023), according to which which function(s) a given concept should fulfil is ultimately context-dependent and should follow a case-by-case determination. On the other hand, it seems reasonable to expect that concepts will admit of degrees of functional fulfilment and, consequently, conceptual adequacy. That is, it may not be the case that all concept will be adequate or inadequate in absolute terms. Rather, functions may be fulfilled in gradients, rather than in binary terms. This is consistent with Nado's (2021) position.

of functionalism (e.g., Haslanger, 2000; Plunkett & Sandell, 2013; Prinzing, 2018; Queloz, 2019; Chalmers, 2020; Haslanger, 2020a, 2020b; Simion & Kelp, 2020; Sundell, 2020; Thomasson, 2020; Belleri, 2021; Miranda Vilchis, 2022; Hopster & Löhr, 2023; Marchiori & Scharp, 2024). Functionalism traditionally operates on the assumption that conceptual engineers ought to concern themselves with what concepts *should* do. In practice, this has sometimes restricted the focus of conceptual engineers to the *intended* functions of concepts, thus systematically neglecting what concepts *could* do, what they *enable* us to do. Recent proposals (e.g., Nado, 2021; Riggs, 2021; Jorem, 2022; Queloz, 2022; Köhler & Veluwenkamp, 2024) have enriched this view by articulating a more context-sensitive and dynamic approach to functionalism.

Building on such developments, this paper introduces the notion of conceptual affordances as a complementary framework for conceptual engineering. Conceptual affordances refer to the potential actions, uses, and thoughts that a concept enables or constrains relative to a user and a given context. A coarse-grained notion akin to affordances has already emerged organically in the conceptual engineering literature bearing other names, e.g., under the heading of *possible functions* (in Nado, 2021). However, as this paper illustrates, a more systematic and robust account can be developed by drawing on the notion of affordance as introduced in ecological psychology and later elaborated in the philosophy of technology scholarship, which engages in similar pursuits as conceptual engineering, with similar normative aims.

Attending to the affordances of concepts allows conceptual engineers to identify not only the fulfilment of intended conceptual functions but also the range of uses that concepts facilitate, including unintended or unforeseen uses and effects of conceptual deployment. Understanding conceptual functions through the lenses provided by the notion of affordance solidifies the move towards an understanding of concepts as tools whose uses may extend beyond their original design, often in ways that are contextually bound and user-dependent (Köhler & Veluwenkamp, 2024). This shifts the focus from what a concept *should* do to what a concept *enables* one to do, thus offering a richer understanding of the roles that a given concept might perform.

This paper argues that affordances can supplement and enrich functionalism in a way that is both theoretically fruitful and ethically necessary. By considering the affordances of concepts, conceptual engineers can better analyse and anticipate both intended and unintended consequences of conceptual use, which is crucial for guiding

conceptual engineering projects effectively. Additionally, in cases where multiple competing concepts meet standards of functional adequacy, attending to their respective affordances offers an additional normative criterion for conceptual selection, allowing for more context-sensitive and ethically responsive choices in conceptual engineering projects.

The paper is structured as follows. Section §4.2 introduces the notion of affordance and discusses its relevance to the functional approach in conceptual engineering. Section §4.3 expands on this idea, illustrating how the integration of conceptual affordances into functionalism can provide a more holistic understanding of conceptual adequacy. Section §4.4 contains the conclusion.

4.2 From conceptual functions to conceptual affordances

Functionalism in conceptual engineering traditionally focusses on what concepts are *supposed* to do. This perspective emphasises assessing concepts primarily by reference to the roles they are intended to fulfil (their proper functions⁵⁵)—or the roles we wish them to fulfil—while often overlooking a broader analysis of what concepts might enable beyond those these functions. However, recent proposals have increasingly highlighted the significance of conceptual functions beyond proper functions, challenging the conventional focus and urging a more comprehensive, context-sensitive approach to conceptual engineering.

⁵⁵ The notion of a *proper* function refers to the function or role that a concept is supposed to serve due to its history of selection, whether in biology, language, or social practice, often grounded in either the original intentions behind the concept or its role in a system of practices or institutions. This idea draws from theories of function in biology and philosophy of language, particularly teleological accounts of function (Millikan, 1984). In this sense, a concept's proper function can be understood as its historically established purpose in fulfilling a specific epistemic, moral, or practical role. Other terms used in the literature include *intended function* and *system function*, which capture slightly different nuances. The *intended* function of a concept typically refers to the purpose a concept was explicitly designed or introduced to fulfil, often associated with the intentions of its creators or early users (Cappelen & Plunkett, 2020). By contrast, a concept's *system* function reflects the role a concept plays within a broader system, regardless of the intentions behind its creation, thus emphasising the functional interdependence between concepts within institutional or linguistic frameworks (Haslanger, 2012). Here, I will use intended, proper, and system functions interchangeably to mean the purpose that a concept is supposed to serve.

For example, Nado (2021) provides a critical perspective on the limitations of focusing solely on proper functions, which she deems ill-advised, arguing instead that other, user-dependent functions should also be of interest to conceptual engineers beyond intended functions. In a similar vein, one may interpret scholars arguing for the assessment of conceptual adequacy based on a case-by-case determination of the relevant function(s) that a given concept should serve within a given context as adhering to a similar position to Nado's (Riggs, 2021; Jorem, 2022). Again, Queloz (2022) highlighted the need for functional accounts to attend to the dynamic interplay between conceptual use and social practice. More recently, Köhler & Veluwenkamp (2024) advocated evaluating concepts by reference to what matters (to us) in a given context, drawing on the notion of *normative function*.

Here, I build on such insights, which align with an affordance-theoretical approach, and suggest that conceptual engineers should draw on conceptual affordances for the assessment of conceptual adequacy. The following sections introduce the notion of affordance, articulate its features, and illustrate how affordances can be applied to the conceptual domain. I begin by discussing the key features of affordances, which will serve as building blocks for integrating this notion into conceptual engineering.

4.2.1 Affordances

The notion of affordance picks out the range of possible actions that an artefact offers to an agent within a given environment (Gibson, 1977a, 1977b, 1979; Gibson, 1982). This notion captures the ways in which artefacts enable or constrain actions, depending not only on their intrinsic features, but also on the characteristics of the environment and the

features of the user.⁵⁶ Affordances thus encompass both intended and unintended uses of the artefact, and emerge from the dynamic interaction between artefact, user, and context.

Being mindful of the uses that artefacts allow—by virtue of a combination of their features, the agents involved in their use, and the context of use—broadens the scope of inquiry to include unintended (yet possible) uses that artefacts or concepts afford. With this in mind, consider the following description of the roles of towels in *The Hitchhiker's Guide to the Galaxy*.

A towel [is] about the most massively useful thing an interstellar hitchhiker can have. *Partly it has great practical value*—you can wrap it around you for warmth[,] you can lie on it on the [beach,] you can sleep under it[,] use it to sail a mini raft[,] wet it for use in hand-to-hand-combat[,] wrap it round your head to ward off noxious fumes [,] you can wave your towel in emergencies as a distress signal, and of course dry yourself off with it if it still seems to be clean enough. *More importantly, a towel has immense psychological value.* [If a non-hitchhiker] discovers that a hitchhiker has his towel with him, he will [happily] lend the hitchhiker [any items] that the hitchhiker might have accidentally “lost”[, thinking] that any man that can hitch the length and breadth of the Galaxy [and] still know

⁵⁶ It should be clarified that the notion of affordance presented in this paper reflects a weak, perception-neutral account of affordances, that is, a view that does not engage with the question of whether affordances must be perceived to exist. This more relaxed view of affordances sets aside meaningful disagreements about the extent to which affordances are fundamentally tied to an agent's perception, as in Gibson's (1977a, 1997b, 1979) original formulation, where perception is crucial to understanding what an artefact (or concept) affords (e.g., McGrenere & Ho, 2000; Chemero, 2003; Chemero, Klein & Cordeiro, 2003; Klenk, 2021). This weak account is employed deliberately, as it allows for a broader application to conceptual engineering without introducing unnecessary complexity. Indeed, my aim here is not to do justice to the richness and depth of the scholarly debates on affordances, nor to engage in such debates, but to provide a more comprehensive framework for understanding how concepts can enable or constrain actions, whether or not such possibilities for actions are actively perceived by the user. Hence, for the purposes of this paper, a weaker notion of affordance suffices to illuminate how concepts can invite or enable certain possibilities for conceptual use, regardless of whether these are consciously perceived by users.

where his towel is, is clearly a man to be reckoned with.
(Adams, 1979, p. 19, emphasis added)

What makes towels so useful for hitchhikers relates in large part to the alternative uses that towels have beyond their traditional role as absorbent pieces of fabric used for drying things, people, or animals. In fact, if non-hitchhikers were aware of all the uses of towels described in the *Guide*, it would not be nearly as noteworthy or remarkable for a hitchhiker to carry a towel, which would undercut the very reasons why towels are so useful for hitchhikers. Rather, these uses, which towels afford, are highly dependent on a combination of the artefact (the towel), the context (intergalactic travel), and the agent involved (a hitchhiker). In the example at hand, the unintended affordances of towels seem to be overall desirable.

However, affordances are a double-edged sword: they enable or constrain our actions well beyond the confines of the intended functions for which they were designed. This means that, by bearing affordances, artefacts will always be able to do more than what we want them to do. Indeed, what happens when designers can predict an affordance, which they deem undesirable, and wish to restrict the possibilities for action such that the artefact will cease to bear such an affordance? Consider the following extract from *The Restaurant at the End of the Universe*.

The designer of the gun had clearly not been instructed to beat about the bush. 'Make it evil' he'd been told. 'Make it totally clear that this gun has a right end and a wrong end. Make it totally clear to anyone standing at the wrong end that things are going badly for them. If that means sticking all sorts of spikes and prongs and blackened bits all over it then so be it. *This is not a gun for hanging over the fireplace or sticking in the umbrella stand, it is a gun for going out and making people miserable with.*' (Adams, 1995, p. 172, emphasis added)

Crucially, unintended, albeit predicted, affordances cannot be eliminated just because one is mindful of them and deems them undesirable. Affordances are, in some salient respect, both unpredictable and beyond the control of the designer. At the same time, it is important to stress that this does not make being mindful of affordances a futile exercise. Indeed, firstly, it is important to emphasise that affordances, while not entirely under the control of the designer, are not *wholly beyond* the designer's influence. That is, even though some affordances might elude designers'

control, e.g., such that they could not fully dictate or eliminate them, designers can still meaningfully influence key dimensions of how an artefact will likely function in specific contexts, helping steer or limit some of the more (un)desirable outcomes. Secondly and relatedly, just because designers cannot have full control over conceptual affordances, this does not undermine the utility of investigating affordances to anticipate a broader range of possible uses of artefacts (beyond their intended functions), which can meaningfully inform design projects. In short, while retaining absolute control over all affordances is unrealistic, this does not mean that any efforts to pre-emptively address, accommodate, or mitigate affordances, whenever possible, will be futile and should not be encouraged.

Having outlined the notion of affordances with regard to artefacts, I suggest that a similar affordance-theoretic perspective can be meaningfully extended to the conceptual domain. Just as artefacts afford possibilities for action depending on their features, users, and contexts, I propose that concepts afford possibilities for action, use, or thought, depending on the interplay among concepts, users, and the contexts of use. In what follows, I explicate and justify this extension by examining how affordances manifest in the conceptual realm.

A step-by-step clarification of this move is in order. If we accept the assumption (widely shared in the conceptual engineering literature) that concepts have functions, it seems reasonable to further assume that conceptual functions are not limited to the functions that concepts *should* serve. Just as artefacts may serve multiple purposes beyond their intended function, concepts may exhibit a broader range of functions than what they are specifically intended to accomplish. While intended functions refer to what concepts *should* do, the broader range of conceptual functions captures what concepts *could* do. At this point, the move from functions (broadly understood) to affordances becomes apparent. Affordances capture what a concept enables us to do, i.e., the potential actions and uses that emerge from the combination of a concept

in a given context and with regard to a given user.⁵⁷ Consequently, if conceptual engineers accept that concepts have functions broadly understood (including both intended and unintended functions), they should plausibly also accept that concepts have affordances—just like artefacts do.

Ultimately, understanding conceptual functions through the lenses of the notion of affordances can provide more solid and comprehensive guidance for conceptual engineering projects, by enabling conceptual engineers to gain a more rounded understanding of the arrays of conceptual uses that are (and should be) relevant for conceptual engineering purposes. This, in turn, can enhance and guide conceptual evaluation, (re-)design, and implementation.

Whether artefactual or conceptual, affordances exhibit three salient features: they are typed (as invitations for use or mere possibilities for use), graded (as stronger or weaker), as well as dynamic and multi-stable (shifting across contexts, users, and times). The following sections

⁵⁷ Here, it is relevant to note that functions and affordances can be understood in terms of each other. On the one hand, affordances could be framed through the lens of functions. In this view, affordances would represent the full range of possible functions a concept has, including but not limited to its intended functions. On the other hand, one could understand functions through the lens of affordances. In this sense, functions would be interpreted as the actions that a concept enables in a given context and with respect to a given user, encompassing not just what the concept *should* enable, but the full range of what it *can* enable in different contexts. Here, I deliberately skip the intermediate step of considering possible functions as a helpful bridge between intended functions and affordances, as this would introduce unnecessary complexity. Specifically, distinguishing between possible functions and affordances is both unnecessary and undesirable. It is unnecessary as there is little benefit to stretching the notion of function when the notion of affordance already captures the full range of what concepts *enable* us to do (their possible functions). Furthermore, it is undesirable as over-inflating the notion of function to collapse onto that of affordance would risk conflating the two concepts and generate confusion. This point will be expanded upon in section §4.3.1.

examine each dimension in turn, first for artefacts, then for concepts.⁵⁸ While the discussion remains neutral with respect to competing theories of concepts, I illustrate how affordances are instantiated across three leading theories of concepts—concepts as mental representations, as inferences, and as prototypes—, thereby underscoring the meta-theoretical neutrality and operational flexibility of the affordance-theoretic approach.

4.2.1.1 *Types*

First, affordances can be differentiated based on their type—that is, whether they are intended invitations for use or (non-)canonical possibilities for use. Affordances are often described in terms of possibilities for action (or use; Gaver, 1991; Glăveanu, 2020; Borghi, 2021; Tollon, 2022; Koutamanis, 2023). These possibilities can manifest in different ways depending on the features of the artefact (or concept), the context of use, and the user. Here, I find it helpful to further qualify affordances by distinguishing between intended and unintended affordances, depending on the involvement of the designer (or lack

⁵⁸ It should be noted that the notion of affordance first originated to shed light onto the possibilities for action offered by artefacts. In this paper, I will focus on conceptual affordances. The move from artefacts from concepts is justified to the extent that concepts are oftentimes taken to be artifactual in some salient respect (Thomasson, 2021; Margolis & Laurence, 2023; Veluwenkamp & van den Hoven, 2023). In this sense, I do not wish to argue that concepts are (socio-technical) artefacts, but merely to adopt this well-established analogy as a heuristic device to illustrate how affordances can inform conceptual engineering (e.g., Chalmers, 2020; Köhler & Veluwenkamp, 2024). For a more in-depth illustration of the similarities between artefacts and concepts that lays the groundwork for such a comparison, see Marchiori (2026).

thereof).⁵⁹ Therefore, one may understand affordances broadly as possibilities for use, which can be further classified as invitations for use (specifically intended uses) and mere possibilities for use (general potential uses).

Affordances understood as *invitations for use* refer to the potential actions that an artefact (or concept) explicitly supports or enables due to the designer's intentions. For example, a wooden chair may afford sitting because its design and structure make it suitable for that purpose. This is an intended affordance, where the artefact's design explicitly invites a particular use. Similarly, the concept CONCEPTUAL ENGINEERING invites a normative approach to conceptual work, e.g., one that encourages tinkering with concepts as one would do artefacts (Chalmers, 2020). However, affordances need not align with the designer's intentions.

Affordances understood as *mere possibilities for use* refer to the general potential actions, uses, and thoughts that a given artefact (or concept) enables. These are the inherent possibilities that arise from the interaction among a given artefact's (or concept's) features, the context of use, and the user. For example, a chair can be used to stand on to reach a high shelf, even if it was not specifically designed for that purpose. However, this depends on a combination of the chair's features (e.g., a small plastic chair may be not well-suited for an adult to step on), the context (e.g., to use the chair, the floor should be stable and free of obstructions), and the user (e.g., a person with impaired mobility may not perceive the chair as at all helpful to reach the shelf).

⁵⁹ This reflects a similar distinction made in the literature between intentionally designed affordances and unintended affordances. By way of example, Norman (1988, revised 2013; 1999) distinguishes between affordances that are intentionally designed and those that are not. Similarly, Shneiderman (1992) touches on the difference between intended and unintended affordances in the context of user interface design, while discussing how designers often build user interfaces that afford certain actions, but users may discover alternative ways of interacting that were not intended by the designers. More recently, Bennett & Bennett (2012) tackle the impact of unintended affordances in user experience, understood in terms of products or interfaces being used in ways that the designers did not envision, thus suggesting a distinction between the two.

Likewise, the concept DISORDER can afford drawing different inferences depending on the context.⁶⁰ In a clinical setting, DISORDER may afford drawing specific inferences, e.g., related to the need for diagnosis and the possibility of treatment. These inferences are shaped by the medical system, the expertise of practitioners, and the tools at their disposal, such as diagnostic tests and therapeutic protocols. In contrast, in a social context, DISORDER may prompt different inferences, such as associations with stigma, exclusion, or deviance. These inferences are influenced by the features of the social environment, such as prevailing negative attitudes toward mental health or other types of disorder, leading to marginalisation or exclusion.

One can further distinguish affordances intended as mere possibilities for use between canonical and non-canonical affordances, i.e., expected and unexpected possibilities for action, respectively (Arnold et al, 2023). Canonical affordances refer to the possibilities for action that, while not explicitly intended by the designer, are nonetheless expected or foreseeable based on the features of the artefact (or concept) and the conventional contexts in which it is employed. For example, although a chair is designed primarily for sitting, using the chair to stand on to reach a high shelf is an expectable, albeit unintended, use of the chair. Similarly, the concept SALAD affords thinking of dishes such as *pasta salads* or *fruit salads*. While the prototype of SALAD centrally features a cold mixture of small, edible components—typically vegetables, traditionally leafy greens —, the salient properties of mixture, cool serving temperature, and combinatory preparation make it expectable that other ingredients, such

⁶⁰ The inferentialist view of concepts understands concepts as inferential roles, e.g., nodes within a web of normative commitments and inferential roles (Sellars, 1953; Brandom, 2009). On this view, to possess a concept is to grasp the inferences that it licenses and what follows from its application within a given discourse. Concepts are thus defined by the inferences that they allow, the social roles they instantiate, and the actions they justify or prohibit. For example, the concept book is not defined merely by its physical properties (e.g., having pages or being readable), but by its place within a network of inferences (e.g., being cited in academic work, being subject to copyright law) and the social practices that those inferences enable.

as pasta or fruit, can be substituted while retaining categorisation as a salad.⁶¹

By contrast, non-canonical affordances arise in ways that are neither intended nor expected. For example, the same chair may afford shelter from the sun with its shadow, it may be used to prop open a door, or it may become part of as a sculpture in an art installation, such as in Joan Miró's *Seated Woman and Child*. The parallel in the conceptual realm may be the following. Under a definitional view of concepts, SOUP would typically be understood through necessary and sufficient conditions such as involving a cooked liquid base combined with edible ingredients. However, in informal settings, such as social media and playful legal or philosophical debates, users may extend the concept SOUP in unexpected ways. For instance, by construing ocean water as broth and seaweed as vegetables, by pointing out the presence of fish, and by treating geothermal and solar processes as forms of cooking, one may extend the concept SOUP to the ocean. This reasoning recontextualises the defining

⁶¹ The prototype theory of concepts holds that concepts are organised around cognitively salient examples or typical feature clusters, rather than being strictly defined by necessary and sufficient conditions. On this view, to possess a concept is to recognise its central and peripheral instances based on similarity to an idealised prototype (Rosch & Mervis, 1975; Lakoff, 1987). For instance, the concept BOOK is prototypically linked to objects with a cover, printed pages, and textual content meant for reading. Variants like audiobooks or eBooks may be considered less typical members of the category. Conceptual affordances—what a concept invites or allows users to think and do—are shaped by these prototypes: actions and expectations are anchored in the features most salient to the prototype and may shift as prototype structures evolve through cultural, technological, or experiential change.

features of SOUP to fit novel cases, generating a non-canonical affordance of the concept.⁶²

4.2.1.2 *Degrees*

Second, affordances can differ in degree, depending on how strongly or weakly they invite or enable specific actions, uses, or thoughts. That is, artefacts (and concepts) can bear higher and lower affordances (Maier & Fadel, 2007). The more intuitive and automatic the use of an artefact (or concept), the higher the affordance. Once again, the determination of the strength of an artefact's (or concept's) affordances depends on the interaction among the artefact (or concept), the context in which it is used, and the person using it.

Therefore, as illustrated in the Hitchhiker's example in Section §4.2.1, one may argue that towels bear higher affordances with regard to their practical uses and lower affordances for their psychological uses. Indeed, the very reason why towels have immense psychological value for interstellar hitchhikers is precisely because they enable specific uses, which are not intuitive, but rather specific to interstellar hitchhikers.⁶³ If non-hitchhikers were to intuitively capture towels' possibilities for use as means to be perceived as being worthy of respect, regardless of one's actual merits, towels would bear higher affordances for their psychological function for hitchhikers and non-hitchhikers alike and, quite ironically, lose that very function for interstellar hitchhikers.

⁶² The definitional view of concepts holds that concepts are structured by a set of necessary and sufficient conditions that jointly determine category membership. On this view, to possess a concept is to know, explicitly or implicitly, the defining features that an object or instance must have to fall under the concept. For example, the concept BACHELOR is traditionally analysed as requiring that the person be male, adult, and unmarried. Conceptual competence, in this framework, consists in correctly applying the concept on the basis of these definitional features. Conceptual affordances thus derive from the logical structure of these definitions. Novel or non-standard applications emerge when definitional features are reinterpreted, stretched, or applied under atypical conditions, often revealing tensions or ambiguities within the definitional schema itself.

⁶³ This is what Gaver (1991) would describe as a *hidden* affordance. That is, despite an artefact bearing a particular possibility for action, such an affordance is not perceived by the user.

Analogously, the concept BOOK helps individuals identify certain objects as suitable for reading, annotating, and knowledge acquisition. However, for users primarily familiar with digital media, some of the traditional affordances of BOOK may be weaker, such as expectations tied to physical interactions with printed volumes, e.g., the action of flipping physical pages. Conversely, other affordances arising from more peripheral or contested inferential paths may become stronger, such as notions of books as means for interactive learning, or audio narration. Similarly, one might observe differing degrees of affordance strength when considering the concepts FUNCTION and AFFORDANCE themselves within the practice of conceptual engineering. Particularly because of its widespread entrenchment in conceptual engineering discourse, the concept FUNCTION strongly affords its application in evaluating the adequacy of concepts. By contrast, while AFFORDANCE also offers a powerful lens for assessing conceptual adequacy, its affordances in this evaluative role are comparatively weaker. This is not due to any inherent deficiency of the notion, but because AFFORDANCE is less established within the conceptual engineering community. While both concepts can be used for the same purpose, their respective affordances differ in strength, e.g., according to users' disciplinary backgrounds, prevailing theoretical frameworks, and the context of inquiry. Thus, affordances' strength is graded and modulated by users' background and the context of use.

4.2.1.3 Dynamism and multi-stability

Third, affordances are dynamic and multi-stable, shifting across users and contexts. That is, the understanding of affordances as invitations for use and possibilities for use, as well as their degree, may differ vastly depending on the users involved and the context of use. Indeed, different users may perceive the same affordances as more or less intense.

Consider a chef's knife. To a professional chef in a kitchen, the knife, with its sharpness, balance, and handle, invites a range of skilled uses, such as cutting, slicing, dicing, and chopping food. In this context, the affordances for the intended uses of the knife will be strong because the knife's design aligns with the user's expertise. Conversely, in the hands of a child, the very same knife's affordances may shift significantly. While the knife still physically affords cutting, slicing, dicing, and chopping, the child's lack of skill and awareness of the knife's potential

functions leads to lower affordances for its intended uses. Again, while interacting with a smartphone, elderly users may perceive higher affordances for making calls, while perceiving low affordances for sending texts, and not perceiving the smartphone's affordances for social media usage. Conversely, younger users may perceive higher affordances for the use of apps, among which social media and games, while perceiving low affordances for making calls.

Therefore, it is crucial to be mindful of the fact that artefacts (and concepts) can be perceived and used differently by different individuals and across various communities and cultures. In the engineering and design literature, such insights have led to research emphasising the importance of user-centred design in creating intuitive affordances for use, whose focus is on designing objects that clearly communicate their potential interactions to users (e.g., Norman, 1988, 1999).

Context also greatly affects affordances. Consider once again the professional chef and their knife. Were the chef no longer in their kitchen, but stranded in the wilderness, their knife may afford a broader range of uses beyond cutting food: it could be used for carving tools, making fire kindling, or self-defence. The survival context amplifies the salience of these affordances, which would be weaker in the everyday setting of a kitchen, while cloaking the knife's intended possibility for use. Again, in a symbolic context, the chef's knife could afford yet another set of possible actions. For example, in certain ceremonial practices, the knife might not be used for cutting at all but rather as a symbol of power or authority.

A similar pattern can be observed in the conceptual realm with the notion of ARTIFICIAL INTELLIGENCE. For the same user, ARTIFICIAL INTELLIGENCE may afford different possibilities for use and thought depending on the context of reference. In a technical setting, such as a machine learning conference, ARTIFICIAL INTELLIGENCE may strongly afford precise inferential and operational uses, invoking notions such as algorithmic modelling, supervised and unsupervised learning, or computational complexity. Here, the concept's affordances are tightly coupled to specialised, technical meanings and actions. In contrast, in a lay conversational context, the very same user may deploy ARTIFICIAL INTELLIGENCE with affordances related to popular narratives, such as

automation, humanoid robots, or existential risk. The concept's affordances dynamically shift according to the context of use.

These examples illustrate the multi-stability of affordances, where the same artefact (or concept) affords a wide range of possibilities for use, shaped by the interaction among the artefact (or concept), the user, and the context of use. As relational properties, affordances are not static, but dynamic in nature, and evolve as technology and user interactions change. That is, as artefacts (or concepts) and users change or interact with new contexts, artefacts' (or concepts') affordances can also change. For example, the dynamic nature of affordances means that today's (non-)canonical possibilities for use could become invitations for use tomorrow—such as when designers find ways to repurpose unintended affordances into intended ones. Relatedly, this means that (conceptual) engineers must continually reassess the possibilities for use of concepts. In the words of Heras-Escribano (2019), the notion of affordances uncovers “a meaningful world of promises and threats” (2019, p. 1). This has salient implications for the (ethical) design and maintenance of artefacts (and concepts).

One additional consideration is in order. When affordances are acted upon, side-effects may emerge. Consider a smartphone. The design of a smartphone affords several interactions: it can be used for making calls, sending messages, taking photos, browsing the internet, and more. Each of these interactions is a possibility for use that the smartphone offers to its users. The phone's touchscreen, buttons, and apps, are all designed with specific affordances in mind. In contrast, while the design of the smartphone does not invite its users to use it to emit electromagnetic radiation, this is an unavoidable side-effect of its electronic components.

The engineering literature on the unintended consequences of artefacts highlights how side-effect are often discovered post-design and require mitigation strategies. As a result, there is a growing trend to anticipate side-effect and either minimise their negative impacts or transform them into beneficial possibilities. When this occurs, side-effects may themselves become affordances. For example, the electromagnetic radiation emitted by the smartphone can become a designed affordance if it is harnessed for wireless charging or data transmission. In such cases,

what was once a side-effect becomes a possibility for use through intentional design modifications.⁶⁴

Taken together, these three features of affordances—i.e., types, degrees, and dynamism—highlight affordances as a rich, context-sensitive framework for understanding how concepts can invite or constrain actions, uses, and thoughts beyond their intended or desired functions.

4.2.2 In which sense do concepts enable or constrict our actions and thoughts? Concepts, language, and society

In the previous sections, I illustrated how, when users are presented with an artefact, they are invited to use it in particular ways. I argued that the same extends to concepts. Intuitively, one of the salient differences between artefactual affordances and conceptual affordances seems to be that (socio-technical) artefacts are only partially determined by their function and are also determined by their physical properties (the so-called “dual-nature” of technical artefacts, Kroes & Meijers, 2006). As such, artefacts’ physical constraints shape and limit the range of affordances that they bear.

At a first glance, such constraints seem to be largely missing in the case of concepts. Still, as pointed out by Houkes and Vermaas (2010) with regard to technical artefacts, the fact that constraints are not clearly visible does not mean that they are absent. Rather, some features of artefacts may be obscured and need to be unveiled. For example, the physical constraints of digital artefacts may not be missing but merely rendered opaque. A similar mechanism may be at play with concepts. Indeed, while concepts do not seem to have noticeable physical

⁶⁴ It is important to distinguish between unintended affordances and side-effects, as they operate at different levels of analysis. Indeed, while affordances refer to the potential actions or uses that an artefact (or concept) enables in relation to a user and a context, side-effects refer to the actual, often unintended, consequences that result from the mere existence or deployment of an artefact (or concept). Although unexpected affordances may lead to unforeseen side-effects, the two should not be conflated: affordances capture the range of possibilities for (inter)action, whereas side-effects represent the realised outcomes of those interactions, often independent of the user’s knowledge and intentions. For example, in the case of concepts, side effects may occur when the introduction or use of a concept leads to unintended restructuring of behaviours, expectations, or institutions.

constraints, they may nevertheless be constrained in some salient respects. In this section, I briefly expand on the extent to which the linguistic items used to express concepts may be interpreted as constituting one of such constraints. Discussing conceptual affordances with regard to the linguistic medium of concepts picks out an interesting duality.

On the one hand, one could understand the linguistic medium as a necessary vessel for concepts, such that one cannot neatly disentangle concepts from the linguistic items used to convey them. On this view, language may be to concepts what materiality is to artefacts. Therefore, similarly to the material conditions of artefacts, the linguistic items through which concepts are articulated may themselves be interpreted as a constraint that restrict or enable concept users' actions and thoughts.⁶⁵ For example, consider how the term “man” may refer to both the concepts MALE HUMAN BEING and HUMAN BEING, while the term “woman” is typically taken to refer solely to FEMALE HUMAN BEING. One may argue that using the term “man” to convey the concept HUMAN BEING constricts our actions and thoughts in ways that using the term “human being” to convey the same concept does not, e.g., to the extent that it can lead to framing male experiences as universal and the norm, but female or non-binary experiences as distinctly confined to the female and non-binary world, respectively.⁶⁶ In other words, to the extent that concepts and the linguistic items used to express them are inevitably interlinked, focusing on the affordances of a concept detached from the affordances arising from the linguistic medium used to convey it could be problematic, as it would inevitably fall short of identifying at least part of the affordances of the concept at hand.

⁶⁵ The idea that the limits of language are the limits of one's world is famously attributed to Wittgenstein (1922). This intuition is in accordance with studies suggesting that language and its structure affects its speakers' worldview and cognition, by shaping thought processes and influencing the perception of the world (Sapir, 1929; Kay & Kempton, 1984; Winawer et al, 2007; Whorf, 2012).

⁶⁶ Indeed, research on gendered language demonstrates that the use of gendered pronouns and words can reinforce gender stereotypes and influence perceptions of gender roles (Boroditsky, 2006; Stahlberg et al, 2011). In Marçal's (2021) words, “[t]he problem are all the people who are not accorded the same right to be universal and how, in turn, this limits our idea of what it means to be human” (Marçal, 2021). This leads to “men's tools [being] allowed to belong to history while women's tools belong to women's history” (Marçal, 2021).

On the other hand, one could argue that, despite concepts being frequently conveyed through language, the linguistic medium can be safely detached from the concept it expresses. However, even in such a case, attempting to identify the affordances of a concept by looking at the concept in a vacuum, detached from the socio-technical-cultural landscape in which it is embedded, would be misguided. Indeed, if conceptual affordances are influenced and co-shaped by a combination of the concept itself, the user, and the environment, then the affordances that a concept bears cannot be meaningfully identified by detaching the concept from the broader context in which it is situated, which includes (but is not limited to) the means (linguistic or otherwise) through which the concept is conveyed. Indeed, consider the ways in which language often acquires a mediating role between people and their concepts, e.g., when people speaking different languages try to use a shared vocabulary to understand and communicate with each other. Understanding and communicating concepts through the means of language can lead to constraints to the extent that linguistic items may be more or less adequate vessels to capture the concept one wants to convey.⁶⁷

In simpler terms, while it is plausible to expect that some affordances might be purely based on inherent features of the concept (if one understands concepts and language as inextricably connected), at least some affordances will arise from extrinsic vectors depending on the interaction between the concept, the user, and the broader context (including the linguistic items through which concepts are conveyed). Ultimately, it seems that whether language is akin to the material conditions of artefacts is not crucially relevant. Indeed, while there may be a benefit in being able to distinguish whether it is the concept itself that carries certain affordances, or whether it is the way the concept is conveyed through language that introduces such affordances, both hypotheses (language-driven affordances and concept-driven

⁶⁷ For example, language fails us when we wish to translate the Dutch concept of “gezelligheid” (roughly, a person, object, or atmosphere, perceived as cozy, inviting, pleasant, convivial), and constrains us when we try to convey the full richness of the Farsi notion of “kirpan” (a religious object reminiscent of a knife, as in Soltandazeh, 2021), or the weighty lightness of the Italian concept of “menefreghismo” (blatant disregard for one’s own duties or the rights of others).

affordances) seem plausible, and not mutually-exclusive.⁶⁸ As Mills (2007) writes,

[I]f the society is one structured by relations of domination and subordination (as of course most societies in human history have been), then in certain areas this conceptual apparatus is likely going to be shaped and inflected in various ways by the basis of the ruling group(s). So crucial concepts may well be misleading *in their inner makeup and their external relation to a larger doxastic architecture*. (Mills, 2007, pp. 24-25, emphasis added)

This suggests that conceptual affordances may be enabled by a combination of intrinsic and extrinsic factors, both of which can constrict certain forms of thought or action while enabling others in alignment with the interests of dominant groups. Moreover, such an integrated approach seems consistent with Haslanger's (2012) and Congdon's (2023) views of concepts and social practices and normative outlooks as mutually co-constitutive.

Specifically, Haslanger (2012) argues that many of our concepts, especially those related to social categories like gender, race, and class, are socially constructed. These concepts are not merely reflections of natural kinds but are created and maintained through social practices and institutions. She furthermore claims that language and discourse play a critical role in shaping our concepts. That is, the ways in which we talk about and categorise the world influence how we think about it. For example, the concept WOMAN is shaped by societal norms, expectations, and language. Importantly, Haslanger (2012) highlights a feedback loop between concepts and social practices. That is, social practices give rise to and sustain certain concepts, which in turn influence and reinforce those practices. For example, the social practice of gender roles reinforces the concept GENDER, which then perpetuates such roles.

Similarly, Congdon (2023) builds on and extends Haslanger's (2012) insights. Congdon (2023) argues that concepts are not just

⁶⁸ Here, I focus on the role of language in the identification of conceptual affordances. I do not wish to exclude that concepts themselves may, at least in some cases, afford possibilities for action and thought independently from the language through which they are conveyed. Similar questions are baked into the notion of affordance beyond its application in the conceptual domain, and mirror a structural feature that is already present when talking about artefactual affordances.

passively shaped by social practices but are actively constructed through them. For example, evolving norms around gender equality influence the concept GENDER, which in turn affects normative expectations and social practices related to gender roles. This mutual shaping underscores the co-constitutive relationship between concepts, social practices, and normative outlooks.

4.3 Conceptual affordances-informed functionalism

This section proceeds in two steps: first, I argue that functionalism should be supplemented by an affordance-theoretic approach; second, I address why functions should not be replaced altogether by affordances. Let us begin with the former. Here, I argue that functionalism should be understood through the lenses provided by the notion of affordances. That is, by considering functions as “the point of concepts”, we are underselling the role that functions can and should have in conceptual engineering projects, by only focusing on what concepts *should do*. This is in line with criticism advanced by Nado (2021), who writes:

[I] don’t think we ought to limit ourselves to speaking of something like the ‘proper’ function of a concept. Tools, concepts included, can be used for purposes other than their ‘proper’ function; I can use a hammer as a paperweight, or a knife as a lever. Weighting papers may not be the proper function of a hammer, but insofar as I use my hammer for that purpose, it makes perfect sense to speak of weighting papers as one the hammer’s (current) functions. (Nado, 2021, S1521).⁶⁹

The current function of the hammer that Nado is referring to in the example can, and perhaps should, be better understood in terms of affordances. Relatedly, the full spectrum of conceptual uses that should be of interest to conceptual engineers (at least) for the assessment of conceptual adequacy is better captured by grounding functionalism in a theory of affordances. That is to say, by moving from the current notion of function (i.e., what concepts *should do*) to a notion of function

⁶⁹ Interestingly, Thomasson (2022) herself draws heavily from Millikan (1984, 1989) in her discussion of functions framed in terms of ‘proper’ vs ‘accidental’ functions. Similarly, the literature on the functions of artefacts in the philosophy of technology also draws heavily from Millikan’s work on functions in (evolutionary) biology.

grounded in a theory of affordances (i.e., what concepts should *enable us to do*), conceptual engineers can achieve greater granularity in their analytic and normative approach to conceptual work. However, one may argue that, just because functions *can* be understood in terms of affordances, it does not necessarily follow that they *should* be. Here, I argue that grounding conceptual functions in a theory of affordances is not only warranted, but desirable, from the perspective of the intellectual and professional integrity of conceptual engineers.

That is, if understanding the notion of function through that of affordance gives us good reasons to think that functions alone may fail to capture the complexity of the uses that concepts can and should enable, as I suggest is the case, choosing not to pursue the investigation of the notion of affordances seems ethically irresponsible. Rather, conceptual engineers should embrace such a complexity, even if this will lead to new issues being uncovered, which may not be meaningfully solved, mitigated, or even addressed by current approaches, and for which new tools and methods may need to be devised, and indeed even if such issues may not be meaningfully solvable *tout court*.

Intuitively, it seems reasonable to expect that such considerations should be even more pressing when the concepts being (re-)engineered are lay concepts or can be expected to have an impact on the broader societal context in which they will be implemented. In such cases, failing to take into account the perceived complexity of conceptual functions and the multi-stability of concepts may lead to a rather grim picture, and may be framed in terms of conceptual engineers failing to exercise their due diligence. However, it should be noted that technical concepts which may not intuitively have significant societal impact may also become societally relevant.

Consider the concept HOLE. The metaphysics of holes has been studied extensively, most notably by Casati and Varzi (1994). Whether Casati and Varzi's work would qualify as conceptual engineering is not crucially relevant. What is relevant here is that their work contributed to the conceptualisation of the technical concept HOLE within the domain of ontology and metaphysics. The societal relevance of expert concepts such as HOLE studied from a decidedly metaphysical perspective, seemingly detached from society, should not be underestimated. Indeed, perhaps unpredictably, a few years after Casati and Varzi's (2004) book was first published, US jurists were animatedly discussing which conception of HOLE should be given precedent when faced with seemingly invalid electoral ballots. This was due to the poor design of the ballots, which

had led a significant number of voters to express two preferences after learning they had inadvertently voted for the wrong candidate, and to some voters encountering difficulties while operating the voting machines, which resulted in some ballots not being punched through completely. Ultimately, the conceptualisation of HOLE became societally relevant and arguably played a non trivial role in determining the outcome of the US Presidential Elections.

Therefore, the process one should undertake to determine what counts as a good enough concept in a given domain should also include an assessment of the possible uses of the concepts, beyond the intended and actual uses, and the dynamism of such uses. Relatedly, it may be preferable to exercise precaution and assume that any conceptually (re-)engineered concepts may have an impact on broader society.

It should also be reiterated that an unrefined notion of affordances has already emerged organically in the conceptual engineering literature bearing other names (such as “possible function” in Nado, 2021, as illustrated above). This is not unlike the case of “conceptual engineering” itself. While many scholars and practitioners engage in normative conceptual work, only part of them refer to it in terms of “normative conceptual work” or “conceptual engineering”. Consider bioethicists conceptualising BRAIN DEATH, legal scholars coining the concept LEGAL PERSONHOOD, and physicists adapting the concept PLANET. When considering such projects from the perspective of normative conceptual work, they would neatly fit within our understanding of “conceptual engineering”. However, it is possible that bioethicists, jurists, and physicists themselves would not refer to their activities in terms of conceptual engineering.

Similarly, I suggest that it is reasonable to assume that (at least) some conceptual engineers may already be using the notion of affordances without referring to it explicitly in these terms. Therefore, tapping into the rich philosophy of technology literature on affordances is warranted and desirable to the extent that it can illuminate a notion that has thus far remained under explored in the conceptual engineering literature, but which can be of great value for conceptual engineers. Taking inspiration from the philosophy of technology scholarship on affordances is particularly promising as the notion of affordances has been investigated in tandem with that of function, and the two concepts have often been mutually integrated. Moreover, engineers and philosophers of technology on the one hand and conceptual engineers on the other engage in similar activities and are motivated by similar goals.

Establishing a parallel between the two bodies of scholarship is further warranted if one considers that conceptual engineers generally take concepts to be *de facto* artifactual in some salient respect (see chapter 3).

4.3.1 Functions (still) matter. Why affordances should not replace functions

In this paper, I have argued that, despite being a helpful starting point to guide conceptual engineering projects, functionalism, i.e., the approach according to which concepts should be deemed adequate to the extent that they adequately fulfil their functions, does not fully capture the dynamic and contingent interplay between concepts and their contexts of use. Specifically, over the previous sections, I have illustrated the notion of affordances, and made the case that it can, and should, inform functionalism. Indeed, affordances uncover a complex web of interdependencies involving concepts, their users, and the context of use, thus moving past merely highlighting what concepts *should* do, and rather spotlighting what concepts *enable* us to do. As such, affordances allow to cast much larger a net and capture much broader a scope of the uses that concepts allow.

In this regard, one objection should be addressed. That is, (i) if it is the case, as I argue it is, that functions alone cannot meaningfully guide conceptual engineering projects, as they can only provide conceptual engineers with too narrow an account of the considerations that should be relevant to evaluate conceptual adequacy, and (ii) if it is the case that the notion of affordances can provide a richer, more holistic framework for such an evaluation, then (iii) should conceptual engineers not discard the functional account altogether in favour of an affordances-based approach to conceptual adequacy? I argue that that is not the case.

That is, the shortcomings of the functional approach should not be bridged by extending or inflating the notion of conceptual function so as to encompass conceptual affordances. Were that the case, conceptual engineers would risk both losing a useful notion of function and still lacking adequate vocabulary to discuss the broader considerations and implications of conceptual engineering beyond those captured by functions. Rather, conceptual engineers should confront the fact that the adequacy of concepts (and, relatedly, of their evaluation, design, and implementation) cannot be understood solely in terms of their functions but require stepping out of the functional bubble and supplementing conceptual functions with conceptual affordances.

Indeed, in conceptual engineering, functions (i.e., the specific roles or purposes that concepts are intended to fulfil) are critical because they provide clear adequacy criteria for concepts: a concept is adequate only if it successfully fulfils the function(s) it *should* serve. This makes functions a useful starting point for guiding conceptual revisions and ensuring that engineered concepts meet their goals. Affordances, on the other hand, introduce breadth and complexity by considering not only what a concept is intended to do (its intended function), but also what the concept enables or constrains users to do within a given context.

Thus, despite affordances providing a more comprehensive and sophisticated framework for the assessment conceptual adequacy in holistic terms, replacing functions with affordances altogether should be discouraged, as this could dilute the precision and purpose-driven nature of conceptual engineering.⁷⁰ Indeed, while functions provide focussed normative guidance for what concepts should achieve, affordances might introduce an overwhelming variety of possible conceptual uses. It follows that understanding functions in terms of affordances may introduce the risk of overextending a concept to the point where it loses its distinctiveness and utility. When a concept is stretched too far, it becomes harder to define and apply consistently. This complexity can dilute the core meaning of concepts, making them less effective in fulfilling their intended purposes.

At the same time, the role of affordances in the assessment of conceptual adequacy should not be downplayed. Indeed, consider an instance in which several competing conceptual proposals might all be deemed adequate, yet differ (even substantially) in the possibilities for action and thought that they afford. For example, while both fake news and disinformation capture the spread of false information, fake news may afford distrust, delegitimisation, and political polarisation, whereas disinformation may afford blame attribution as well as regulatory, institutional, and security-oriented responses. In such cases, just as sets of reasons for adopting one version of a concept over another can be weighed with decisive reasons favouring the concept whose reasons are weightier than those of any alternative (as argued by Köhler & Veluwenkamp, 2024), attending to the affordances of competing concepts

⁷⁰ Cosentino (2021) reaches similar conclusions through a different route.

can similarly aid in selecting among multiple adequate candidates, functioning as a potential tie-breaker.⁷¹

Ultimately, grounding functionalism in a theory of affordances can enrich conceptual engineers' understanding of the interaction between concepts, users, and contexts and its significance for conceptual engineering. By attending to affordances, conceptual engineers can move beyond purely functional evaluations to consider the broader socio-technical, institutional, and discursive trajectories that different concepts are likely to enable. Moreover, because conceptual affordances shape not only immediate use but also long-term pathways of meaning and institutionalisation, assessing affordances can help anticipate how concepts might evolve and what futures they make possible or foreclose.

4.4 Conclusion

Conceptual engineering is a normative approach to conceptual work, the purpose of which is the improvement of our concepts. To improve concepts, conceptual engineers need to have a measure of what an adequate concept amounts to. Functionalism offers such a standard. Following the functional account, concepts have functions and are adequate to the extent that they can fulfil their functions.

While traditional functional approaches have centred narrowly on the intended functions of concepts, recent contributions to the literature have enriched this view, advocating a more context-sensitive and dynamic understanding of conceptual functions (e.g., Nado, 2021; Riggs, 2021; Jorem, 2022; Queloz, 2022; Köhler & Veluwenkamp, 2024). In this paper, I built on such contributions and introduced the notion of *conceptual affordances*, understood as the potential actions, uses, or thoughts that a concept enables or constrains relative to a given user and within a given context. Conceptual affordances elaborate and refine the coarse-grained notion of *possible functions* that has emerged in the conceptual engineering literature, by tapping on the literature on artefactual functions, which draws from similar intuitions and pursues similar normative aims (Nado, 2021).

I argued that the move towards an affordance-informed functional approach to conceptual adequacy is not only theoretically

⁷¹ Future work will focus on the articulation of practical guidelines to facilitate the operationalisation of these insights.

enriching, but also crucial to address the ethical dimensions of conceptual engineering. Indeed, an affordance-theoretical approach systematically allows to consider a subset of the actions and thoughts that concepts enable and constrain, which should be deemed relevant for the purpose of conceptual engineering, but are not adequately captured by functions alone. This enables considering a wide range of (mis)uses of concepts beyond their intended roles, which should nevertheless be of interest to conceptual engineers and be considered in the assessment of conceptual adequacy.

Ultimately, an affordance-informed functional approach to conceptual engineering allows conceptual engineers to more skilfully navigate normative conceptual work, by enriching and refining the understanding of the roles of concepts and shifting the focus from what concepts *should* do to what they should *enable* us to do. Moreover, when multiple competing concepts are functionally adequate, their differing affordances can serve as a critical tie-breaking factor in conceptual selection, thereby enabling conceptual engineers to make more normatively sensitive and contextually attuned choices. Future work will focus on illustrating how such a framework can be translated into operationalisable guidelines for conceptual engineers.

Chapter 5

CONCEPTUAL RESILIENCE. CONCEPTUAL ENGINEERING AS PREVENTION AND CURE⁷²

5.1 Introduction — 5.2 From conceptual stability to conceptual resilience — 5.3 A graded account of conceptual resilience — 5.4 Modes of resilience. How do concepts exhibit resilience? — 5.4.1 Conceptual Resilience as Immutability (CRI) — 5.4.2 Conceptual Resilience as Adaptation (CRA) — 5.4.3 Relation between CRI and CRA — 5.5 (Re-)engineering for resilience. Conceptual engineering as prevention and cure — 5.5.1 Thick and thin normative concepts — 5.5.2 Levels of analysis. Individual concepts, conceptual clusters, conceptual scheme — 5.5.2.1 Individual concepts — 5.5.2.2 Conceptual clusters — 5.5.2.3 Conceptual scheme — 5.6 Conclusion

⁷² Sections §5.2-§5.4.3 of this chapter have appeared as an article in Marchiori, S., & Sta. Maria, J. (2026). Conceptualising conceptual resilience. A comparative approach. *Philosophy & Technology*, 39(45). SM took the lead and was responsible for the conceptualisation. SM wrote sections §5.2, §5.3, §5.4, and §5.4.3. Both authors jointly wrote sections §5.4.1 and §5.4.2.

5.1 Introduction

Recent work on conceptual engineering in the philosophy of technology has increasingly turned its attention to the phenomenon of *conceptual disruption*, understood as the interruption in the functional adequacy of concepts⁷³ (originally, Löhr, 2023; subsequently, Hopster & Löhr, 2023;

⁷³ The account of conceptual resilience advanced in this paper is intentionally non-committal regarding which view of concepts one ought to adopt. Rather than adopting a substantive metaphysical account of what concepts are, this paper recognises that competing theories of concepts (e.g., definitional, inferential, prototypical) are often motivated by distinct explanatory aims and background assumptions (Margolis & Lawrence, 2023). Committing to any one of these would entangle the paper in debates that are tangential to its purpose. Instead, the discussion proceeds from a minimal and broadly compatible assumption: that concepts can be treated as functional kinds (Lalumera, 2009). On this view, concepts are individuated and re-identified by their functions. This functional account is methodologically justified in at least two ways. On the one hand, the appeal of this view lies in its compatibility with a range of competing theories of concepts. This allows the discussion to proceed while remaining neutral across rival theories of concepts and still providing a clear criterion for identity and change. On the other hand, understanding concepts as a functional kind also aligns the present discussion on conceptual resilience with existing accounts of conceptual disruption, which are themselves framed in functional terms. Thus, adopting a functional view of concepts provides continuity between the discussions on disruption and resilience: if disruption marks the loss of functional adequacy due to pressures, resilience marks its preservation despite pressures. In this sense, the functional individuation of concepts is particularly well suited for articulating the dynamics of both conceptual disruption and conceptual resilience.

Hopster et al, 2023; Marchiori & Scharp, 2024; van de Poel et al, 2023).⁷⁴ On such an account, concepts have functions and are adequate to the extent that they fulfil their functions; conversely, concepts are disrupted

⁷⁴ Different accounts of conceptual disruption have been put forth in the literature (Hopster & Löhr, 2023; Hopster et al, 2023; Löhr, 2022; 2023; Marchiori & Scharp, 2024), all of which are compatible with, if not explicitly endorsing, a functional approach. Here, we choose to draw from Marchiori & Scharp's (2024) account of conceptual disruption as an "interruption in the [...] functioning of concepts" (Marchiori & Scharp, 2024, p. 1). On this account, conceptual disruption can take two main forms (conceptual gaps and conceptual conflicts) and can vary in severity (from mild to severe) depending on how urgently it calls for a conceptual engineering intervention. *Conceptual gaps* arise when no existing concept adequately applies to a phenomenon that ought to be captured at a certain level of precision. Conversely, *conceptual conflicts* occur in three main ways, as inconsistencies (i) among concepts' constitutive principles, (ii) between concepts' constitutive principles and facts about the world, or (iii) between concepts' constitutive principles and values or social norms (a *conceptual misalignment*, following Hopster & Löhr, 2023). Moreover, gaps and conflicts are not static, but may occur at different times in the very same process of conceptual disruption and develop into one another. Specifically, conceptual conflicts may be followed by conceptual gaps. For example, when new science and technology is introduced, this may lead to conceptual conflicts when existing concepts clash with new facts or practices. This may prompt attempts to redefine or adjust such concepts, which in turn may generate a gap, requiring the introduction of a new concept altogether. Conversely, a gap may reveal deeper tensions within our conceptual framework that eventually manifest as a conflict. For example, the invention of the mechanical ventilator first lead to a conflict between the concepts LIFE and DEATH, which was later followed by a conceptual gap that led to the creation of the new concept BRAIN DEATH. Following Marchiori & Scharp (2024), conceptual disruption affects concepts both individually and collectively and may manifests differently depending on the level at which it occurs: at the level of individual concepts, clusters of related concepts, or the broader conceptual scheme as a whole. For the purpose of this paper, the focus lies on conceptual disruption affecting individual concepts.

when they fail to do so.⁷⁵ For example, concepts can become contested (e.g., failing to serve as a shared epistemic tool, such as in the case of DEMOCRACY; Gallie, 1955), misleading (e.g., introducing distortions into discourse, as with rigid conceptualisations of MARRIAGE that fail to accommodate social change; Anderson, 2013; Fineman, 2001; 2006), or otherwise defective (e.g., by being historically tainted, fuelling unrealistic expectations, distorting value judgments, and enabling conceptual appropriation, such as in the case of ARTIFICIAL INTELLIGENCE illustrated in Hopster, 2024a). This can lead, among others, to breakdowns in

⁷⁵ Which functions concepts (should) fulfil, as well as how to assess adequate functional fulfilment, are the object of ongoing discussion in the conceptual engineering literature (e.g., Cappelen, 2018; Köhler & Veluwenkamp, 2024; Quelo, 2019, 2022; Riggs, 2021; Thomasson, 2021, 2022). The literature has offered divergent accounts of conceptual functions, including but not limited to: proper functions (Millikan, 1989), etiological functions (Prinz, 2018; Simion & Kelp, 2020), system functions (Haslanger, 2020; Thomasson, 2020), practical, evaluative, animating, and inferential functions (Quelo, 2019), normative functions (Köhler & Veluwenkamp, 2024). In this sense, there are many ways in which one can tease out an account of conceptual functions. However, the present discussion does not hinge on the adoption of any particular account. The aim here is not to settle the question of what conceptual functions *are* or what they *should be*, but to clarify how the notion of resilience might be understood *given* a broadly functional orientation to conceptual adequacy. The argument proceeds at a general level: whatever one takes conceptual functions to be, it remains meaningful to ask under what conditions the fulfilment of those functions can be sustained under pressure. I thus remain neutral with respect to which (if any) of the existing accounts advanced in the literature captures conceptual functions best, since the general framework of resilience developed in this paper can be instantiated across them, and rather work with a deliberately inclusive understanding of conceptual functions. For the purpose of illustration, this paper follows Riggs (2021) in adopting a deflationary functional account, whereby which function(s) a given concept should fulfil is (at least in part) context-dependent and should be determined on a case-by-case basis. It also follows Nado (2021) and Marchiori (2025) in rejecting an all-or-nothing stance towards functional fulfilment, i.e., concepts are not functionally adequate or inadequate in absolute terms; rather, they are more or less adequate. This point is expanded in section §5.3 when discussing the degrees of conceptual resilience. I recognise that the precise contours of what conceptual adequacy (and, *mutatis mutandis*, resilience) amounts to may vary depending on how one construes these functions, with different functional accounts plausibly leading to resilience being more or less easily achieved or achievable. While this is not the focus of the current paper, future work will explore how specific functional accounts may yield correspondingly distinct models of conceptual resilience.

communication, epistemic confusion, and normative uncertainty (Hopster, 2024b).

Conceptual disruption is a central concern for the philosophy of technology and bears significance for our engagement with technology in at least two ways. First, technology has an impact on the socio-technical landscape in which it is designed, developed, and deployed. As recent discussions illustrate, such an impact extends to concepts themselves (e.g., Hopster & Löhr, 2023; Hopster et al, 2023; Löhr, 2023; Nickel, 2020; van de Poel et al, 2023). Relatedly, when new technologies are introduced, they may prompt instances of conceptual disruption. For example, the advent of autonomous vehicles has led to (ongoing) sustained debates about the (in)adequacy of existing notions of responsibility (Goetze, 2022; Hindricks & Veluwenkamp, 2023; Matthias, 2004; Santoni de Sio & Mecacci, 2021). More recently, the proliferation of large language models has prompted renewed scrutiny of the adequacy of concepts related to writing (Floridi, 2025), authorship (Bajohr, 2024; Bimo, 2024; Coeckelbergh & Gunkel, 2025; Huang et al, 2025), and intellectual property (Gervais, 2025). Second, the very concepts that underpin and structure our analysis, evaluation, and governance of technology may be disrupted in a way that hinders our ability to do so effectively. Consequently, the relevance of conceptual disruption for the philosophy of technology extends beyond technology-induced disruptions alone.

The focus on conceptual disruption is proving useful as a diagnostic tool to signal that conceptual engineering (the evaluation, design, and implementation of concepts; Chalmers, 2020) may be warranted. Much of the existing (albeit still limited) literature on conceptual disruption has pursued this diagnostic function in tandem with a curative orientation, proposing strategies for restoring conceptual adequacy through conceptual engineering interventions once disruption has occurred (Hopster & Löhr, 2023; Hopster et al, 2023; Löhr, 2023; Marchiori & Scharp, 2024; van de Poel et al, 2023). These approaches have yielded valuable insights into how and why concepts are disrupted, and into how such a disruption may be overcome, whenever deemed desirable. Yet, to the extent that the literature on conceptual disruption has so far largely linked the investigation of when and how concepts are disrupted under pressure to restorative approaches, it has largely overlooked the complementary dimension of how concepts can resist disruption. Relatedly, the investigation of how concepts remain (more or less) functionally adequate *under* pressure remains widely under-explored. This asymmetry invites a complimentary line of inquiry.

This paper takes up this invitation. That is, there may be a benefit in articulating how concepts can resist pressures and disturbances (technology-induced or otherwise), beyond how they may fail to do so. Moreover, if there is a strong sense in which conceptual disruption may be undesirable, as has been suggested by Lundgren (2024),⁷⁶ further investigating how concepts may be more or less equipped to deal with disturbances and resist conceptual disruption should be of interest to conceptual engineers. To this end, the paper introduces the notion of *conceptual resilience*, understood as the ability of concepts to resist conceptual disruption by maintaining continuous functional adequacy despite tensions, pressures, or other disturbances. This builds on and expands recent work by Lundgren (2024) on the notion of *conceptual stability*. While Lundgren (2024) advocates for static resistance to disruption through stability, this paper also accounts for a dynamic mode of resilience through adaptation. In doing so, the paper provides a more comprehensive illustration of how concepts remain adequate under pressure, thus refining the tools available to conceptual engineers for engaging with conceptual adequacy under conditions of disturbances. Importantly, the account of conceptual resilience developed in this paper is non-committal on the desirability of (re-)engineering for resilience, and rather provides a basis from which both advocates and critics of resilience can evaluate its desirability. The second part of the chapter

⁷⁶ Indeed, Lundgren (2024) argues that, all other things being equal, it is desirable to avoid disruption. In Lundgren's words, "if we can avoid any conceptual disruption, then that has clear benefits" as "a stable concept is *ceteris paribus* better than a disruptable concept, given that it avoids the need for future conceptual work on that concept" (Lundgren, 2024, p. 2). Two considerations are in order. First, the literature on conceptual disruption does not attribute a positive or negative valence to the phenomenon of resilience in a vacuum, rather suggesting that resilience can be positive, negative, or neutral, depending on how it manifests in a given context (Hopster & Löhr, 2023; Hopster et al, 2023; Löhr, 2023; Marchiori & Scharp, 2024). Indeed, much of the literature is informed by discussions surrounding the neighbouring notion of *social disruption* (Hopster, 2021) in the philosophy of technology, which is itself explicitly noncommittal with regards to the determination of its desirability, which should be assessed on a case-by-case basis (Marchiori et al, 2025; van de Poel et al, 2023). Second, even granting that conceptual disruption may be undesirable (whether in specific contexts or in general), it does not follow that (re-)engineering the disrupted or disruptable concept for resilience will itself be desirable. In this sense, the current paper aims to ground future work on the desirability of (re-)engineering concepts for resilience.

builds on the notion of conceptual resilience to address the question of desirability. Should conceptual engineers (re-)engineer for resilience, and if so, under what conditions? It is argued that the answer depends both on the type of concept being (re-)engineered (thick or thin) and on the level of analysis (individual concept or conceptual scheme).

The chapter is structured as follows. Section §5.2 illustrates the notion of conceptual stability as introduced by Lundgren (2024) and argues for a move from stability to the more comprehensive notion of conceptual resilience, of which stability is a facet. Section §5.3 proposes that conceptual resilience should not be understood as a binary property, but as a capacity that can be exhibited in degrees. Section §5.4 articulates how concepts can exhibit resilience by distinguishing between different modes of resilience, drawing from interdisciplinary work to introduce two heuristic framings—Conceptual Resilience as Immutability (CRI) and Conceptual Resilience as Adaptation (CRA)—which illuminate contrasting yet complementary modes through which concepts preserve functional adequacy under pressure. Section 5 poses the question of desirability, illustrating the distinction between thick and thin normative concepts, and the relevance of the levels of analysis employed in the determination of desirability. Section §5.6 sketches a tentative answer to the question of whether conceptual engineers should try to prevent future disruptions by (re-)engineering concepts for resilience. Section §5.7 contains the conclusion.

5.2 From conceptual stability to conceptual resilience

Much of the existing literature on conceptual disruption has concentrated on identifying when and how concepts are disrupted under pressure, and how such disruptions can be addressed through conceptual engineering interventions (Hopster & Löhr, 2023, Hopster et al, 2023; Löhr, 2023; Marchiori & Scharp, 2024). Recent work by Lundgren (2024) shifts the attention from restorative to preventative conceptual engineering. Rather than analysing conceptual disruptions after they occurred, Lundgren suggests that some concepts can be deliberately designed to resist disruption in the first place. Lundgren (2024) observes that technological developments, normative critique, and conceptual challenges frequently force revisions or replacements of concepts. If conceptual engineers could construct concepts that resist such pressures, the need for future revisions might be mitigated or even eliminated.

Lundgren (2024) frames the discussion in terms of conceptual *stability*, suggesting that some concepts can be designed in ways that ensure they remain stable across contexts, avoiding counterexamples, normative critique, and conceptual disruption.⁷⁷ Specifically, Lundgren proposes that this can be achieved by formulating concepts at a sufficiently high level of abstraction. His case study of INFORMATION SECURITY illustrates this strategy: rather than defining SECURITY in terms of fixed properties that may be challenged by technological developments, the Appropriate Access definition relativises security to stakeholder needs, ensuring that the concept remains applicable even as specific security concerns evolve. Following Lundgren's account, conceptual *stability* refers to the preservation of a concept's functional adequacy across different contexts through preemptive (re-)design.

In this paper, we build on and extend Lundgren's (2024) proposal by offering a more comprehensive and explanatorily rich account of the ability of concepts to resist conceptual disruption. Indeed, while Lundgren's model provides a compelling account of how some concepts can be (re-)engineered in such a way that makes them resistant to disruption, it does not fully capture the broader range of ways in which concepts remain adequate despite disturbances. Specifically, this chapter argues that *conceptual stability* and *conceptual disruption* alone cannot fully capture the spectrum of ways in which concepts can be affected by tensions, pressures, and disturbances, which need not raise to the level of a conceptual disruption to be of interest to conceptual engineers. A richer understanding should be explored of how concepts behave under pressure, providing the necessary vocabulary to describe not just how concepts are disrupted under pressure, but how well they can endure such pressures and disturbances while retaining functional adequacy.

With such concerns in mind, the chapter hereby introduces the notion of *conceptual resilience* to describe the ability of concepts to resist conceptual disruption by maintaining continuous functional adequacy

⁷⁷ Lundgren (2024) clarifies that a distinction could be drawn between conceptual *stability*, which describes concepts which remain adequate *despite* disruptions, and conceptual *undisruptability*, which describes concepts which *cannot* be disrupted. Ultimately, both terms are used as synonyms in Lundgren (2024).

despite tensions, pressures, or other disturbances.⁷⁸ *Despite tensions, pressures, or other disturbances* is crucial here. Indeed, a concept may be functionally adequate and yet not well-suited to resist pressures. For example, consider a concept that is adequate because it has never been challenged, but would not be able to maintain functional adequacy if challenged. There seems to be a salient difference between having the capacity to resist conceptual disruption and not having been disrupted.

Consider an analogy with virtues. A person's bravery is not fully apparent unless they are placed in a situation requiring bravery. Similarly, a concept might appear to resist disruption because it has never been challenged in the first place; however, this does not entail that it possesses the capacity to withstand significant challenges. Consider the concept of PRIME NUMBER. A number p is said to be prime if (i) $p > 1$, and (ii) p has no positive divisors except 1 and p (Hardy & Wright, 1960, p. 2). That is, primes are defined as numbers that can only be divided by themselves and by 1. While a great deal of research has been (as is still being) conducted on primes, such as research aimed at understanding what primes are (e.g., Jameson, 2003; Maynard, 2019; Saidak, 2006) and how they are distributed (e.g., Conrey, 2003; Tao, 2011), developing new mathematical methods to study them (e.g., O'Neill, 2009; Selberg, 1949), applying them in cryptography (e.g., Ivy et al, 2012; Shemanske, 2017; Zhou & Tang, 2011), and exploring connections between primes and other advanced areas of mathematics (e.g., Lozano-Robledo, 2019;

⁷⁸ The literature on conceptual disruption that this paper builds on is itself either explicitly operating from a functional account of concepts (Marchiori & Scharp, 2024) or nevertheless compatible with such a position (Hopster & Löhr, 2023, Hopster et al, 2023). While we deem adopting a functional lens to ground conceptual resilience overall desirable, this also leads to possible downstream complications. That is, if concepts are a functional kind (as we assume is the case in this paper), and are therefore identified by their functions, discrepancies in the determination of a given concept's function(s) by different actors may lead to such actors referring to different concepts to the extent that such function(s) may be understood as not equivalent. That is to say, if one understands the function of GOOD as "providing normative guidance with respect to desirable actions and outcomes", they may not be willing to accept a concept of GOOD whose function is identified as "providing the standard for ethical behaviour". While we do not think this is an overriding issue of functionalism, it may introduce challenges, particularly in the form of metalinguistic disputes (e.g., Chalmers, 2011; Plunkett, 2015; Thomasson, 2017; Plunkett & Sundell, 2023). Consequently, conceptual engineers should be mindful of this point.

Nathanson, 1996), such a concept remains functionally adequate and unchallenged in mathematics scholarship.⁷⁹

Thus, resisting disruption is alone insufficient for the attribution of conceptual resilience, as resilience must account for the concept's capacity to adequately fulfil its functions under strain. In this sense, conceptual resilience is best understood as a disposition rather than a static state, and more is needed to determine whether a concept is resilient. Consider three buildings: the first two buildings are constructed in an earthquake-prone seismic zone, of which only one is constructed with anti-seismic design, and another building is constructed in a non-seismic zone with no special structural reinforcements. Following an earthquake, the building in the earthquake-prone seismic zone which was built without structural reinforcements collapses. We can say that the building was disrupted. However, if both the other buildings remain standing, can we say that they are both resilient? Probably not. The building in the seismic zone demonstrates resilience by withstanding disturbances, while the other building's stability offers no insight into its ability to endure disturbances. Similarly, the presence of disturbances is essential to assess whether concepts are resilient.

This chapter proposes that framing such a phenomenon in terms of *resilience* rather than *stability* better accommodates the plurality of conceptual responses to pressures and disturbances, particularly those that involve adaptive transformation rather than static continuity (as in Lundgren, 2024), a phenomenon that the notion of stability may overlook. The concept of resilience has been extensively theorised across multiple disciplines, where it generally denotes the capacity of a system, entity, or structure to withstand, absorb, or recover from external pressures and disturbances. In ecology, resilience is often understood as the ability of an ecosystem to return to equilibrium after perturbation, either by resisting change (engineering resilience) or by adapting to new conditions (adaptive resilience) (Holling, 1973). Psychology and cognitive science, in turn, frame resilience as an individual's ability to cope with adversity, maintaining psychological stability or adjusting effectively to stressors (Bonanno, 2004). In engineering, resilience is associated with fault tolerance and robustness, referring to the ability of structures or systems to continue functioning despite failures or unexpected disruptions. This broader theoretical background informs this account of *conceptual* resilience.

⁷⁹ Many thanks to Stefan Buijsman for double-checking this example.

The distinction between *conceptual stability* and *conceptual resilience* is not merely terminological but reflects a substantive theoretical commitment. If we were to speak only in terms of conceptual stability, we would risk overlooking cases where concepts persist by adapting to disturbances rather than by resisting such disturbances. This approach, therefore, does not reject conceptual stability but subsumes it as one possible mode of conceptual resilience. While some concepts achieve resilience through stability, others achieve it through flexibility, and this account seeks to make this plurality of modes of resilience explicit. Unlike *stability*, which suggests a continuation of the same state, *resilience* encompasses continuous functional adequacy as both persistence and adaptability. Framing the discussion in terms of conceptual resilience rather than stability allows us to recognise that not all concepts endure by remaining fixed—many persist precisely because they mutate and adapt in response to pressures without losing their function, and without the need for conceptual engineering interventions. Ultimately, the advantage of this framework is that it allows conceptual engineers to assess not only *whether* a concept is resilient but *how* it exhibits resilience, enabling a more nuanced evaluation of conceptual adequacy over time, and providing insights into the considerations that should guide conceptual engineering projects aimed at the (re-)design of concepts capable of resisting disruption.

The next section builds on the notion of conceptual resilience and explores the continuum of conceptual resilience, illustrating how some concepts exhibit resilience in degrees, as opposed to in binary terms. This will set the stage for the heuristic distinction between the different modes of conceptual resilience (Conceptual Resilience as Immutability and Conceptual Resilience as Adaptation), which will be developed in detail in section §5.4.

5.3 A graded account of conceptual resilience

The previous section introduced the notion of conceptual resilience to describe the ability of concepts to remain functionally adequate despite pressures and disturbances. This section expands on one feature of conceptual resilience, namely its non-binary nature. It argues that resilience should not be understood as an all-or-nothing property, but rather as a matter of degrees.

Conceptual stability, as described by Lundgren (2024), focusses on cases of maximal resistance to disruption. According to such a

position, either a concept resists disruption or it does not. Indeed, as illustrated in the previous section, Lundgren (2023) introduces the notion of *conceptual stability* to capture both “*undisruptable concepts* (i.e., concepts that cannot be disrupted), and *stable concepts* (i.e., concepts that remain stable in a situation of disruption)” (p. 1). For the purposes of his paper, Lundgren (2024) explicitly treats undisruptable and stable concepts as interchangeable. A charitable reading of this move suggests that conceptual stability is conceived as a binary property rather than a gradable one. While one might, in principle, distinguish between degrees of stability, e.g., between concepts that are contingently stable under certain pressures and those that are stably adequate across all conditions, the notion of *undisruptability* allows no such gradation. Therefore, to equate the two is to align stability with a property that admits only of all-or-nothing states. This suggests a certain absoluteness in concepts’ functional adequacy: stable concepts are *disruption-proof*.⁸⁰

This chapter suggests that a similar binary classification should not be generalised beyond conceptual stability to conceptual resilience. Rather, it seems preferable to understand conceptual resilience as the property of being *disruption-resistant*, not *disruption-proof*. What this means in practice is that stable concepts as described by Lundgren (2024) will display the highest possible degree of resilience according to this framework, but being disruption-proof will not be the measure of conceptual resilience. This allows for a nuanced understanding of resilience that goes beyond a binary approach of resilient versus non-resilient concepts. Concepts displaying a low degree of resilience are significantly vulnerable to disruptions and can easily become obsolete or inadequate in the face of changing circumstances, while concepts with a high degree of resilience can withstand significant disturbances and still continue to adequately fulfil their functions.

Consider the concept PRIVACY. PRIVACY extends far beyond the confines of individual data protection; it also includes decisional privacy, spatial privacy, and contextual integrity, each of which may be

⁸⁰ It should be noted that Lundgren (2024) does not suggest that conceptual stability should be understood in absolute terms. However, whether or not Lundgren (2024) would endorse a binary account of conceptual conceptual stability is not crucially relevant here. Rather, we are interested in understanding whether this may be a desirable framing, were this to be generalised beyond conceptual stability to conceptual resilience. Consequently, the criticism of a binary approach to conceptual stability and, relatedly, conceptual resilience might not apply to Lundgren (2024).

differentially impacted by technological, legal, and cultural developments (Nissenbaum, 2009; 2011; 2018; Solove, 2010). Other examples include *AUTONOMY*, whose applicability varies across medical, legal, and technological contexts (Dworkin, 1988; Mackenzie & Stoljar, 2000); *JUSTICE*, which takes on different dimensions depending on its function in distributive or corrective frameworks (Rawls, 1971; Young, 1990); and *TRANSPARENCY*, which ranges in depth and clarity depending on institutional or epistemic context (Ananny & Crawford, 2018; Floridi, 2023). This also suggests that conceptual resilience may be evaluated along several dimensions. For example, one might imagine that conceptual resilience may be assessed, beyond the ability of a concept to preserve functional continuity, through a contextual dimension, assessing the range of situations across which the concept remains functionally adequate. Indeed, a concept might be resilient by only retaining functional adequacy within a narrower range of contexts. In this sense, consider the legal concept *LIABILITY*, which was supplemented by the new notion *STRICT LIABILITY*, introduced to ensure accountability in cases where intent or negligence could not or need not be proven, e.g., with regards to pet owners' liability for damages caused by their pets (Comporti, 2009), and which is currently being discussed as a liability regime for artificial intelligence systems and other emerging technologies (Ćerka et al, 2015; de Fonseca et al, 2023; McDonald, 2023; Wendehorst, 2020). This addition preserved the original function of *LIABILITY* (allocating responsibility) while narrowing its scope of application after introducing *STRICT LIABILITY* to apply to circumstances to which the concept *LIABILITY* originally applied. The concept of *LIABILITY* thus retained functional adequacy but within a smaller contextual range. Relatedly, assessing resilience may involve recognising trade-offs between maintaining functional adequacy and retaining breadth of application, which may vary depending on a combination of the concept, the context of use, and the users.⁸¹ This seems to support the benefits of a graded model of resilience, whereby concepts lie on a continuum extending from a concept being disrupted to a concept being disruption-proof.

⁸¹ As already noted, further work is needed to develop a systematic criterion for functional adequacy. However, because this account of conceptual resilience is compatible with different functional theories, we refrain from proposing such a standard in this paper, as the relevant criterion for adequacy will likely depend on the functional framework one adopts.

Moreover, practical considerations suggest that a graded approach to conceptual resilience is desirable for conceptual engineers. Indeed, conceptual engineering is a normative approach to conceptual work aimed at improve concepts, but such improvement typically proceeds by degrees. That is to say, when intervention is warranted,⁸² conceptual engineers rarely (if ever) (re-)engineer concepts so that they are *perfectly* able to fulfil their functions; rather, improvements are often incremental.⁸³ Relatedly, assuming (as this chapter does) that a notion able to pick out a concept's ability to resist disruption will be of interest to conceptual engineers (e.g., for evaluation and design purposes), it seems misguided to expect that conceptual engineers will find such a notion useful when it only picks out the exceptional cases where concepts are disruption-proof. Holding too high a bar for conceptual resilience by only allowing to pick out concepts where the property of resisting disruption is exhibited in the highest degree would overly restrict the benefits of having such a notion in the first place. Indeed, this would reduce the practical utility of referring to resilience enough as to make such a notion uninteresting for the purpose of most conceptual engineering projects. Just like conceptual adequacy is almost exclusively assessed in relative terms (e.g., Nado, 2021; Marchiori, 2025), it seems plausible to expect that a useful notion of conceptual resilience will pick out concepts exhibiting varying degrees of resistance to disruption.

In this sense, understanding conceptual resilience as a matter of degree, with some concepts being more or less resilient depending on how well they perform their functions under pressure, is particularly important in the context of conceptual engineering. Identifying the degree to which concepts exhibit resilience enables conceptual engineers to assess their relative vulnerability to disruption and, in turn, develop more effective strategies for maintaining their functionality. A nuanced, gradated understanding of resilience thus facilitates a more targeted

⁸² It is, of course, possible that some concepts may be functionally adequate and not require (re-)engineering.

⁸³ Here, the point is that conceptual adequacy is rarely achieved in absolute terms. Even radical proposals (such as Haslanger, 2000; 2020), are themselves subject to subsequent refinement and debate, illustrating that conceptual engineering operates as an ongoing process or continuous (re-)evaluation and improvement, rather than a one-time correction (e.g., Dembroff, 2024; Díaz-León, 2025; Haslanger, 2018; Jenkins, 2016; Marques, 2025; Saul, 2006).

approach to conceptual design and revision, ensuring that challenges, while inevitable, do not necessarily become catastrophic.

5.4 Modes of resilience. How do concepts exhibit resilience?

The previous sections have proposed to conceptualise *conceptual resilience* as the capacity of concepts to maintain their functional adequacy despite tensions and disturbances. It has also been argued that resilience should be understood as a property that admits of degrees. Together, these considerations provide valuable guidance for conceptual engineering, in that they equip conceptual engineers with a starting point for the identification, assessment, and (re-)design of resilient concepts.

However, the question remains open of *how* concepts exhibit resilience. That is, is a concept resilient because it maintains its functional adequacy by being impervious to disturbances, or because it preserves its functional adequacy by being malleable and adaptable in response to disturbances? An explicit articulation of the modes in which resilience manifests vis-à-vis concepts could yield a richer understanding of the phenomenon of conceptual resilience. Such clarity could in turn inform conceptual engineering interventions, by providing valuable insights for conceptual engineers seeking to design or refine concepts for (or against) resilience. Once more, the chapter explores whether similar investigations into different modes of resilience are present in neighbouring literatures discussing related notions of resilience and, if that is the case, whether further clarity can be gained by tapping into such insights.

Drawing from, and synthesising, the rich tapestry of conceptualisations of RESILIENCE,⁸⁴ two understandings of resilience often emerge, which reflect two approaches to the same notion, namely resilience as immutability and resilience as adaptive capacity (Gunderson, 2000). Among the facets of resilience as immutability are: (a)

⁸⁴ Resilience has been widely studied in numerous domains, including psychology (Bonnano, 2004; Caldeira & Timmins, 2016; Çuhadar, 2014; de Terte et al, 2014), ecology (Cretney, 2014; Gunderson, 2000; Holling, 1973, 1996; Walker & Salt, 2006; Walker et al, 2004), disaster studies (Manyena, 2006; Patel et al, 2017), engineering (Bruneau et al, 2003; Hosseini et al, 2016); complex systems (Allen & Holling, 2008), management (Hamel & Välikangas, 2003; Lengnick-Hall et al, 2011; Vogus & Sutcliffe, 2007), urban studies and planning (Leichenko, 2011; Meerow et al, 2016).

persistence, i.e., ability to persist in the face of disturbance without altering the core function; and (b) resistance, i.e., withstanding disturbances without significant change (Holling, 1973; Pimm, 1984; Gunderson, 2000; Alexander, 2013). Among the dimensions of resilience as adaptation are: (a) adaptation, i.e., capacity to adapt to new conditions and evolve in response to disturbances (Rutter, 1987; Berkes & Folke, 1998; Luthar et al., 2000; Carpenter et al., 2001; Rose, 2004; Walker et al., 2004; Vogus & Sutcliffe, 2007; Hill et al., 2008; Norris et al., 2008; Folke et al., 2010; Magis, 2010; Bhama et al., 2011); (b) coping, i.e., implementing short-term strategies to manage the immediate effects of disturbances (Southwick & Charney, 2012); (c) flexibility, i.e., ability to change processes or structures in response to disturbances (Briguglio et al., 2009; Keck & Sakdapolrak, 2013); (d) learning, i.e., the process of learning from disturbances to improve future resilience (Southwick & Charney, 2012); (e) recovery, i.e., returning to a prior state after disturbance (Gunderson, 2000; Carpenter et al., 2001; Rose, 2004, 2007; Hill et al., 2008; Magis, 2010; Manyena et al., 2011); (f) redundancy, i.e., having multiple pathways or components to ensure function is maintained if one part fails; and (g) transformation, i.e., potential to transform in order to maintain or improve a function in the face of disturbances (Folke et al., 2010; Westley et al., 2011). Lastly, one facet, that of anticipation, i.e., the ability to foresee potential disturbances and prevent loss of function through mitigation and proactive measures, is shared by both approaches (Hamel & Välikangas, 2003).⁸⁵

Here, the chapter builds on this intuition and suggest that such a distinction can be productively applied to the phenomenon of *conceptual* resilience. It is proposed that resilience vis-à-vis concepts could be meaningfully conceptualised as a phenomenon that can be framed in terms of immutability and adaptation, relative to concepts' functional adequacy. It is argued that such ways of framing conceptual resilience can be helpful heuristics for understanding how concepts exhibit resilience and, relatedly, can provide helpful pointers with respect to how to assess and (re-)engineer for resilience. On the one hand, conceptual resilience may manifest in static terms. This chapter calls this *Conceptual Resilience as Immutability* (CRI). According to CRI, conceptual resilience

⁸⁵ It should be noted that this is not intended as an exhaustive overview of the conceptualisations of the notion of resilience in the literature on resilience but are merely meant to illustrate the richness of the discussion surrounding the concept of resilience.

refers to the ability of concepts to remain unchanged, i.e., to retain continuity in how they perform their functions in the face of disturbances. Lundgren's (2024) framework aligns closely with CRI, emphasising how concepts can be (re-)engineered to avoid disruption. On the other hand, conceptual resilience may manifest in dynamic terms. This chapter calls this *Conceptual Resilience as Adaptation* (CRA). According to CRA, conceptual resilience refers to the capacity of concepts to adapt and respond to disturbances while maintaining their core functions.⁸⁶ Both facets of resilience will be illustrated in turn.

5.4.1 Conceptual Resilience as Immutability (CRI)

A static account of conceptual resilience, associated with the notion of Conceptual Resilience as Immutability (CRI), emphasises the ability of a concept to remain unchanged despite external pressures. From this perspective, a concept is resilient if it can preserve its function(s) across different contexts and over time. This approach aligns with philosophical traditions that value stability and continuity in conceptual frameworks. In some cases, the immutability of a concept is necessary for it to serve as a reliable standard or foundation within a given system. For example, in legal contexts, the concept CONTRACT may need to remain relatively stable to provide a consistent basis for contractual obligations and rights. Here, the resilience of the concept is demonstrated by its ability to resist redefinition or reinterpretation, which could undermine its role in guiding legal reasoning and practice. In this static model, resilience is linked to the concept's capacity to maintain its functional role without requiring modification. The stability of the concept ensures that it continues to meet the needs of its users, providing a reliable framework for thought, communication, or action. However, the functional account also recognises that this immutability must be justified by the concept's continued functional adequacy. If a concept becomes outdated or irrelevant due to its rigidity, it would no longer be considered resilient

⁸⁶ As already mentioned, despite using similar terminology, this account of conceptual resilience as adaptation differs from Hopster & Löhr's (2023) notion of conceptual adaptation. Conceptual resilience through adaptation captures the property of concepts remaining adequate despite pressures, whereas conceptual adaptation in Hopster & Löhr's (2023) sense points to a stance towards the conceptual engineering process. Therefore, a concept may be resilient even in the absence of a conceptual engineering intervention, whereas conceptual adaptation occurs within a conceptual engineering process.

from a functional perspective, as it would fail to fulfil its intended role effectively.

To better understand Conceptual Resilience as Immutability (CRI), consider the concept GOOD as understood in the Platonic tradition. The concept GOOD is often thought to be resilient because it continues to fulfil its normative function: guiding action and evaluating desirability across different contexts. But what grounds this resilience? The Platonic tradition provides one answer by positing that the concept GOOD refers to a timeless and unchanging ideal (the *Form of the Good*) which subsumes all instances of goodness under a unified essence. According to this view, the concept is equipped with an intention or definition that preemptively includes all possible valid occurrences of its extension. This capacity to encompass all relevant instantiations is precisely what secures its ongoing functional adequacy.

On such a view, the concept is or becomes immutable: if its definition is already complete, there is no need to revise it in the face of new circumstances. The concept thereby remains adequate not because it adapts, but because it never encounters a challenge that calls for adaptation. Resilience here is thus not dynamic but static: the concept retains its function over time by virtue of not requiring change. In this way, its immutability is a feature, not a limitation, one that ensures continued use without the risk of disruption.

This understanding is echoed in Lundgren's (2024) account of conceptual stability. Lundgren proposes that a concept such as "appropriate access" can be deliberately engineered to remain stable by formulating it at a high level of abstraction. This strategy aims to insulate the concept from disruption by counterexamples or technological shifts. The concept retains its relevance precisely because it resists revision thus demonstrating a form of resilience as immutability.

Both the Platonic and Lundgren's (2024) accounts point to an understanding of conceptual resilience where unchangeability is a strength, not a defect. These approaches suggest that when a concept's function is sufficiently general and its scope carefully designed, it may remain stable across contexts and over time. This is the hallmark of CRI: functional adequacy secured through constancy rather than adaptability.

5.4.2 Conceptual Resilience as Adaptability (CRA)

A dynamic account of conceptual resilience, associated with Conceptual Resilience as Adaptation (CRA), emphasises the concept's ability to

evolve and adapt in response to changing circumstances. Under this model, a concept is resilient if it can incorporate new information, adjust its boundaries, or modify its application to maintain its relevance and effectiveness. From a functional perspective, a dynamic account of resilience is particularly suited to contexts where flexibility and adaptability are key to maintaining functional adequacy. Concepts operating within rapidly changing environments, such as technology ethics or environmental policy, may require frequent adjustments to remain effective. For example, SUSTAINABILITY in environmental ethics may need to evolve in response to new scientific insights or societal shifts. Here, resilience is demonstrated by the concept's capacity to adapt while continuing to guide action and decision-making effectively.

Consider once more the concept GOOD. Unlike the Platonic tradition, which posits a fixed, essentialist definition of GOOD that captures its extension through immutable criteria, the Confucian tradition departs from Platonic essentialism and offers a dynamic, context-sensitive understanding of GOOD. In Confucian thought, the concept GOOD—as expressed through the notion of REN (仁), which can be understood as goodness, humanness, or benevolence—does not rest on a fixed, exhaustive definition.⁸⁷ Rather, it is characterised by its open-endedness, and its resilience stems from its capacity to adjust to varying normative demands.

Unlike the Platonic approach, Confucius did not attempt to articulate a singular, universal definition of REN. When his students asked about its meaning, he offered distinct descriptions tailored to their

⁸⁷ We do not intend to suggest that the idea of CRA be found only in the Confucian tradition or only in the broader tradition of Chinese philosophy. Indeed, the CRA frame echoes Moore's (1959 [1903]) open question argument, which suggests that any proposed definition of GOOD" remains open to normative challenge. While Moore's concern is meta-ethical, it similarly points to the conceptual openness and indefinability of GOOD.

moral development and contextual circumstances.⁸⁸ He described goodness as “hesitancy in speaking” to one disciple, “caring for others” to another, and “ritual propriety” in yet another instance. He also adapted to his interlocutors (Ashmore, 2016; D’Ambrosio, 2016, 2019). For example, when talking to Sima Niu, Confucius describes goodness as hesitancy in speaking because he believes that the former’s loquaciousness is detrimental to moral development. This pedagogical strategy reflects more than rhetorical flexibility: it embodies a conception of GOOD that adapts to particularities rather than relying on generality.⁸⁹ The absence of a fixed definition does not weaken the concept’s function; instead, it allows it to remain normatively effective across situations by modulating its application without compromising its role.

This adaptability is central to what the chapter terms Conceptual Resilience as Adaptation (CRA). The Confucian model suggests that a

⁸⁸ A parallel can be drawn between the Platonic and the Confucian traditions. In *Euthyphro*, Socrates asks the titular character what piety is. Euthyphro answers, “prosecuting those who commit an injustice, such as murder or temple robbery” (*Euthyphro*, 5d/ Plato, ed. Reeve, 2012, p. 7). He also says that Zeus chaining his wicked father, Cronos, is piety. Socrates then tells him, “What I urged you to do wasn’t to teach me about one or two of the many pieties, but rather about the form itself, by virtue of which all the pieties are pious. (*Euth.*, 6d-e/ Plato, ed. Reeve, 2012, p. 8)” What Socrates implies is that, in order for the multiple and variegated instantiations of piety to be understood as pious, there must be a singular idea of piety that serves as the unchanging standard by which all these instantiations are judged the way they are (i.e., as tokens of piety). It is the explication of this idea in terms of a definition that Socrates asks from Euthyphro. Now, consider the following examples in the *Analects*: “Zhonggong asked about Goodness [i.e., REN 仁]. The Master [Confucius] said, ‘When in public, comport yourself as if you were receiving an important guest, and in your management of the common people, behave as if you were overseeing a great sacrifice.’” (*Analects* 12.2/Confucius, trans. Slingerland, 2004, p. 126); “Sima Niu asked about Goodness. The Master said, ‘The Good person is hesitant to speak.’” (*Analects* 12.3/Confucius, trans. Slingerland, 2004, p. 126); and “Fan Chi asked about Goodness. The Master said, ‘Care for others.’” (*Analects* 12.22/Confucius, trans. Slingerland, 2004, p. 136).

⁸⁹ Confucius’ refusal to provide universal or exhaustive definitions of ethical ideas such as REN is also understood by scholars as part of his way of providing optimal, practical, moral guidance to his students (Xiao, 2008, Rogacz 2022). That is to say, Confucius considered the activity of exhaustively defining concepts to be an unhelpful exercise since it cannot be ultimately achieved and can potentially distract students from the more important task of cultivating ethical virtue.

concept can remain resilient not by resisting change, but by incorporating it. The concept GOOD, understood through REN, is structurally open-ended, which allows it to accommodate emerging circumstances while continuing to provide moral guidance. In this way, resilience is achieved not through rigidity but through flexibility, not through preemptive completeness but through ongoing responsiveness. Importantly, this adaptability is not arbitrary. Confucius' descriptions of REN are always contextually anchored and morally instructive, aimed at cultivating virtue in line with his students' specific situations. The lack of an exhaustive definition enables the concept to remain functionally adequate, because its normative orientation is preserved even as its instantiations shift. Thus, the Confucian approach illustrates how a concept can remain resilient by adapting rather than remaining unchanged. It maintains its function through situated elasticity, exemplifying CRA. This adaptive form of resilience demonstrates how a concept's openness to re-articulation can itself be a strategy for enduring relevance and normative effectiveness.

5.4.3 Relation between CRI and CRA

It should be emphasised that, by shedding light onto resilience as immutability and resilience as adaptation, this chapter does not wish to argue that such facets should pick out a dichotomy. Importantly, such ways of framing resilience are neither to be taken as mutually exclusive nor jointly exhaustive to describe the ways in which concepts may resist conceptual disruption.

First, such facets are not mutually incompatible, and it is possible that the same instance of conceptual resilience may be described as CRI or CRA by different actors. For example, as Jacobs (1999) suggests in his analysis of the concept of sustainable development, some concepts may remain resilient not through pure immutability or adaptability, but through a layered structure combining a robust normative core with flexible operational features. This dual structure enables such concepts to absorb pressures without losing coherence. Ultimately, what the chapter wishes to highlight when drawing the distinction between a static and a dynamic understanding of resilience is the extent to which one's perspective may greatly impact and shape how one understands and approaches resilience in the first place.

Second, CRI and CRA are not jointly exhaustive. While the literature converges significantly on the static and dynamic nature of

resilience as a helpful heuristic through which to understand the notion of resilience, this focus on static and dynamic conceptual resilience does not exclude that additional facets of conceptual resilience may emerge in the future, which may be of interest to conceptual engineers. As such, this conceptualisation is not rigid with respect to the essential features of resilience. Were additional facets of resilience to come to the fore, which would change or refine our understanding of the concept itself, this framework already accounts for and accommodates such a contingency (see Figure 5.1).

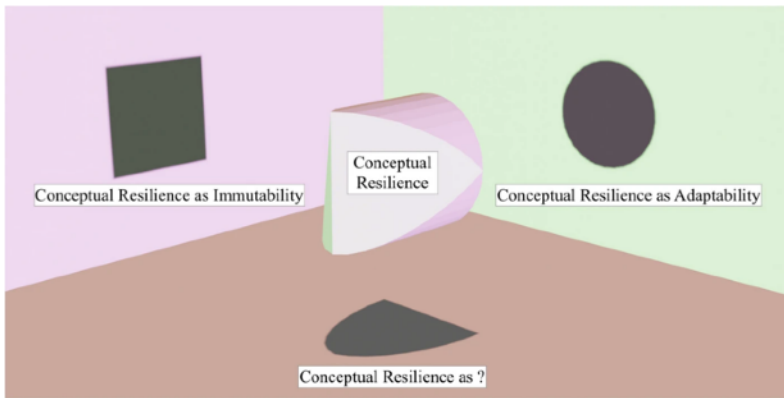


Figure 5.1. Facets of conceptual resilience

Ultimately, the static and dynamic framings of resilience may be mapped onto the continuum of resilience illustrated in sections 2 and 3, such that concepts exhibit resilience in degrees (low to high), and such resilience can be meaningfully understood through different frames (e.g., Conceptual Resilience as Immutability and Conceptual Resilience as Adaptation).

Three brief concluding considerations are in order. First, this chapter does not aim to be prescriptive and suggest that any one frame (CRI or CRA) should be preferred over the other. Second, and relatedly, while concepts resulting from CRI and CRA approaches to conceptual (re-)engineering may exhibit resilience in similar ways, they may differ in their philosophical orientation. For example, Lundgren (2024) explicitly distances itself from bottom-up approaches, more akin to a CRA frame, arguing that they risk excessive relativity (Lundgren, 2024, p. 10). Third and last, it should be noted that different theoretical commitments may lead to different interpretations of what displays resilience, whether

concepts themselves or conceptions. As already discussed, the account of conceptual resilience put forth in this paper can be applied across different theoretical accounts of concepts. Depending on one's preferred view of concepts, resilience may be understood as operating at the level of concepts or at the level of conceptions. For example, is it the concept GOOD that exhibits resilience, or rather specific conceptions of GOOD, such as PLATONIC-GOOD or CONFUCIAN-GOOD? On some accounts, it may be preferable to say that the concept of GOOD itself is resilient, while other accounts may find it more accurate to describe resilience as a property of conceptions (e.g. PLATONIC-GOOD or CONFUCIAN-GOOD). Accordingly, conceptual resilience is a property that can be instantiated across diverse theories of concepts and at different levels (concepts or conceptions).

5.5 (Re-)engineering for resilience. Conceptual engineering as prevention and cure

In the preceding sections, the notion of conceptual resilience was introduced as the capacity of concepts to maintain continuous functional adequacy despite pressures and disturbances. Such analysis was largely diagnostic: it clarified what conceptual resilience is, how it differs from conceptual stability, and how it admits of degrees and modes (CRI, CRA). The natural next step is to ask whether, and, if so, under which conditions, conceptual engineers should actively (re-)engineer for resilience. That is, should conceptual engineering further preventative

pursuits alongside curative ones, by intentionally (re-)engineering for resilience? To what extent is conceptual resilience desirable, if at all?⁹⁰

The following sections will illustrate two considerations that should guide the determination of the desirability of (re-)engineering concepts for resilience. First, the desirability of (re-)engineering for conceptual resilience is (at least in part) constrained by its feasibility. That is, depending on the functions that they should fulfil, different concepts may be, by default or by design, more or less prone to disruption. Relatedly, different concepts may be more or less compatible with resilience. Examining the distinction between thick and thin normative concepts (Williams, 1985) can illuminate how such concepts relate differently to resilience and guide the assessment of the extent to which either should be (re-)engineered for resilience.

Second, the notion of conceptual resilience builds on the phenomenon of conceptual disruption as illustrated in Marchiori & Scharp's (2024), where conceptual disruption is described as manifesting at three levels of granularity, i.e., affecting individual concepts, conceptual clusters, and conceptual scheme. It seems reasonable to expect

⁹⁰ There are clear affinities between the present discussion of *conceptual resilience* and Hopster and Löhr's (2023) account of *conceptual adaptation*. Conceptual adaptation designates an *approach* to conceptual engineering taken in response to (backward-looking conceptual adaptation), or anticipation of (forward-looking conceptual adaptation), conceptual disruption, with the aim of "preserving conceptual functions to adapt to new circumstances, irrespective of whether such adaptations qualify as ameliorative" (p. 3). By contrast, conceptual resilience captures a *property* of concepts, i.e., their capacity to remain functionally adequate under pressure. The two notions intersect insofar as conceptual adaptation can be understood as the *means by which* conceptual engineers may (re-)engineer concepts for resilience. Indeed, if one takes resilience to be a desirable property of concepts, conceptual adaptation would constitute the mechanism through which such resilience may be intentionally achieved. Accordingly, this contribution can be framed as a continuation and specification of Hopster & Löhr's (2023) work on conceptual adaptation, of which we here isolate and develop the proactive, forward-looking dimension. At the same time, conceptual resilience need not follow from conceptual adaptation, as it is a property that concepts can possess independently from conceptual engineering interventions. Ultimately, Lundgren's account of *conceptual stability* captures one route to resilience (immutability under pressure) whereas Hopster and Löhr's notion of *conceptual adaptation* encompasses a wider range of responses, both proactive and reactive, aimed at maintaining or restoring conceptual adequacy relative to a (present or expected) conceptual disruption. The current proposal occupies an intermediate position between these two contributions.

that, similarly to how conceptual disruption manifests differently across different levels, the desirability of conceptual resilience may itself be at least in part contingent on the level of abstraction that one adopts: individual concept, conceptual cluster, or conceptual scheme. Both considerations will be addressed in turn.

5.5.1 Thick and thin normative concepts

The distinction between thick and thin normative concepts (Williams, 1985; Blackburn, 1992; Dancy, 1993) offers a useful heuristic for assessing the desirability of (re-)engineering for conceptual resilience. Thin normative concepts (from here onwards, *thin concepts*), such as GOOD, BAD, RIGHT, WRONG, are evaluative, but not descriptive, concepts. They are evaluative in that they express approval or disapproval, but they do so without providing specific information about the context or the reasons behind the evaluation. For example, arguing that a given conduct is *wrong* conveys disapproval, but does not specify why it is wrong or what features make it so. Therefore, thin concepts are characterised by their broad applicability and minimal descriptive content. Conversely, thick normative concepts (from here onwards, *thick concepts*), such as INFORMED CONSENT and PRIVACY, are both rich in descriptive content and evaluative (Väyrynen, 2013; Gert, 2016). By combining both descriptive and evaluative content, thick concepts enable one to not only make an evaluative judgement, but also to provide information about what led to the evaluation in the first place, e.g., in terms of context or reasons. The thick/thin distinction has profound implications for normative discourse, in which thick and thin concepts serve distinct, but complementary, functions.

Thin concepts function as broad evaluative tools that facilitate general normative discourse and high-level ethical theorising. They are particularly useful in constructing normative frameworks and theories due to their abstract nature and broad applicability. Thin concepts provide common ground for normative reasoning and debate, and are often used in public normative debates due to their broad applicability, as they allow for general discussions about morality without delving into the specifics of particular actions or contexts. Being more abstract and less context-dependent, they are well-suited to provide ethical foundations in dynamic and rapidly changing concepts. For example, in constructing a utilitarian framework, the concept UTILITY serves as a thin concept guiding normative evaluation based on the outcomes of actions.

This abstract evaluative function is essential for developing and applying utilitarian principles across different contexts without being restricted by specific descriptive details.

Thick concepts, on the other hand, provide detailed normative guidance in specific contexts. They combine evaluative and descriptive functions, making them indispensable for nuanced normative judgement. Thick concepts enrich normative discourse by offering context-specific evaluations that are informed by particular traits, actions, and circumstances. Because of their rich descriptive content, thick concepts are rather inflexible, but provide deep contextual integrity and the nuanced understanding needed to navigate complex normative landscapes, offering detailed insight into why certain actions or traits are morally significant. For example, the concept of CHIVALRY in medieval times encapsulated not only the behaviour expected of knights, such as bravery, courtesy, and honour, but also detailed the specific contexts in which these traits were to be applied, such as in battle or in interactions with women and vulnerable parties. This concept provided detailed guidance for recognising, evaluating, and cultivating the moral significance of a knight's actions, deeply rooted in the social and cultural milieu of the time. Similarly, GENEROSITY is a thick concept in that it goes beyond a simple evaluation of giving and encompasses the motivations, intentions, and social context behind an act of giving. Generosity is not merely about the act itself but involves understanding the circumstances and the genuine willingness to help others, highlighting the moral significance of selfless actions.

In short, thick and thin concepts fulfil different but complementary functions. Thin concepts provide the evaluative abstraction needed for universal normative principles and theoretical coherence, while thick concepts offer the descriptive richness required for practical normative judgement and personal ethical growth. Understanding thick and thin concepts through the lenses of functionalism can also allow to address some criticisms of the thick/thin distinction. Indeed, historically, as illustrated above, thick and thin concepts have been articulated as oppositional in philosophical discourse (Williams, 1985). However, a non negligible part of the scholarship calls into question this rigid binary classification, challenging the conventional view that concepts can be neatly categorised as either purely thick or purely thin, e.g., arguing that not all thick concepts possess equal levels of descriptiveness, or that concepts can exhibit descriptive flexibility depending on their contextual applications (Gibbard & Blackburn, 1992;

Blackburn, 1993; Gibbard, 2003; Väyrynen, 2013). Here, I embrace such a position and employ a continuum perspective, where concepts are situated along a spectrum of descriptive and evaluative content.⁹¹

To further elucidate this continuum perspective, we can examine specific examples. Terms like RIGHT and WRONG represent some of the closest approximations of purely thin concepts, offering evaluative judgments without stipulating specific criteria for what constitutes RIGHT or WRONG. They function as broad moral markers without descriptive content. Concepts such as JUSTICE and VIRTUE may be relatively thin but still embody minimal descriptive content influenced by context. Moreover, terms like INTEGRITY and COURAGE embody a balance of descriptive and evaluative content. These concepts describe certain behaviours while also conveying evaluative significance, although their definitions remain somewhat open to contextual interpretation. In contrast, concepts such as CRUELTY, RACISM, and MISOGYNY are highly descriptive and evaluative, encompassing specific behaviours and attitudes that carry robust moral judgments. For example, CRUELTY signifies particular harmful actions, making it both descriptively rich and evaluatively loaded. By focusing on the functional roles of normative concepts, it becomes less crucial to draw a rigid line between descriptive and evaluative content.

Ultimately, thick and thin normative concepts do not stand in the same relation to resilience. Given their high level of abstraction and minimal descriptive commitments, it seems reasonable to expect that thin(er) concepts will be structurally better suited to display resilience and, relatedly, to tolerate resilience-oriented (re-)engineering interventions. Thick concepts, by contrast, draw their normative import from dense descriptive anchoring in particular practices and contexts. Therefore, attempts to render such concepts (more) resilient, e.g., by broadening their scope or relaxing their descriptive content, will

⁹¹ Although not traditionally included in the thick/thin continuum, I understand the continuum as extending from *purely evaluative concepts* on the one end—i.e., concepts providing evaluative content devoid of descriptions (e.g., GOOD, RIGHT)—to *purely descriptive concepts* on the other end—i.e., concepts providing factual descriptions devoid of evaluative content (e.g., WATER, NUMBER). In this sense, concepts located at the extremes of this continuum present a limiting case in which a quantitative reduction in either evaluative or descriptive content yields a qualitative shift in kind. When the relevant dimension is driven to zero, the concept ceases to be merely thinner or thicker and instead becomes purely evaluative or purely descriptive, respectively.

plausibly tend to dilute their thickness. Thus, the result would not be the same concept rendered more resilient, but a thinner successor that no longer fulfils the function the original concept was expected to serve. In this sense, (re-)engineering thick concepts for resilience may be undesirable to the extent that this would imply discarding thick concepts in favour of thinner ones.

5.5.2 Levels of analysis. Individual concepts, conceptual clusters, conceptual scheme

As articulated in the framework by Marchiori and Scharp (2024), conceptual disruption can occur across various levels, impacting not only individual concepts but also groups of interrelated concepts (*conceptual clusters*) and, more broadly, the overarching *conceptual scheme*. As the notion of conceptual resilience adopted in this paper builds on the notion of conceptual disruption advanced by Marchiori and Scharp (2024), it seems reasonable that resilience itself, as well as its desirability, should be examined across these three levels.

5.5.2.1 Individual concepts

Let us begin with individual concepts. As discussed, individual concepts display resilience to the extent that they retain continuous functional adequacy despite pressures. Moreover, as illustrated in section §5.3, conceptual resilience is not a binary property but rather one that exists along a continuum. Similarly, normative concepts do not fall strictly into thick or thin categories, but rather lie on a spectrum extending from thick to (purely) thin concepts. This section elaborates on the desirability of resilience engineering for individual concepts.

The field of AI governance presents a fertile ground for examining the desirability of (re-)engineering normative concepts for resilience. Indeed, given the rapid advancement and pervasive influence of AI, it is crucial to assess whether normative concepts that are impacted by AI should be engineered for resilience to remain effective and useful despite changes in an ever-evolving socio-technical landscape. The distinction between thick and thin concepts can help guide such a determination.

On the one hand, thin concepts such as GOODNESS and FAIRNESS play a crucial role in AI governance due to their evaluative nature. These concepts are inherently flexible and can be adapted to various scenarios,

providing broad ethical guidelines that can evolve alongside technological progress. For example, FAIRNESS in AI can encompass issues of bias, discrimination, and equitable access to AI benefits, regardless of specific technological contexts. Thin concepts benefit from resilience engineering because their minimal descriptive content (if present at all) allows them to adapt to a variety of contexts while maintaining their evaluative purpose, ensuring they remain pertinent across diverse and changing socio-technical landscapes. Thus, engineering thin concepts for resilience ensures that these concepts can withstand disruptions and maintain their ethical relevance in addressing current and emerging normative challenges related to AI.

On the other hand, thick concepts in AI governance, such as INFORMED CONSENT and DATA PRIVACY, carry rich descriptive and evaluative content, which is tightly linked to specific contexts and practices. For example, INFORMED CONSENT in AI refers to the clear and comprehensive disclosure of AI system capabilities, data usage, and potential impacts on users. This concept is deeply embedded in regulatory frameworks and requires precise definitions to protect users' rights effectively. However, the inflexibility of thick concepts like INFORMED CONSENT makes them less adaptable to the rapidly changing socio-technical landscape. Indeed, thickness in concepts can lead to a lack of adaptability, making it difficult to address new ethical challenges and technological developments promptly. For example, a rigid definition of TRANSPARENCY might not accommodate emerging AI systems' unique characteristics, rather necessitating frequent updates to remain relevant. At the same time, efforts to engineer these concepts for resilience may result in overly nebulous approaches to AI governance that manage to keep pace with technological advancements at the expense of crisp and context-sensitive normative directions, thus leading to gaps in the governance of AI. For example, the concept of ACCOUNTABILITY might need to be nuanced enough to address the varying responsibilities of AI developers, users, and stakeholders across different sectors. However, for such a concept to retain the adequate level of precision to be pragmatically helpful for the governance of AI, the notion of ACCOUNTABILITY may need to be adapted often.

In this sense, I do not subscribe to the view that conceptual resilience is always desirable, as argued by Lundgren (2024).⁹² Rather, I claim that, at least when normative concepts are concerned, conceptual resilience is not immediately desirable, nor sometimes compatible, with thick concepts. That is to say, if the function of thick concepts is that of capturing a specific phenomenon or nuance, such a fine-grained purpose makes the concept more prone to conceptual disruption. At the same time, making a thick concept more resilient would inevitably translate to the concept being less apt to fulfil the role we think it should fulfil. Therefore, it seems like we cannot have it both ways. Either we have a normative concept that is better equipped to resist conceptual disruption, or we have a thick normative concept. As I illustrated with the example of normative concepts used in AI governance, I argue that thicker normative concepts are useful in ways that thinner normative concepts are not. Therefore, (sometimes) there are good reasons to think that we may need thick concepts more than we may need highly resilient concepts.

Therefore, while conceptual resilience provides stability and consistency, it can also hinder the contextual sensitivity required for effective AI governance. Relatedly, while thin concepts benefit from high resilience to maintain a stable normative framework, thick concepts in fields like AI ethics require less resilience and greater granularity to be effective in guiding ethical conduct. Thus, while highly resilient concepts provide stability and long-term reliability amid changing circumstances, the pursuit of conceptual resilience should be approached with caution.

A parallel can be drawn here between the undesirability of resilience for thick concepts and the notion of antifragility, originating from financial contexts (Taleb, 2012; Ramezani & Camarinha-Matos, 2020), which can be fruitfully transposed to conceptual engineering to denote a disposition to exploit disruption as a driver of functional adequacy. Conceptual antifragility can be understood as a positive-sum response to disruption: while resilient concepts resist pressures, antifragile concepts benefit from such pressures, which catalyse growth. Consider the concept SUSTAINABILITY. It could be argued that early

⁹² Lundgren (2024) writes: "If I can avoid any conceptual disruption, then that has clear benefits. [...] [I]f the task of philosophers and academics is partly aimed at providing and improving classifications for academia, or elsewhere, then a stable concept is *ceteris paribus* better than a disruptable concept, given that it avoids the need for future conceptual work on that concept" (Lundgren, 2024, p. 2).

formulations focussed narrowly on conservation, but successive disruptions (e.g., climate crises, industrial shifts, social-justice critiques) did not merely challenge and restore it, but rather transformed it into a richer, multidimensional concept incorporating e.g., ecological, economic, and social considerations. The concept's adequacy improved through pressures that would have destabilised a less plastic notion. Similarly, the concept GENDER has gained nuance precisely because disruptive pressures (such as feminist and queer critiques, technological developments in reproductive practices) prompted its reconceptualisation. In such cases, disruption is not merely endured or repaired, but becomes a constitutive part within the process of conceptual improvement.

Indeed, recognising antifragility allows us to better determine when resilience may be actively undesirable. If a concept is incomplete, rapidly evolving, and expected to face sustained disruption, (re-)engineering it for resilience may suppress the very pressures that would improve it. Attempting to lock such concepts into stable forms too early would risk entrenching their current inadequacies. In such cases, (re-)engineering for antifragility, i.e., encouraging disruption-driven evolution, is preferable to (re-)engineering for resilience. Conversely, if a concept is foundational, widely relied upon, and difficult to replace, then resilience may be preferable.

Ultimately, in the context of AI governance, the desirability of conceptual resilience varies significantly between thick and thin concepts, and depends on the concept's role within the broader conceptual scheme. While thick(er) concepts provide depth and specificity, they are more vulnerable to disruption. Conversely, thin(ner) concepts benefit from maintaining their applicability across varied and evolving circumstances. Thus, engineering thick concepts for resilience might be less desirable compared to thin concepts, which need to function effectively over time. Specifically, while thick concepts should not be engineered for high resilience due to their context-dependence and inflexibility, thin concepts benefit from resilience engineering that emphasises adaptation. This ensures that thin concepts can provide robust and enduring ethical guidance amidst the dynamic and rapidly evolving landscape of AI design, development, deployment, and use.

5.5.2.2 Conceptual clusters

Let us now consider conceptual clusters. Although conceptual clusters are relevant to the dynamics of conceptual disruption and are acknowledged within the literature as a significant intermediary level between individual concepts and the broader conceptual scheme (Löhr, 2023), they present unique definitional and methodological challenges for the investigation at hand.

On the one hand, the primary challenge lies in the relative indeterminacy and heterogeneity of conceptual clusters. Conceptual clusters can be loosely defined as groups of concepts that are interrelated or frequently co-occur within similar cognitive or linguistic contexts. Unlike individual concepts, which can be sharply identified as discrete units, or a conceptual scheme, which encompasses a comprehensive conceptual repertoires, clusters lack consistent defining features that would make them homogenous or reliably identifiable as cohesive entities, while also not being as consistently bounded as a conceptual scheme. The absence of clear structural or functional boundaries complicates their analysis as discrete units. This complicates efforts to evaluate resilience meaningfully at the cluster level, as a conceptual cluster's responses to disruption may vary significantly depending on the particular interplay of concepts involved.

On the other hand, the nature of conceptual clusters is fluid. Indeed, constituent concepts may fluctuate or overlap with other clusters based on varying contexts, and interrelations among concepts in a cluster can shift. This can be based (among others) on: (a) *contextual usage*, where concepts may align differently in varying social or disciplinary contexts; (b) *pragmatic goals*, since specific tasks or problem-solving contexts can make certain interrelations more salient; (c) *cultural norms*, as certain associations become more or less prominent across cultures; (d) *historical changes*, where evolving understandings or dominant paradigms alter the connections within a cluster; (e) *linguistic variations*, with different languages emphasising alternative groupings; (f) *technological and scientific developments*, as research findings reshape conceptual boundaries; (g) *political discourse*, where ideological perspectives can reconfigure conceptual alliances; (g) *aesthetic preferences*, where preferences in art or design influence how certain concepts align or cluster; (h) *emotional associations*, as concepts may cluster differently depending on affective connotations. The dynamic nature of clusters renders them less stable as objects of resilience assessment.

Because of this heterogeneity, while examining *disruption* at the level of a cluster can be illuminating, since multiple interconnected concepts may be disrupted simultaneously, there is no equally tractable notion of resilience that could reliably guide the assessment of the desirability of (re-)engineering conceptual clusters for resilience. For this reason, discussions of the desirability of conceptual resilience are better directed at the level of individual concepts and at the level of the conceptual scheme as a whole. Lastly, the case could be made that insights gained at the level of individual concepts and (perhaps especially) of the conceptual scheme may extend to clusters in specific instances. However, due to the lack of consistent structural or functional boundaries, any extrapolation of insights to the level of conceptual clusters would be tentative and speculative.

5.5.2.3 *Conceptual scheme*

Until this point, we have been discussing conceptual resilience in relation to individual concepts. Lastly, let us consider a conceptual scheme. The functional adequacy of a conceptual scheme concerns its capacity to structure an interconnected repertoire of concepts in a way that sustains coherent reasoning and supports reliable inference. A conceptual scheme therefore answers to systemic rather than local standards of adequacy, features that are not reducible to the properties of any particular concept. As conceptual resilience tracks continuous functional adequacy despite disturbances, and functional adequacy manifests differently in the case of individual concepts and a conceptual scheme, it seems plausible to expect that conceptual resilience itself should manifest differently for a conceptual scheme than for the individual concepts that compose it.⁹³

That is, the resilience of a conceptual scheme is not reducible to, nor exhausted by, the resilience of its constituent concepts; rather, it depends on the scheme's capacity to maintain coherent functional organisation under pressure. A resilient scheme need not (and should not) be one in which most concepts are themselves highly resilient. Instead, it requires an appropriate distribution of highly and scarcely

⁹³ For the purpose of this discussion, I will assume for illustrative reasons that each of the scheme's concepts is itself displaying a degree of resilience, whether high or low, that is appropriate to its function. This will enable us to better focus on the question of desirability: to what extent may a scheme composed of functionally adequate concepts, each with an appropriate degree of resilience, nevertheless exhibit a problematic degree of resilience?

resilient concepts: too many resilient (and therefore typically thinner) concepts render the scheme overly permissive, unable to capture context-specific distinctions or support fine-grained normative judgments; too few resilient concepts render the scheme brittle, prone to cascading failures when novel disruptions arise.

A scheme is resilient when its structural architecture accommodates both stability-preserving concepts that secure continuity across contexts and deliberately more fragile, thick concepts that anchor context-sensitive evaluation. The resilience of the scheme therefore derives from the way these heterogeneous concepts are integrated, not from any aggregate of the individual concepts' resilience. Therefore, for a scheme to continuously fulfil its function, it should be able to offer consistent normative grounding, which is especially valuable in contexts in which continuity in normative standards is crucial. Ultimately, it seems desirable to (re-)engineer our conceptual scheme for resilience, as some degree of resilience is necessary if a scheme is to retain its functional integrity across changing circumstances, and there is a benefit in such a continuity. Yet, conceptual engineers should be wary of an over- and under-resilient scheme.

On the one hand, an over-resilient scheme composed predominantly of highly resilient concepts (typically thinner, cross-contextual concepts) may become excessively coarse-grained, and preserve functional continuity at the cost of losing sensitivity to context-specific distinctions, thereby limiting its capacity to both serve a normative foundation for conduct and register emerging normative considerations. Indeed, a scheme containing too many highly resilient concepts tends toward structural over-stabilisation: excessive reliance on resilient concepts produces a conceptual architecture that preserves high-level evaluative orientation while lacking the descriptive articulation needed to register emerging normative distinctions. Thin concepts adapt easily to changing circumstances, but their very abstraction limits the scheme's diagnostic sensitivity: they maintain evaluative polarity while offering little traction for identifying new morally salient patterns. The result is a form of evaluative inertia. The scheme continues to make sense of the world in the same broad evaluative terms while becoming increasingly blind to the grounds on which revision may be warranted. In this sense, the scheme remains anchored in an inherited evaluative architecture because it lacks the thick conceptual resources needed to detect and articulate misalignment.

On the other hand, an under-resilient scheme composed of too few resilient concepts (predominantly thicker, context-sensitive concepts) may risk excessive rigidity, e.g., limiting its adaptability to new moral insights, and volatility. While scarcely resilience concepts are essential for fine-grained evaluation, a scheme built primarily out of such concepts lacks the stable evaluative anchors provided by more resilient concepts and becomes excessively sensitive to local perturbations, such that even modest contextual shifts can trigger large-scale conceptual disruption. Such a monolithic conceptual scheme where concepts are largely prone to disruption would be too susceptible to cascading disruptions and ultimately too volatile to provide stable foundations for continuous normative grounding (Hermann, 2019; Sauer et al, 2021; Sauer, 2023), as well as too rigid to allow for normative growth, innovation, and progress (Raz, 1994; Pleasants, 2018; Eriksen, 2020; Danaher & Hopster, 2022; Klenk et al, 2022; Danaher & Sætra, 2023).⁹⁴

Both extremes undermine the capacity of the scheme to sustain its function, preserving continuity where continuity is normatively valuable, and facilitating revision where normative insights warrant change. Ultimately, an intermediate approach is desirable, one that aims for an ecosystem of concepts lying at different points on the resilience spectrum, in which some concepts remain scarcely resilient and amenable to change, while others maintain a higher degree of resilience to offer a stable normative foundation. A resilient conceptual scheme therefore requires a calibrated distribution of concepts with varying degrees of

⁹⁴ Moreover, without sufficiently resilient concepts securing cross-contextual continuity, the normative landscape would be too easily unsettled by anticipated or perceived future pressures and the scene lacks stable anchors for deliberation. This echoes debates on moral progress about the risks and benefits of balancing the preservation of present normative values with the anticipation of future needs, to avoid the pitfalls of normative presentism and futurism (Swierstra, Stemerding, & Boenink, 2009; Boenink, Swierstra, & Stemerding, 2010; Swierstra, 2015). A conceptual scheme where all concepts are scarcely resilient risks falling into normative futurism, i.e., the perspective according to which concepts should continually change in anticipation of future challenges, without being grounded in stable values. This may lead to normative uncertainty, where concepts are in constant flux and cohesion is undermined. Having an ecosystem of concepts with different resilience levels avoids the pitfalls of both normative presentism (over-resilience) and normative futurism (under-resilience). Highly resilient concepts provide continuity and normative security, while scarcely resilient concepts allow for normative dynamism, enabling society to recalibrate its normative compass as needed.

resilience. Highly resilient concepts secure diachronic stability and provide orientation across contexts; scarcely resilient concepts supply the descriptive traction needed to identify new and emerging normatively salient considerations and justify targeted revision, so that our context-specific concepts can evolve to meet new challenges. This approach promotes a dynamic conceptual landscape that can accommodate both various degrees of resilience, ensuring that our concepts remain fit for their purposes in an ever-evolving socio-technical and normative landscape.

5.6 Conclusion

Recent work on conceptual engineering in the philosophy of technology increasingly focusses on conceptual disruption, understood as an interruption in functional adequacy following tensions and disturbances affecting concepts, and the need for conceptual engineers to remedy such disruptions. This paper addressed a complementary and neglected phenomenon, which we referred to as conceptual resilience, i.e., the ability of concepts to maintain functional adequacy despite tensions and disturbances. Contrary to existing work, which conceives resistance to disruption in absolute terms, we argued that conceptual resilience is best understood as a property that admits of degrees of resistance to disruption.

While such insights allow for intuitive assessments of whether a concept is (more or less) resistant to disruption, they do not, on their own, suffice to explain how a concept exhibits this property. To address this gap, we turned to neighbouring literatures that grapple with similar tensions, often framing resilience either as robustness—resistance to change—or as adaptability—flexible response to challenges. These distinctions, widely acknowledged in the literature on resilience, map well onto this heuristic of conceptual resilience, understood as immutability and adaptation.

Accordingly, we introduced two heuristic framings: Conceptual Resilience as Immutability (CRI) and Conceptual Resilience as Adaptation (CRA). Importantly, we do not consider such static and dynamic framings of resilience as mutually exclusive—rather, we argue that resilient concepts can simultaneously exhibit both immutability and adaptability, depending on the framing of one’s analysis—nor as jointly exhaustive. The concept GOOD provides a compelling example of how the same concept can be interpreted as embodying both static and dynamic

resilience, depending on one's framing. On the one hand, the concept's resilience can be construed in terms of the immutability of its essential definition, which embodies a stable, unchanging essence that persists across different contexts and times. On the other hand, the concept's resilience can be considered as a corollary of the concept's ability to adapt, by virtue of its lack of exhaustive definition, to its various instantiations.

Ultimately, what we aim to show is that conceptual engineers can fruitfully draw upon the conceptualisation of resilience that is prevalent in domains beyond the philosophical one. Consistently with the broader literature on resilience, we suggested that different approaches to understanding how concepts avoid disruption (through immutability or adaptation) should be seen not as distinct *types* of resilience, but as complementary analytic *lenses* or *frames* through which one can illuminate different dimensions of a unitary phenomenon.

Having clarified what resilience is and how it may be exhibited, we then turned to the issue of desirability: should conceptual engineers aim to (re-)engineer concepts *for* resilience and, if so, under what conditions? Two considerations are central here. First, the distinction between thick and thin normative concepts matters: resilience is more commonly found, and often more desirable, in thinner concepts, than in thicker ones. Second, the level of analysis matters: while resilience may be desirable for some individual concepts, a resilient conceptual scheme will typically combine both more resilient and less resilient concepts, depending on their functions and contexts of use. Conversely, conceptual clusters are too heterogeneous for any assessments of the desirability of resilience.

The implication is not that resilience is desirable or undesirable *tout court*, but that its desirability is conditional and resilience should not be pursued as a default. Instead, conceptual engineers should ensure that disruption is neither universally suppressed (as excessive robustness would do) nor universally amplified (as excessive antifragility would do). Specifically, (re-)engineering for resilience may be desirable in the case of some individual (thin) concepts, while it is generally undesirable in the case of thick(er) concepts, and is undesirable as a feature of all the concepts of a conceptual scheme, for the functional adequacy of which conceptual engineers should rather aim for a combination of more resilient and less resilient normative concepts. The goal of conceptual engineering should be to reach a state of equilibrium, ensuring that our conceptual repertoire can ensure some level of stability while also being

flexible enough to an ever-evolving socio-technical landscape and respond to new challenges. The desirability of a similar balance is particularly evident in the context of AI governance, where foundational ethical principles must coexist with context-specific concepts to ensure effective AI governance. To this end, it is reasonable to expect that conceptual engineers should aim to strike a balance between highly resilient thin concepts, able to provide solid foundations for normative thinking and withstand the test of time, and less resilient thick concepts, able to provide precise guidance within a specific socio-technical context.

PART II:

SYNTHESIS

All methods are wrong, but some are useful.

George E.P. Box

Chapter 6

NAVIGATING CONCEPTUAL DISRUPTION WITH CONCEPTUAL ENGINEERING. SYNTHESIS AND OPERATIONALISATION

6.1 Introduction — 6.2 Synthesis. Theoretical foundations — 6.2.1 Conceptual adequacy. Functions and affordances — 6.2.1.1 Concepts, conceptions, sub-concepts — 6.2.2 Conceptual disruption — 6.2.2.1 Levels, types, severity — 6.2.3 Conceptual engineering interventions — 6.2.3.1 Conceptual evaluation and conceptual preservation — 6.2.3.2 De novo engineering, conceptual re-engineering, and conceptual abandonment — 6.2.3.3 Conceptual implementation — 6.3 Operationalisation — 6.3.1 Case study. AUTHORSHIP and translation — 6.3.2 Case study. AUTHORSHIP and AI(-assisted) translation — 6.4 Conclusion

6.1 Introduction

Philosophers of technology are faced with a recurrent challenge: technological innovation frequently unsettles our concepts, causing them to no longer adequately fulfil their functions. This is known as a conceptual disruption (Löhr, 2022). The philosophy of technology scholarship has increasingly turned to conceptual engineering as an approach for addressing and bridging instances of conceptual disruption (Hopster, 2024; Hopster & Löhr, 2023; Hopster et al, 2023; Löhr, 2022; 2023a; 2023b; Marchiori & Scharp, 2024). This tends to operate from a broadly functional account of concepts, linking conceptual adequacy to the ability of a concept to fulfil their functions. However, by focusing on the uses that concepts are intended to fulfil or do fulfil, functionalism alone risks overlooking the broader range of actions and uses that concepts enable or constrain. To this end, it has been suggested that functionalism should be supplemented with the notion of conceptual affordances, pointing to the actions, uses, and thoughts, that a concept enables or constrains relative to a given user and within a given context. Synthesising and expanding on work at the intersection of conceptual disruption and conceptual engineering in the philosophy of technology, this closing chapter answers the question that motivated this investigation. That is, how should we (re)engineer concepts that have been disrupted in relation to socially disruptive technologies?

The chapter advances an affordances-informed functional framework to navigate conceptual disruption with conceptual engineering. It does so in two moves: synthesis and operationalisation. First, it synthesises insights from the literature and from previous chapters of this dissertation into the foundational dimensions that should guide such a process. Second, these dimensions are operationalised into a cyclical, iterative process, intended to guide conceptual engineers from the suspicion of conceptual disruption through to adequacy assessment, and conceptual engineering intervention. By synthesising theoretical insights into a structured process, this chapter provides a practical toolkit for navigating conceptual disruption in rapidly evolving socio-technical landscapes and makes conceptual engineering more operationalisable for philosophers of technology. Lastly, the chapter applies this cyclical process to two related cases involving the concept of AUTHORSHIP: first, the historical disruption created by human translation; second, the contemporary challenges posed by AI-assisted translation.

6.2 Synthesis. Theoretical foundations

6.2.1 Conceptual adequacy. Functions and affordances

The assessment of conceptual adequacy lies at the heart of any conceptual engineering project. To determine whether a concept calls for preservation, revision, replacement, or abandonment, we must first establish the criteria by which adequacy is to be judged. The dominant approach within conceptual engineering is functionalism. Functionalism about concepts holds that concepts are adequate to the extent that they fulfil their functions (Haslanger, 2000, 2020a; Chalmers, 2020; Thomasson, 2020; Simion & Kelp, 2020; Belleri, 2023). Different accounts have been advanced in the literature which offer competing views as to what these functions might or should be (Jorem, 2022; Nado, 2021; Queloz, 2019; Riggs, 2021), but they generally converge on the idea that concepts serve a range of purposes, such as epistemic, practical, and communicative aims, and that conceptual adequacy is (and should be) evaluated by identifying the relevant functions concepts should fulfil and assessing the extent to which concepts perform them. This is indicative of a broader commitment within the conceptual engineering literature, which similarly overwhelmingly operates from a functional account of conceptual adequacy (Belleri, 2021; Chalmers, 2020; Haslanger, 2000, 2020a, 2020b; Queloz, 2021; Plunkett & Sandell, 2013; Prinzing, 2018; Simion & Kelp, 2020; Thomasson, 2020, 2021).

This functional lens has guided much of the recent work on conceptual disruption in philosophy of technology. When scholars observe that concepts like RESPONSIBILITY or AUTONOMY may not be adequately fulfilling the roles they should be serving, such as in contexts involving autonomous AI systems, the diagnosis of disruption follows directly from identifying this functional breakdown (Hopster & Löhr, 2023; Hopster, 2024). Consider the concept of AUTHORSHIP. Functional analysis reveals that AUTHORSHIP serves multiple purposes across different domains. In legal contexts, it serves to allocate intellectual property rights, determining who may claim ownership over a creative work and who may benefit from its exploitation (Lauber-Rönsberg & Hetmank, 2019). In academic and literary contexts, it enables attributing ideas and expressions to specific individuals, thereby e.g. enabling proper acknowledgment and facilitating discourse about influence and originality (Berensmeyer et al., 2012; Mazzi, 2024; Nimmer, 2001). In cultural contexts, AUTHORSHIP carries symbolic weight, marking certain

Type	Description
Conceptual functions	The roles and purposes that a concept (should) fulfil
Conceptual affordances	The possibilities for action that a concept enables or constrains within a given context and relative to specific stakeholders

Table 6.1. Conceptual functions and affordances

forms of creative labour as prestigious and culturally valuable. These functions are not merely descriptive observations about how the concept happens to be used; they represent normative commitments about what the concept ought to enable. When we judge that AUTHORSHIP is functioning adequately, we mean that it successfully performs these functions, i.e., it allocates rights appropriately, attributes creative labour accurately, and structures cultural recognition fairly.

Yet functionalism alone, while essential, risks being incomplete. As Nado (2021) argues, focusing exclusively on a concept's functions may overlook the broader range of actions and uses that concepts enable or constrain in practice. A concept may fulfil its intended function while simultaneously enabling harmful uses, or it may fail its primary function yet serve valuable purposes that were not part of its original design. To address this limitation, I have proposed in chapter 4 that functionalism could be supplemented with attention to *conceptual affordances* (Marchiori, 2025). Drawing on Gibson's (1977a, 1977b, 1979, 1982) notion of affordances in ecological psychology—i.e., the possibilities for action that are available to agents in an environment—conceptual affordances refer to the potential actions, uses, and thoughts that a concept enables or constrains relative to a given user and within a given context (see Table 6.1).

The affordances framework introduces refinements to the functional analysis. First, it enables distinguishing between *functions*—i.e., the actions and uses that concepts should fulfil and do fulfil in practice—and *affordances*—i.e., the potential actions and uses that concepts enable relative to specific users and within a given context. This distinction helps revealing cases where concepts enable unintended uses or fail to deliver on their intended purposes. Second, the affordances lens emphasises the relational and contextual nature of conceptual adequacy. What a concept affords depends not only on its internal structure but on who is using it, for what purposes, and within what (e.g., institutional,

socio-technical) contexts.⁹⁵ The same concept may afford different possibilities to different users: for example, AUTHORSHIP may afford legal protection to established writers while affording exclusion to collaborative or non-traditional creators, such as ghostwriters.

Attending to conceptual affordances to assess conceptual adequacy also reveals dimensions of adequacy that functionalism alone might miss. For example, consider how, beyond its functions of allocating rights and attributing creative labour, AUTHORSHIP affords various practical and symbolic uses. Historically, the concept has afforded individual creators a form of legal and economic autonomy, enabling them to benefit from their work. Yet it has simultaneously constrained recognition of collaborative and derivative creative practices (e.g., translation, adaptation, compilation) by centring the notion of the solitary, original author (Venuti, 1995). These affordances had profound practical consequences. When translators contested their exclusion from authorial status, they were not merely arguing about whether the concept of AUTHORSHIP was fulfilling its function; they were challenging which creative practices the concept enabled or marginalised, and whether these affordances aligned with defensible normative commitments about recognition and compensation.

The historical debate over translators qualifying as authors illuminates the relationship between functions and affordances. Berensmeyer et al. (2012) identify a tension in concepts of AUTHORSHIP between what they term *“strong” author concepts*, in which the author is understood as autonomous creator and absolute ruler over meaning, and *“weak” author concepts*, in which authorship is more distributed, heteronomous, and embedded in networks of cultural production. This distinction maps onto different affordance profiles. Strong authorship concepts afford greater legal and symbolic protection to individual originators but constrain recognition of the collaborative and derivative dimensions of creative work. Weak authorship concepts afford more inclusive attribution but may dilute the legal protections and economic benefits that strong authorship concepts enables. Neither strong nor weak authorship concepts are simply adequate or inadequate in

⁹⁵ Similar considerations have been advanced in the conceptual engineering literature by proponents of a deflationary account of conceptual functions, according to which the types of conceptual functions that should be of interest to conceptual engineers and guide conceptual engineering projects are those determined on a case-by-case basis, depending on the context of use (Riggs, 2021; Jorem, 2022).

functional terms; rather, each of such concepts enables and constrains different possibilities for action, and such affordances should be normatively assessed relative to the values and goals of the communities employing the concept.

Love's (2002) taxonomy of "authemes" (i.e., precursory, executive, declarative, and revisionary authorship) further illustrates how attention to affordances enriches functional analysis and, relatedly, conceptual adequacy assessment. Love's (2002) framework recognises that "authorship" does not serve a monolithic function but points to a set of linked activities that may be performed by different individuals in different combinations. By disaggregating these activities, Love's (2002) approach enables a more fine-grained assessment of what particular conceptions of AUTHORSHIP afford. A concept that restricts authorship to executive creation (the physical act of composition) affords different forms of recognition and compensation than one that includes precursory authorship (incorporation of earlier work) or revisionary authorship (subsequent editing and refinement). The adequacy of any particular conceptualisation depends not merely on whether it performs "the function" of authorship but on whether the specific affordances it enables align with our normative judgments about which forms of creative labour merit recognition and protection. Integrating affordances into functional analysis thus provides a more comprehensive evaluative framework for conceptual adequacy.

6.2.1.1 Concepts, conceptions, sub-concepts

An important methodological question concerns the appropriate target for the framework advanced in this chapter and specifically whether context-dependent adequacy involves a single concept with different performances across contexts, or distinct concepts (or conceptions, or subconcepts) that may share a term. On one approach, AUTHORSHIP in legal contexts and AUTHORSHIP in cultural contexts is the same concept employed in different contexts. On this view, the concept might be adequate in one context while disrupted in another, and interventions must navigate trade-offs among different domains of application. The examples in this chapter adopt this approach: AUTHORSHIP is treated as a single concept that must perform multiple functions across legal, commercial, and socio-cultural contexts, and the analysis examines how well it performs these various functions and what tensions may arise when its ability to perform some functions adequately may hinder its

ability to adequately fulfil some other functions.⁹⁶ On another approach, 'authorship' may label different concepts. In this sense, LEGAL AUTHORSHIP and SOCIO-CULTURAL AUTHORSHIP are related but distinct concepts, and each must be assessed independently of the other. On this view, disruption in one context need not indicate disruption in another, and interventions can be domain-specific. A third possibility involves sub-concepts (Cohen & Murphy, 1984): AUTHORSHIP might be an overarching concept with subconcepts like LEGAL AUTHORSHIP, SOCIO-CULTURAL AUTHORSHIP, and COMMERCIAL AUTHORSHIP, where each subconcept inherits some structure from the parent concept while having domain-specific constitutive principles⁹⁷ and functional requirements. For example, LEGAL AUTHORSHIP might include constitutive principles about rights allocation that SOCIO-CULTURAL AUTHORSHIP lacks, while both share constitutive principles about creative contribution. Sub-concepts allow for both unity (they derive from and relate to a common concept) and differentiation (they have distinct adequacy conditions appropriate to their domains). The concept/conception distinction offers yet another way to capture this relationship (Lalumera, 2014; Rey, 2010; Woodfield, 1991): legal and cultural uses might involve different conceptions of a shared concept AUTHORSHIP, where conceptions are particular ways of understanding or implementing the concept that emphasise different features and serve different purposes while remaining recognisably about the same underlying phenomenon.

Ultimately, while I choose for illustrative purposes to focus on one of such views, the framework developed in this chapter is compatible with all these approaches. The stages of the cycle apply regardless of

⁹⁶ As will be illustrated in the case studies, while treating AUTHORSHIP as a single unified concept has advantages for illustrative purposes, it also carries a methodological cost worth addressing. That is, interventions targeting generic concepts motivated by conceptual disruptions in one domain may affect all domains in which the concept operates (van de Poel, 2025). This is problematic whenever a concept (such as AUTHORSHIP) structures practices across several domains, and a conceptual engineering intervention may solve the issue in the specific context under investigation while leading to unintended and undesirable consequences for (some of the) other practices that the concepts governs. This is one reason why the choice between unified concepts, conceptions, and sub-concepts matters not only to structure the discussion surrounding conceptual disruption, but also in terms of framing any subsequent conceptual engineering intervention.

⁹⁷ See section §2.3 for an illustration of constitutive principles.

whether the target are unified concepts, distinct concepts, sub-concepts, or conceptions. What matters is making explicit which functions concepts should perform in which contexts, assessing whether they perform them adequately, and which affordances they bear, and selecting interventions appropriate to the diagnosed problems (if any). The examples in sections §6.3.1 and §6.3.2 treat AUTHORSHIP as a single concept operating across contexts to illustrate how the framework handles cases where adequacy conditions vary by domain and interventions must balance competing functional requirements. However, the analysis could equally be reframed in terms of distinct conceptions of authorship, or sub-concepts of a super-concept AUTHORSHIP, without fundamentally altering the diagnostic and intervention approach the framework provides. Whether we treat context-dependent variation as involving: (a) a single concept applied across contexts with trade-offs in functional performance, (b) distinct concepts or conceptions assessed independently, (c) sub-concepts with partial inheritance and domain-specific requirements, or (d) conceptions of a concept, the key methodological commitment remains constant: adequacy must be assessed relative to specific functions, contexts, and stakeholder communities.

6.2.2 Conceptual disruption

The notion of conceptual disruption provides a critical diagnostic category for identifying when concepts may require a conceptual engineering intervention. First introduced by Löhr (2022) to capture classificatory uncertainty arising from technological change, the phenomenon has become central to discussions of how philosophy can respond to socially disruptive technologies (Löhr, 2023a, 2023b; Hopster & Löhr, 2023; Hopster et al., 2023). Chapter 2 illustrated a systematic account of conceptual disruption understood as the interruption in the normal functioning of concepts, individually or collectively (clusters of interrelated concepts, or an entire conceptual scheme) (Marchiori & Scharp, 2024). This definition explicitly links disruption to functionalism: a conceptual disruption occurs when concepts fail to adequately perform the functions we expect them to fulfil. The presence or suspicion of disruption thus serves as a signal that conceptual adequacy may be compromised and that the concept becomes a candidate for conceptual engineering intervention.

The diagnostic value of conceptual disruption is twofold. First, it provides an early warning system. When persistent disagreement arises

about how to apply a concept, when new practices strain existing definitions, or when normative unease emerges about a concept's current usage, these signals may indicate underlying functional breakdown. The suspicion of disruption directs attention to concepts that warrant closer examination, even when it remains unclear whether intervention will ultimately be required. Second, and crucially, the specific manner in which disruption manifests can inform the selection of appropriate responses. Different types of disruption point toward different underlying problems in the relationship between concepts and the phenomena they are meant to capture, and these different problems call for different conceptual engineering responses.

6.2.2.1 *Types, levels, severity*

In this chapter, I build on the phenomenon of conceptual disruption as presented in chapter 2, which distinguished two primary types of conceptual disruption: *conceptual gaps* and *conceptual conflicts* (see Table 6.2). Conceptual gaps occur when no existing concept adequately captures a phenomenon at the necessary level of precision. The phenomenon exists and requires conceptual articulation for purposes of e.g., reasoning, communication, or governance, but our current conceptual repertoire lacks the resources to represent it adequately. That is, many phenomena fit into categories at multiple levels of description, but a gap exists when we lack concepts at the relevant level of specificity for the aims at hand. For example, when astronomers hypothesised the existence of planets orbiting other stars, existing concepts like PLANET (restricted to objects orbiting our sun) and CELESTIAL OBJECT (too general) failed to capture the phenomenon at the appropriate level of precision, prompting the introduction of EXOPLANET.

The category of conceptual conflicts encompasses what earlier literature (Hopster & Löhr, 2023; Hopster et al., 2023) termed *conceptual overlaps*, i.e., cases in which multiple concepts seem to apply but generate classificatory uncertainty, and *conceptual misalignments*, i.e., cases in which concepts conflict with our values and norms. Unlike such accounts of overlaps and misalignments, conflicts can occur not only between multiple concepts but also within a single concept when its constitutive principles—which are part of the content of a concept and guide the application of the concept (Scharp, 2013)⁹⁸—prove mutually inconsistent

⁹⁸ See section §2.3 for an illustration of constitutive principles.

or inconsistent with norms or facts about the world (Marchiori & Scharp, 2024). For example, conceptual conflicts arise when there is some known fact that is inconsistent with constitutive principles for one or more concepts, e.g., when empirical discoveries, technological developments, or social changes reveal that a concept's constitutive principles are inconsistent either with norms or facts about the world, or with each other. When ventilator technology enabled patients to breathe artificially despite having no brain activity, this created a conflict: the constitutive principle that death involves cessation of breathing contradicted the constitutive principle that someone with a beating heart is not dead. These two principles could not both be maintained in light of the new socio-technical landscape.

Disruption type	Description
Conceptual gap	No existing concept adequately conveys a given phenomenon at the necessary level of precision
Conceptual conflict	Inconsistency: A) among constitutive principles of one or more concepts; B) between constitutive principles and facts about the world; C) between constitutive principles and values or social norms
Conceptual overlap	Type of conceptual conflict (A or B)
Conceptual misalignment	Type of conceptual conflict (C)
No disruption	The concept is adequate

Table 6.2. Conceptual disruption types

The concept of AUTHORSHIP has experienced disruptions of both types throughout its history. The case of translation initially presented as a conceptual conflict. On the one hand, translators were excluded from the category of authors by constitutive principles tying authorship to original composition rather than linguistic transformation of existing work. However, they were simultaneously included by other constitutive principles of authorship linking authorial creation with the displaying of what Love (2002) calls "executive authorship", i.e., the actual making of a text. As Venuti (1995) documents, copyright law resolved this conflict by maintaining a narrow conception of authorship that excluded translators, effectively choosing which constitutive principles to preserve at the cost of constraining recognition of forms of creative labour that many believed merited acknowledgment.

The historical development of authorship concepts also reveals how disruption can propagate to related concepts through inferential connections. The question of translation disrupted not only AUTHORSHIP but also related concepts including ORIGINALITY, CREATIVITY, and INTELLECTUAL PROPERTY (Nimmer, 2001). If translators could claim authorship, what constituted original creation? If translation involved creativity, how should creative labour be distinguished from mere technical skill? These questions indicate cluster-level disruption: the problem was not confined to a single concept but affected a network of interrelated concepts. Berensmeyer et al. (2012) observe that concepts of authorship have historically oscillated between "strong" and "weak" forms, with periods of transition marked by widespread uncertainty about the proper allocation of creative credit. Such transitions suggest moments when entire frameworks for understanding textual production undergo strain, indicating potential disruption at the conceptual scheme level.

The urgency of the conceptual engineering response also contributes to determining which conceptual engineering intervention is most suitable in a given situation. The urgency of conceptual engineering intervention depends on the severity of disruption, the stakes of inaction, and the availability of provisional solutions. As developed in Chapter 2, disruptions can be classified as mild, moderate, or severe (Marchiori & Scharp, 2024). Mild disruptions, typically gaps at the scheme level where we simply lack a needed concept, can be monitored⁹⁹ and addressed gradually, to the extent that the absence creates inconvenience but allows

⁹⁹ As mentioned in chapter 5 and as will be discussed when illustrating the cyclical approach to navigate conceptual disruption with conceptual engineering, conceptual engineering interventions are hardly (if ever) final, and rather proceed through gradual improvements. Namely, it is possible that an instance of conceptual disruption may become more problematic over time—such that a conceptual engineering intervention may not be deemed necessary from the outset, but will become more and more pressing e.g., as the extent of the disruption deepens or broadens—or that a conceptual engineering intervention may not successfully eradicate future instances of conceptual disruption. In this sense, ongoing monitoring may be warranted and desirable to ensure that changes in the adequacy of the concepts may be discovered sooner rather than later. This is not unlike monitoring after medical interventions: despite the medical concern being provisionally solved, it is often advisable to continue monitoring the situation for months or years to ensure that any lingering issues, side effects, or new manifestations of the issue being targeted will be spotted in a timely manner.

work to continue with approximations or workarounds. Moderate disruptions, such as gaps at the cluster or concept level, or conflicts that admit provisional accommodations, require intervention but do not demand immediate action, as partial solutions can mitigate problems while more adequate responses are developed. Severe disruptions, particularly conflicts where constitutive principles prove fundamentally inconsistent and no provisional solution can meaningfully bridge the gap, demand urgent intervention due to high costs of inaction.

The assessment of urgency must also consider the dynamics of technological change: when technologies develop rapidly and their deployment is widespread, disruptions that might otherwise be moderate may become severe due to the speed at which problems compound and the difficulty of implementing interventions once problematic practices become entrenched. The urgency dimension does not determine which intervention strategy to pursue (preservation, re-engineering, de novo engineering, or abandonment), but rather affects the timeline for deliberation and the acceptability of provisional measures. High urgency may require accepting imperfect solutions that can be implemented quickly, while low urgency permits extended evaluation and the development of more carefully considered responses. Importantly, urgency assessments themselves require ongoing monitoring: as contexts evolve and technologies develop, disruptions initially judged mild or moderate may become severe, necessitating escalation of intervention efforts.

Understanding conceptual disruption as a diagnostic signal that conceptual adequacy has been compromised and a conceptual engineering intervention may be needed thus requires attending to multiple dimensions, including the type of disruption (gap or conflict) and the level at which the disruption operates (concept, conceptual cluster, conceptual scheme). Moreover, the combination of the type and level of disruption impact the severity of the conceptual disruption, i.e., the urgency of a conceptual engineering intervention, which may extend from mild to severe (Marchiori & Scharp, 2024) (see Table 6.3).

	Level of disruption		Type of disruption	Severity of disruption
Conceptual disruption	Level 1	Conceptual scheme	Conceptual gap	Mild disruption
	Level 2	Conceptual cluster	Conceptual gap	Moderate disruption
			Conceptual conflict	Severe disruption
	Level 3	Individual concept	Conceptual gap	Moderate disruption
			Conceptual conflict	Severe disruption

Table 6.3. Conceptual disruption. Levels, types, severity (from Marchiori & Scharp, 2024)

While these dimensions do not mandate specific intervention strategies, they provide helpful insights to structure the considerations informing a suitable conceptual engineering response.

6.2.3 Conceptual engineering interventions

Once disruption has been diagnosed, the question becomes: what should be done? Conceptual engineering offers a range of intervention strategies, each appropriate to different types and severities of disruption. The conceptual engineering literature distinguishes several core interventions (Cappelen & Plunkett, 2020; Chalmers, 2020; Thomasson, 2020, 2021): *preservation*, i.e., retaining a concept as is; *evaluation*, i.e., subjecting it to normative analysis, which precedes, follows, or guides each of the other activities; *de novo engineering*, i.e., introducing a new concept; *re-engineering*, i.e., revising an existing concept; *abandonment*, i.e., eliminating a concept from use; and *implementation*, i.e., instantiating concepts in socio-technical and institutional practices and landscapes (see Table 6.4).

Intervention	Description	Example
Preservation	The concept is retained as is	E.g., "squirrel", "flower"
Evaluation	Normative analysis	E.g., "authorship" vis-à-vis translation
De novo engineering	Design of a new concept	E.g., "brain death"
Re-engineering	Re-design of a concept	E.g., "planet"
Abandonment	The concept is eliminated	E.g., racial slurs
Implementation	Uptake of the conceptual engineering proposal	E.g., "strict moral liability"

Table 6.4. Conceptual engineering interventions

While this taxonomy is now standard, what remains less developed is systematic guidance for matching interventions to specific disruption profiles. Importantly, such interventions are not to be understood as alternatives among which only one should be chosen, but rather as different ways of performing conceptual engineering, each of which can be more or less suitable for a given situation, and many of which may be used at different stages throughout a singular conceptual engineering project.

6.2.3.1 *Conceptual evaluation and conceptual preservation*

Before turning to interventions that (re-)engineer or replace concepts, it is important to recognise two baseline responses that apply when either conceptual disruption or a suitable conceptual engineering response are absent or uncertain. First, conceptual evaluation, i.e., subjecting concepts to normative analysis, plays a crucial role in determining which conceptual engineering intervention should be pursued, and becomes especially important when disruption is suspected but not yet confirmed, as well as to determine a suitable intervention strategy when disruption is present. For example, in the situation of a suspicion of disruption, evaluation enables systematic investigation of whether adequacy has genuinely been compromised. This may involve mapping the concept's current functions and affordances, consulting stakeholders about difficulties they encounter, examining how the concept performs across different contexts, and assessing whether provisional accommodations might suffice. Evaluation is also appropriate when disruption is confirmed but its type, level, and severity remain unclear, or when the stakes of intervention are high enough to warrant extended deliberation. In practice, evaluation should be understood as an ongoing component of responsible conceptual engineering, including conceptual maintenance post intervention: even concepts that currently appear adequate merit periodic reassessment as contexts evolve and new challenges emerge.

Second, not every suspicion of disruption indicates that a conceptual disruption is present and that it requiring intervention. For example, disagreements about application may reflect legitimate ambiguity in borderline cases rather than conceptual disruptions, normative unease may stem from misuse of an adequate concept rather than inadequacy of the concept itself, and new phenomena may be adequately captured by existing concepts at appropriate levels of precision, even if initial uncertainty creates the impression of a gap. In these cases, when evaluation reveals that a suspected disruption is unfounded and a concept continues to fulfil its functions adequately, making intervention unnecessary or counterproductive, conceptual preservation, i.e., retaining a concept as is, is the appropriate response.

6.2.3.2 *De novo engineering, conceptual re-engineering, and conceptual abandonment*

Let us now consider cases in which evaluation confirms that disruption is present and intervention is warranted. Though not to be taken as an invariable rule, some considerations can be drawn from disruption types and recommended conceptual engineering interventions. As suggested in chapter 2, when a disruption manifests as a conceptual gap, this typically calls for *de novo* engineering: when no existing concept adequately captures a phenomenon at the necessary level of precision, introduction rather than revision is required (Marchiori & Scharp, 2024). The phenomenon of sexual harassment, for example, existed as a pattern of behaviour long before the concept `SEXUAL HARASSMENT` was coined in the 1970s (Congdon, 2024). Prior concepts (such as flirtation, rudeness, assault) captured aspects of the phenomenon but missed its distinctive character as a form of gender- or sex-based discrimination that need not involve physical contact or explicit *quid pro quo* demands. The introduction of `SEXUAL HARASSMENT` as a distinct concept enabled recognition, discussion, and regulation of a form of harm that existing concepts obscured (Haslanger, 2000). Similarly, when exoplanets were hypothesised, existing concepts like `PLANET` and `CELESTIAL OBJECT` applied at some level of description but lacked the appropriate level of precision, warranting introduction of a new concept rather than revision of existing ones (Marchiori & Scharp, 2024).

Conceptual conflicts, by contrast, may call for re-engineering, *de novo* engineering, or abandonment. The connection between types of conceptual conflicts and intervention strategies can be made more precise by attending to the locus of the conflict. When a conflict is internal to a single concept, i.e., when its constitutive principles are mutually inconsistent, re-engineering and abandonment are the most suitable available responses. For example, abandonment becomes appropriate when conflicts run so deep that revision cannot salvage the concept. Chapter 2 suggests that some conceptual conflicts reveal not merely false constitutive principles but principles that conflict fundamentally with our values and norms (i.e., a conceptual misalignment; Hopster & Löhr, 2023; Hopster et al, 2023). In such cases, re-engineering may be impossible without discontinuity, i.e., without changing the concept so radically that it ceases to be the same concept (Cappelen & Plunkett, 2020). Slurs provide a clear illustration: the primary function of slurs are to demean and subordinate, and these functions appear built into the concepts in

ways that resist amelioration. Attempting to re-engineer a slur to remove its demeaning function while preserving some classificatory role faces the question of whether the resulting concept would retain sufficient continuity with the original to count as revision rather than replacement. When a concept's functions are harmful and undesirable, abandonment may be the only viable response. Moreover, technology can sometimes render concepts obsolete by making the phenomena they track disappear or by revealing inadequacies stemming from historical contingent practices that have since been updated, prompting discontinuation rather than revision or replacement.

When a conflict arises between two or more concepts whose constitutive principles are individually consistent but mutually incompatible, the range of viable conceptual engineering responses widens: re-engineering, de novo engineering, and abandonment are all available, as the disruption points to a concern that is relational, rather than internal to any one single concept. For example, one may respond to conceptual conflicts with de novo engineering, by opting to design a new concept to address the conflict, instead of trying to bridge the conflict by re-engineering one or more of the concepts involved. This choice is illustrated in the case of BRAIN DEATH.

Lastly, when a conflict manifests between a single concept and facts, norms, or values, all three responses are available (re-engineering, de novo engineering, and abandonment), but they suit different configurations of the problem.¹⁰⁰ When dealing with a *concept-fact conflict*, re-engineering may be appropriate when the conflict is pervasive enough that the concept may need to be revised across the entire domain to remain useful, but is not so inadequate that revisions may not salvage it (thus calling for abandonment). De novo engineering may be appropriate when the conflict is localised in such a way that a new concept may address and bridge the disruption, while the original concept can continue to fulfil its function(s) in a more limited set of cases. For example, when discoveries of numerous Kuiper belt objects created conflict with the constitutive principle that there are only a few planets, astronomers re-engineered PLANET by adding a new constitutive principle (planets must clear debris from their orbital neighbourhood),

¹⁰⁰ It may also be the case that conceptual disruptions may also be bridged by approaches that do not target the concepts themselves, but the broader socio-technical landscape in which they are used.

thereby eliminating the conflict while maintaining the concept's classificatory function.

Conversely, when a concept conflicts with norms or values, i.e., when the functions or affordances of a concept (or both) are something that we have reason to disavow, re-engineering the concept or abandoning it seem the only available conceptual engineering interventions. Indeed, as the issue manifests as an existing concept actively producing outcomes (e.g., classifications, inferences, affordances) that are incompatible with our normative commitments, introducing a new concept through de novo engineering does not neutralise the impact of the existing concept. Discriminatory concepts illustrate this case: to the extent that the discrimination is a contingent feature of the concept, the concept may be re-engineered so as to continue to fulfil its function without perpetuating the discriminatory outcomes. For example, the historically gendered affordances of a concept like HUMAN KIND, which has contingently come to centre male experiences as the human universal, thereby marginalising non-male perspectives, can in principle be re-engineered to feature such exclusionary affordances less prominently. Conversely, if the discrimination is foundational to a concept, re-engineering is no longer available as it would lead to a new concept altogether. In that case, abandonment becomes the best response. A racial slur, for example, cannot be re-engineered to remove its demeaning function without radically changing it, since the subordination of its target group is not an incidental feature of the concept but constitutive of what the concept does and should do.

6.2.3.3 *Conceptual implementation*

Conceptual implementation deserves attention as a follow-up of re-engineering, de novo engineering, or abandonment, particularly in contexts involving socially disruptive technologies. While the interventions discussed above focus on the conceptual work of preservation, revision, introduction, or elimination, implementation concerns the practical work of instantiating concepts e.g., in institutional practices, legal frameworks, and socio-technical landscapes. That is, a concept may be adequately engineered in principle yet fail to fulfil its functions in practice due to inadequate implementation mechanisms. For example, implementation may involve designing and applying

procedures and educational programs to increase concept uptake.¹⁰¹ For instance, implementing a revised concept of INFORMED CONSENT in contexts involving AI-driven medical diagnosis requires not merely conceptual clarity about what consent entails but also interface design that makes relevant information comprehensible, institutional protocols that ensure adequate deliberation time, and monitoring systems that verify compliance. In such contexts, conceptual implementation may require coordination between conceptual engineering and complex system design, as concepts must be operationalisable within existing technological infrastructures, and the socio-technical landscape may need modification to enable functional adequacy of the concept. In some cases, a concept may be suitably (re-)engineered but inadequately implemented, leading to an implementation gap.¹⁰² Other times, implementation challenges may reveal that engineered concepts are not sufficiently operationalisable, requiring iteration back to earlier stages of the cycle. Implementation should thus be understood both as a distinct intervention phase following conceptual engineering work and as an ongoing process requiring monitoring and refinement as contexts evolve.

6.3 Operationalisation

I propose that the insights from the literature on conceptual disruption and conceptual engineering synthesised in the previous sections are best operationalised through a cyclical approach. The cycle is intended as a structured approach that makes explicit the reasoning required at each phase of a conceptual engineering project while preserving the flexibility necessary for context-sensitive judgment.

¹⁰¹ In the literature, some have argued that the implementation of concepts is fundamentally difficult or even impossible (Gibbons, 2021; Jorem, 2021; Nimtz, 2021; 2024; Queloz & Bieber, 2022). This is known as the implementation challenge, and is often addressed as one of the main limitations of conceptual engineering projects. That is, even if conceptual engineers succeed to (re-)design improved concepts, the whole conceptual engineering enterprise may be undermined if such conceptual proposals cannot be adequately operationalised and the uptake remains limited.

¹⁰² Veluwenkamp & van den Hoven (2023) address this concern and suggest that the implementation challenge may be mitigated by embedding concepts within the design of technological artefacts.

The procedural approach developed here draws inspiration from the ETICA framework (Stahl, 2011; Stahl et al., 2012), originally developed for ethical assessment of emerging information and communication technologies but applicable more broadly to (anticipatory) technology ethics (e.g., Brey, 2012). The ETICA approach distinguishes between identification and evaluation stages, recognising that normative assessment must be preceded by careful descriptive work to specify what is being assessed and under what conditions. This distinction maps onto the relationship between conceptual analysis and conceptual engineering (Cappelen & Plunkett, 2020): conceptual engineering is normative conceptual work, but it presupposes descriptive conceptual work, such as understanding of how concepts currently function and how they relate to other concepts and to the phenomena they purport to capture. The cycle operationalised here structures both the descriptive and normative dimensions of conceptual engineering work, specifying what information must be gathered, what judgments must be made, and how these elements fit together to support suitable intervention decisions.

The seven stages of the cycle are: (1) *suspicion of disruption*, recognising signals that a concept may no longer be functioning adequately; (2) *context and stakeholder specification*, delimiting the scope of the inquiry, identifying the domains in which the concept operates and the relevant stakeholders; (3) *functional and affordance mapping*, mapping the functions the concept fulfils and the affordances the concept bears; (4) *adequacy and disruption assessment*, evaluating whether the mapped functions and affordances are desirable, determining the desirable functional and affordance profile of the concept, and diagnosing the type and severity of any disruption present; (5) *conceptual engineering intervention*, selecting and specifying an appropriate conceptual engineering response informed by the disruption assessment; (6) *conceptual implementation*, implementing the (re-)engineered concept within the relevant context; and (7) *evaluation and monitoring*, assessing whether the intervention and its implementation have restored adequacy and (if deemed necessary) establishing procedures for ongoing monitoring. These stages are arranged sequentially, the output of each stage provides input for the next, but the cycle is recursive rather than linear. For example, developments at later stages may require returning to earlier steps to revise initial judgments, gather additional information, or reconsider adequacy criteria in light of what has been learned through attempted intervention.

This recursive structure reflects important features of conceptual engineering as a practice. First, conceptual adequacy is not static. Technological developments, social changes, and shifts in normative commitments can alter the context in which concepts operate, potentially undermining interventions that initially restored adequacy or revealing new disruptions in concepts previously judged unproblematic. Ongoing monitoring is therefore not merely a check on whether an intervention succeeded but an integral component of responsible conceptual maintenance. Second, the relationship between concepts and the phenomena they capture is often complex and partially obscure. Initial suspicions of disruption may prove unfounded upon closer analysis, or the type of disruption may differ from what early signals suggested. As illustrated in chapter 2, it is also possible that instances of disruption may lead to new and different forms of disruption following conceptual engineering interventions (such as conflicts leading to gaps¹⁰³), ultimately prompting the need to start the process anew. Relatedly, this approach must allow for revision of diagnoses and responses as understanding deepens and circumstances change. Third, conceptual engineering interventions are often contested. Different stakeholders may have divergent views about which affordances are desirable, whether disruption is present, and what response is appropriate. Making each stage explicit facilitates transparency about where disagreement arises and what considerations inform different judgments, enabling more productive debate about how to proceed.

The cycle is illustrated visually in Figure 6.1, which depicts the seven stages arranged in a circle with arrows indicating the primary flow of reasoning. Dotted lines from later stages back to earlier ones represent potential feedback loops: adequacy and disruption assessment may prompt reconsideration of context and stakeholder specification if the diagnosis reveals that relevant stakeholders or domains were initially overlooked; at the conceptual intervention stage, it may be required to return to the adequacy and disruption assessment if proposed responses would alter the functions and affordances the concept should fulfil;

¹⁰³ In chapter 2, it was argued that re-engineering to resolve a conflict often creates a new gap. For example, the revised concept of PLANET no longer applied to Pluto and similar objects, generating a gap that prompted de novo engineering of DWARF PLANET. This dynamic relationship, where conflict resolution through re-engineering produces gaps that require introduction of new concepts, recurs across multiple cases and informs how conceptual disruptions cascade and propagate.

evaluation may reveal that disruption was misdiagnosed or that the intervention introduced new problems, necessitating a return to assessment and (re)design. The cyclical arrangement emphasises that conceptual engineering is not a one-time fix but an ongoing process of refinement and adjustment as concepts, technologies, and the socio-technical landscape they operate in evolve and co-shape each other.

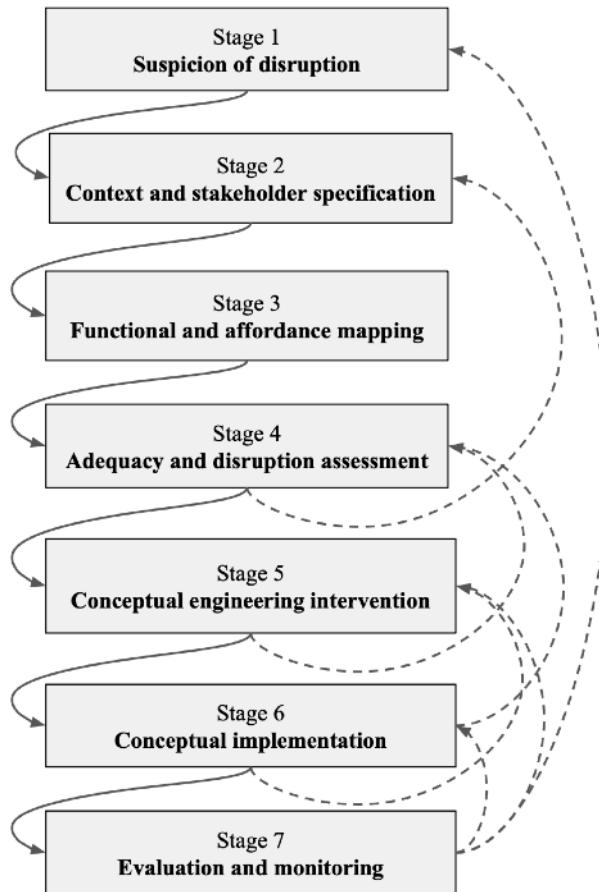


Figure 6.1. Navigating conceptual disruption with conceptual engineering. Cycle

To demonstrate how this cycle operates in practice, the following sections apply it to the example of AUTHORSHIP, examining the historical disruption surrounding translation, walking through each stage of the cycle to reconstruct how the disruption might have been diagnosed and addressed using this framework. This historical application illustrates how this approach can be operationalised in a case where the outcome is known, allowing assessment of whether the approach would have supported appropriate intervention; it also establishes a baseline for comparison with the second example, where the approach is applied to current questions about AI and authorship, this time operating speculatively to explore how the cycle might guide intervention design for an ongoing, unresolved suspected disruption.

Before proceeding, two methodological notes are warranted. First, the historical reconstruction to follow is necessarily interpretive and simplifying a rich debate (e.g., Bennett, 2020; Garzelli, 2011; Jansen, 2019; Lee, 2020; Malkov, 2024; Pagano et al, 2024; Summers, 2017; Venuti, 1995; Yan & Du, 2020; Zeller, 2000). The analysis in this chapter aims to reconstruct how the problem might have been approached using the approach illustrated here. The goal is not historical accuracy of the actual deliberation processes but illustration of how the approach structures reasoning about disruption and conceptual engineering intervention. Second, the contemporary application to follow is explicitly speculative. The AI and authorship case is ongoing and contested, such that it may even be contested whether a *suspicion* of disruption is present. Ultimately, the analysis explores possible pathways through the cycle rather than prescribing a definitive answer, demonstrating how this approach can guide reasoning even in conditions of uncertainty and disagreement, and structure deliberation to make reasoning, considerations, and trade-offs transparent and explicit.

6.3.1 Case study. AUTHORSHIP and translation

The question of whether translators qualify as authors has a long and contested history in (among others) copyright law, literary theory, and publishing practice. While translators have always performed creative labour, conceptual frameworks have often denied them authorial status, treating translation as reproduction rather than creation. This section reconstructs how the conceptual disruption surrounding the concept of authorship vis-à-vis the practice of translation might have been navigated using the seven-stage approach developed above. The

reconstruction draws on historical scholarship about the development of copyright law and concepts of authorship (Berensmeyer et al., 2012; Lee, 2020; Love, 2002; Venuti, 1995) to illustrate how each stage of the cycle would structure analysis and inform intervention.

Stage 1. Suspicion of disruption

The suspicion that AUTHORSHIP as applied to translation might be disrupted arose gradually as translation became increasingly central to literary production and economic exchange. By the eighteenth and nineteenth centuries, the proliferation of translated works across European markets generated practical questions that existing concepts struggled to address (Tahir Gürçağlar, 2010). Publishers faced uncertainty about who held rights to translated texts. Translators contested their exclusion from economic benefits and reputational credit. Legal disputes arose when multiple translators produced competing versions of the same source text, raising questions about whether later translations infringed on earlier ones. These practical difficulties, persistent disagreement about classification, uncertainty in rights allocation, normative unease about recognition and compensation, constituted signals of possible conceptual disruption. That is, the modern concept of AUTHORSHIP considered in this case study, developed primarily around paradigm cases of original composition, did not seem to perform its function adequately when applied to translation contexts, which gave rise to suspicion of disruption (Garzelli, 2011). It seems reasonable to expect that some stakeholders, e.g., publishers and original authors who benefited from existing arrangements, saw no problem requiring conceptual attention (e.g., Yan & Du, 2020). The dominant concept of AUTHORSHIP continued to perform its legal function of allocating rights, even if this allocation systematically disadvantaged translators. Translators, by contrast, experienced what they perceived as inadequacy of existing concepts acutely: their labour was economically and culturally devalued precisely because the dominant concept of AUTHORSHIP

excluded them from the protections and recognitions it afforded.¹⁰⁴ Whether this exclusion represented a form of conceptual inadequacy or a defensible normative choice about which forms of creative labour authorship should recognise was itself part of what was contested, and this framework does not presuppose an answer in this regard.

Stage 2. Context and stakeholder specification

Translation operated across multiple overlapping contexts, each with distinct stakeholders and distinct functional requirements for the concept of AUTHORSHIP. In legal contexts, AUTHORSHIP functioned to allocate intellectual property rights under emerging copyright regimes. The Statute of Anne (1710) in England and subsequent copyright legislation in other jurisdictions established that authors held exclusive rights to print and sell their works, but left ambiguous whether translators counted as authors of their translations. Publishers, original authors, translators, and courts all had stakes in how this question was resolved, and the function of clear rights allocation was paramount. In literary and academic contexts, AUTHORSHIP served to attribute ideas, establish reputational credit, and structure critical discourse about influence and originality (Nimmer, 2001). Here, the function of accurate attribution was central. Translators, literary critics, readers, and scholars participated in debates about the proper recognition of translation as creative achievement. In commercial contexts, AUTHORSHIP structured market categories and compensation practices, with economic functions driving stakeholder interests. Publishers, booksellers, translators, and reading publics all operated with implicit or explicit assumptions about the relationship between original and translated texts.

These contexts were not isolated; decisions in one domain affected others. Legal determinations about copyright influenced publishing economics, which in turn shaped cultural valuations of translation labour. The stakeholders were diverse and their interests sometimes conflicted: publishers might prefer to deny translators

¹⁰⁴ It seems plausible that a similar asymmetry in suspicion may be characteristic of many conceptual disruptions, with those disadvantaged by existing conceptual arrangements being typically first to diagnose an inadequacy, while those who benefit from the status quo having incentives to preserve the concept as is. Importantly, while this points towards a suspicion of disruption, it does not yet allow to establish the extent to which a grievance is normatively justified and a concept genuinely inadequate.

authorial status to avoid sharing profits or navigating complex rights arrangements, while translators sought recognition to improve their economic position and cultural standing. Original authors had mixed interests: some valued translators' work and supported their claims, while others worried that acknowledging translators as authors might dilute their own ownership or control. Specifying this complex context was essential for understanding what functions AUTHORSHIP needed to perform across different domains and what was at stake in how the disruption was addressed.

Stage 3. Functional and affordance mapping

To map the functional affordances profile of AUTHORSHIP in translation contexts, it is necessary to specify what functions the concept performed across relevant domains and what it enabled or constrained for different stakeholders. In legal contexts, AUTHORSHIP served to (among others) allocate property rights in ways that incentivised creative production and enabled the enjoyment of the fruits of one's labour as well as orderly market transactions (Gosseries, 2008; Hettinger, 1989; Tavani, 2017). The concept enabled clear determination of who holds what rights and who may benefit from exploitation of creative works. In literary contexts, functions included accurately attributing creative labour, facilitating critical discussion of textual relationships and influence, and recognising achievements deserving of cultural esteem. Here, the concept enabled proper acknowledgment of contributions and structure discourse about creativity and originality (Nimmer, 2001). In commercial contexts, functions included coordinating market transactions, structuring compensation practices, and managing relationships between creators, intermediaries, and audiences.

Attending to the affordances of AUTHORSHIP alongside its functions enriches this mapping, by illustrating which practices the concept enabled and foreclosed. The dominant conception of AUTHORSHIP afforded (especially, individual) original creators legal protection as well as economic benefits and social prestige, while constraining recognition of collaborative and derivative creative practices. In particular, this afforded clear right allocations and (economic) control over derivative works, which benefited publishers and original authors. For translators, the concept often afforded neither legal protection nor social recognition, effectively cloaking their creative contributions.

The functional mapping revealed that AUTHORSHIP was performing different and competing functions for different stakeholders. For some, particularly translators, the concept performed the functions of recognising all forms of creative labour, affording inclusive attribution, and ensuring fair compensation for skilled linguistic work. For others, including original authors and publishers, the concept performed functions of marking original creation as special, maintaining clear property boundaries, and protecting the economic interests of original creators. This maps onto what Berensmeyer et al. (2012) identify a tension between "strong" (centring autonomous individual creativity) and "weak" (recognising distributed, collaborative creative processes) authorship concepts. "Strong" authorship concepts perform functions that protect individual originators and mark originality as a special achievement. "Weak" authorship concepts perform functions that recognise diverse contributions and distribute credit more inclusively. Neither concept performed the full range of functions across all stakeholders. The affordances lens adds precision to this picture, by making explicit the possibilities for action and thought that different conceptualisations of AUTHORSHIP enable and constrain, and more clearly identifying what is at stake when choosing one over the other.

Stage 4. Adequacy and disruption assessment

Assessing conceptual adequacy requires first determining which functions AUTHORSHIP should perform across the relevant contexts.¹⁰⁵ From the perspective of translators and translation advocates, it was crucial for AUTHORSHIP to perform functions of recognising and compensating translation labour. Venuti (1995) documents extensive

¹⁰⁵ The claim that conceptual adequacy is assessed based on which functions a given concept should fulfil is not intended to invite a subjectivist reading, whereby a concept is deemed inadequate simply because it produces outcomes a given analyst or stakeholder dislikes. Indeed, how to determine which functions should serve as the standard against which adequacy is assessed in the first place is a normative exercise, and not every function-specification is equally defensible. The task of conceptual engineering is to adjudicate between competing specifications on principled grounds. A full account of how the relevant functions are to be selected and how competing function-specifications are to be adjudicated (including questions of procedure, authority, and legitimacy) remains an open question in the literature and will be the object of future work (e.g., Jorem, 2022; Marchiori, 2025; Nado, 2021; Riggs, 2021).

arguments that translation involves creativity, skill, and interpretive judgment comparable to original composition, and that the concept's failure to afford translators recognition represented functional inadequacy. The concept performed its legal function of allocating rights, but did so in ways that failed to recognise a form of creative contribution that many believed merited acknowledgment. From the perspective of original authors and publishers operating under strong individualistic conceptions of authorship, the concept was functioning adequately: it allocated rights clearly, coordinated market transactions efficiently, and preserved the special status of original creation that the concept was intended to mark (e.g., Wilson & Gerber, 2012). On this view, the exclusion of translators was not a failure but a defensible normative choice about the scope and purpose of authorial recognition. The disagreement was not about whether the functions of the concept were being performed, but about which functions the concept should perform and whether its current functional profile aligned with defensible societal values.

Assessment of whether AUTHORSHIP satisfied adequacy conditions revealed a pattern of contested functional performance. In legal contexts, the concept functioned to allocate rights, but did so in ways that many judged unfair: translators performed creative labour comparable to composition but were denied the economic benefits and legal protections that AUTHORSHIP afforded to original authors. The concept successfully supported market transactions by providing clear rules about ownership, but these rules failed to perform functions of recognising creative contribution that many believed important. In literary and cultural contexts, the dominant concept of AUTHORSHIP performed functions of marking original composition as a special achievement but failed to perform functions of recognising translation as creative work deserving acknowledgment. While some translators gained reputational credit, translation labour was often cloaked and translators were often rendered invisible in discussions of literary production.

The translation case exhibited primarily conceptual conflict. The constitutive principle that authors create original works, i.e., that authorship requires origination rather than transformation, conflicted with the empirical fact that translators engage in creative decision-making, exercise interpretive judgment, and produce texts through skilled linguistic labour. Indeed, translation shared features with both original authorship (creative decisions, textual production, expression) and non-authorial activities (derivation from pre-existing content,

linguistic transformation rather than origination). The concept AUTHORSHIP, structured around paradigm cases of original composition, contained constitutive principles that excluded translation despite its creative dimensions. Venuti (1995) characterises the development of exclusive authorial copyright as depending on "an individualistic concept of original authorship that negates the translator's work", suggesting that the legal concept's constitutive principles were structured to exclude translators rather than being merely ambiguous about their status.¹⁰⁶

The disruption manifested primarily at the individual concept level, with AUTHORSHIP itself being the focus of contestation, though cluster-level disruptions were also present. Indeed, questions about translators' status prompted adjacent questions about ORIGINALITY (does translation involve original creation or mere reproduction?), CREATIVITY (is linguistic skill in translation a form of creativity or technical competence?), and DERIVATIVE WORK (should translations be understood as derivative and therefore subordinate to original texts?). The severity of disruption was moderate. Translators experienced genuine harm through economic disadvantage and lack of recognition, and the inadequacy of existing concepts created inefficiencies in rights management and coordination problems in literary markets. However, literary production continued, partial accommodations emerged through contractual arrangements and informal norms, and the problems did not rise to the level of crisis that would demand immediate intervention. The concept continued to perform core legal and economic functions even as it failed to perform functions of inclusive recognition that translation advocates judged important, though whether those functions were ones that AUTHORSHIP should have been performing remained contested.

Stage 5. Conceptual engineering intervention

Given the diagnosis of conceptual conflict at moderate severity, several intervention strategies were available. Preservation, retaining AUTHORSHIP in its narrow form centred on original composition, would have maintained existing functions for clear rights allocation and economic predictability, while perpetuating the exclusion of translators

¹⁰⁶ Whether this exclusion reflected a defensible commitment to marking original creation as a distinctively valuable achievement or an unjustifiable marginalisation of creative labour, is precisely the type of question that this stage of the cycle aims to address.

which generated normative concerns. This was effectively the path initially taken: copyright law explicitly denied translators authorial status in many jurisdictions through the eighteenth and nineteenth centuries, preserving the concept's performance of legal and economic functions while accepting its failure to perform functions of inclusive recognition. Re-engineering AUTHORSHIP to include translators would have addressed the conflict by revising constitutive principles to encompass both original composition and creative linguistic transformation. This would enable the concept to recognise translators, but at the potential cost to the concept's capacity to mark original creation as distinctively valuable—a function that some stakeholders judged central to AUTHORSHIP's purpose. This could have taken various forms: revising the concept to encompass "collective authorship" as Venuti (1995) suggests, distinguishing "degrees" or "types" of authorship to accommodate both original and derivative creation, or broadening the notion of originality to include the creative dimensions of translation. Re-engineering would enable the concept to perform recognition functions while potentially complicating its legal and economic functions by creating ambiguity about rights allocation.

De novo engineering, introducing new concepts alongside AUTHORSHIP, offered a middle path, allowing for the creation of concepts like DERIVATIVE WORK or TRANSLATOR'S RIGHTS that would enable recognition and protection for translators without collapsing the distinction between original and translated texts. This strategy preserves AUTHORSHIP's existing functions while introducing new concepts to perform functions that AUTHORSHIP failed to fulfil, thus offering a middle path that could avoid forcing a choice between competing normative commitment: AUTHORSHIP could retain its existing scope and functions, while new concepts could be introduced to provide coverage for the functions that AUTHORSHIP's current conceptualisation left unserved.¹⁰⁷ Abandonment was not seriously considered: AUTHORSHIP continued to perform important legal and economic functions, and the disruption did

¹⁰⁷ It is also worth noting that (re-)engineering concepts to perform functions that existing concepts leave unserved need not always take the form of de novo engineering. In some cases, existing neighbouring concepts may be re-engineered to perform the required functions, avoiding the costs (e.g., in terms of implementation) associated with introducing novel concepts from scratch. This points to a more general feature of conceptual engineering: when multiple functions need to be served and multiple neighbouring concepts are available, there may be some latitude in how such functions are distributed across the available concepts.

not reveal constitutive principles so problematic that elimination was warranted.

Regardless of which conceptual intervention was chosen, implementation would have been crucial to functional success. Re-engineering AUTHORSHIP to include translators would require not merely conceptual revision but institutional changes: copyright offices would need procedures for registering translator claims, courts would need guidance for adjudicating disputes between translators and original authors, and publishing contracts would need frameworks for allocating rights and royalties. De novo engineering, introducing concepts like DERIVATIVE WORK, would similarly require implementation infrastructure: legal definitions would have to be codified in statute, professional associations would need to develop guidelines for applying new categories, and educational institutions would have to train lawyers, publishers, and translators in the new conceptual framework.

The choice among interventions involved trade-offs. Preservation maintained functional clarity for legal and economic purposes but accepted failure to perform recognition functions that many judged important. Re-engineering to include translators would enable inclusive recognition but might reduce the clarity of rights allocation and complicate legal coordination. De novo engineering could perform both exclusive and inclusive functions through multiple concepts but at the cost of conceptual complexity. The actual historical trajectory involved partial and uneven implementation of multiple strategies: some jurisdictions introduced neighbouring concepts granting translators limited rights (de novo engineering), academic and literary communities developed informal norms of crediting translators (social rather than conceptual intervention), and over time the concept of DERIVATIVE WORKS emerged in copyright law to provide a framework that acknowledged translation as creative while maintaining distinctions from original authorship (combined de novo engineering and partial re-engineering) (Lee, 2020).

Stage 6. Conceptual implementation

The interventions that actually occurred, i.e., the piecemeal introduction of translators' rights, the emergence of derivative work categories, the evolution of cultural norms around translation credit, partially restored functional adequacy but left residual tensions. Notably, many implementation challenges persisted even where concepts were

adequately engineered. For example, translators in some jurisdictions had legal recognition but lacked effective mechanisms for enforcing their rights, and cultural norms about crediting translators evolved slowly despite conceptual frameworks that should have supported recognition. These implementation gaps meant that well-designed conceptual interventions often failed to perform their intended functions in practice.

Stage 7. Evaluation and monitoring

Monitoring over the twentieth century revealed that questions of translation and authorship did not resolve definitively but evolved with changing technologies and practices. The emergence of machine translation in the late twentieth and early twenty-first centuries introduced new challenges that echoed earlier debates (Lacruz Mantecón, 2023): if human translators struggled for recognition as authors, what functional status should be accorded to AI(-assisted) translation? Does AI(-assisted) translation perform any of the functions that led to demands for recognising human translators? The persistence of these questions suggests that the conceptual interventions, while reducing the severity of the initial disruption and enabling concepts to perform additional functions, did not fully restore adequacy across all contexts and all stakeholders. Monitoring revealed that even partial interventions leave residual functional tensions requiring ongoing adjustment as the socio-technical landscape, its practices and values evolve.

6.3.2 Case study. AUTHORSHIP and AI(-assisted) translation

The emergence of large language models and AI(-assisted) translation systems capable of generating sophisticated text represents a new chapter in the longstanding question of translation and authorship, and creates renewed uncertainty about whether AUTHORSHIP continues to perform its functions adequately. While human translators have gradually gained recognition as creative contributors, though still subordinated to the original authors, AI translation systems like DeepL, Google Translate, and LLMs raise novel questions about whether and how AUTHORSHIP can perform its functions when translation involves minimal or no human creative labour. This section applies the seven-stage approach speculatively to the contemporary AI translation case, exploring how the cycle might guide conceptual engineering in response to an ongoing and unresolved disruption. The analysis necessarily remains tentative: unlike

the historical translation case where outcomes are known, the AI and translation question is actively contested, with no settled resolution and a rapidly evolving socio-technical landscape.

Stage 1. Suspicion of disruption

The suspicion that AI(-assisted) translation disrupts AUTHORSHIP manifests across multiple domains. In legal contexts, publishers and rights holders face uncertainty about who holds copyright over AI-translated texts. For example, when a novel is translated by an AI system with minimal human intervention, perhaps only prompt engineering and light editing, existing frameworks struggle to allocate rights clearly. Does copyright vest in the original author, the person who prompted the translation, the editor who refined outputs, the creator of the AI system, or does the translation enter the public domain? In commercial publishing, these questions have practical urgency. For example, international rights sales depend on clear ownership structures, and translation markets function through established conventions about who controls derivative works and how profits are distributed (Lacruz Mantecón, 2023).

Moreover, questions arise about recognition and value. Literary translation by humans has been understood, albeit incompletely and unevenly, as involving creative labour worthy of acknowledgment. Venuti (1995) documented how translators fought for recognition against conceptions of translation as mechanical reproduction. Now, when AI systems can produce fluent translations rapidly and cheaply, what becomes of these recognitions? If publishers can generate translations algorithmically, does this vindicate older views that translation involves no genuine authorship, or does it create a new problem where neither human translators nor AI systems (or the human using AI systems for the translation) are recognised as contributors? If AUTHORSHIP cannot perform its function of marking creative labour as culturally valuable when that labour is performed or assisted by AI, and if this leads to treating all AI(-assisted) translation as unworthy of recognition, the concept may fail to perform functions stakeholders have come to depend on.

The suspicion is more acute than in the translation case because AI challenges not just the boundaries of AUTHORSHIP but the constitutive principles connecting authorship to human capacities. Translators were indisputably human agents performing intentional creative labour; the

question was whether their specific form of labour qualified as authorship. Conversely, AI systems are commonly understood as lacking consciousness and intentionality and the phenomenology of creative experience that has historically been central to authorship concepts. Lang (2022) observes that concepts of authorship have centred on "agency/creativity, intention/authority, and self-presentation/self-construction", dimensions that AI systems arguably lack. Yet AI outputs are similar to human-authored texts, sometimes indistinguishable without information about provenance. This similarity combined with ontological difference generates suspicion that existing authorship concepts may not perform their functions adequately in the new socio-technical landscape.

The suspicion is heightened by AI's challenge to distinctions that previously seemed stable. When human translators advocated for authorial recognition, they could point to their intentional creative choices, their interpretive judgments, and their lived understanding of language and culture. AI systems are generally held to lack these qualities, yet may produce (or meaningfully assist in the production of) outputs that are comparable to human translations. This seems to put pressure on the concept of AUTHORSHIP, which is generally held to require human agency, consciousness, and creativity. Insofar as AI(-assisted) translations do not satisfy such requirements, they reasonably cannot qualify as authored works; yet they perform roles in literary and commercial markets that have historically required authorship. The tension signals potential disruption.

Stage 2. Context and stakeholder specification

AI translation operates across overlapping contexts with distinct stakeholders. In legal contexts, AUTHORSHIP functions to allocate intellectual property rights under copyright law. Jurisdictions often tie copyright frameworks to human authorship, creating ambiguity about AI-translated texts (He, 2019; Watiktinnakorn et al, 2023). Original authors of source texts have interests in controlling derivative works and benefiting from international editions. Human translators, now potentially competing with AI-aided human translators and AI systems, have interests in maintaining frameworks that recognise and protect their labour. AI developers seek clarity about rights to outputs generated by their systems, potentially claiming ownership based on their creation of the technology. Publishers navigate between these competing interests

while seeking legal certainty to structure contracts and manage rights efficiently.

In commercial publishing contexts, AUTHORSHIP structures market categories and compensation practices. Publishers traditionally pay human translators royalties or fees, reflecting recognition of their creative contribution. When translation is performed by AI, should compensation go to the AI developer, the original author, someone who prompts and edits the AI output, or no one? The functional requirement is not merely determining who gets paid but maintaining market structures that enable coordination between stakeholders. For example, if ownership is unclear, rights sales may become difficult to negotiate, creating friction in international publishing markets.

In socio-cultural contexts, AUTHORSHIP marks which forms of labour deserve recognition and cultural esteem. Human translators have partially secured recognition through decades (or centuries) of advocacy, though they remain less visible than original authors. The emergence of AI(-assisted) translation creates a dilemma: if AI-generated translations are treated as authored works, this might undermine arguments that translation requires distinctively human creativity; if they are treated as unprotected mechanical outputs, this might reinforce views that translation generally (including human translation) involves no genuine authorship. The stakeholders here include human translators concerned about their professional status, readers and critics who engage with translated works, and the broader socio-cultural landscape that depends on concepts such as AUTHORSHIP to structure (attributions of) recognition and value.

Lastly, these contexts are interconnected. For example, legal determinations about copyright affect economic incentives for both human translators and AI developers, which shapes what translation services are available and at what cost, which influences cultural practices around translated works. The stakeholders have partly competing and partly aligned interests. Original authors want to control and profit from translations of their works, but some may prefer human translators for quality or ethical reasons while others may favour AI for cost and speed. Human translators want recognition and fair compensation, which may require frameworks that distinguish their contributions from AI outputs, yet they may also use AI tools as assistants, complicating neat distinctions. Publishers want legal clarity and cost efficiency, which AI translation enables, but they also depend on

requirements of quality and cultural legitimacy that human expertise may better accommodate.

Stage 3. Functional and affordance mapping

Mapping the functional and affordance profile of AUTHORSHIP as it operates in AI(-enabled) translation contexts requires identifying which functions this concept currently performs across the relevant domains and what it enables and constrains for different stakeholders. For example, in legal contexts, AUTHORSHIP can be understood as performing the functions of allocating property rights in ways that enable orderly market transactions, provide sufficient incentive for valuable creative activity, and align with defensible normative principles about who deserves to benefit from creative outputs. In commercial contexts, the case could be made that AUTHORSHIP contributes to structuring compensation fairly, coordinating transactions efficiently, and maintaining trust between stakeholders, such as original authors, translators, editors, publishers, and readers. Lastly, in socio-cultural contexts, AUTHORSHIP performs the functions of recognising creative labour and structuring discourse about literary production.

Moreover, attending to the affordances of AUTHORSHIP reveals what the concept enables and constrains for different stakeholders. For example, the current concept of AUTHORSHIP affords original authors clear rights over derivative outputs, including control over translations of their works. The current ambiguity about the status of AI(-enabled) translation also affords flexibility to translators and publishers, enabling the usage of AI in translation practices if desired, while retaining the understanding of the human translator involved as the authorial (or authorial-adjacent) figure. For AI developers, the current concept affords neither recognition nor liability.

Stage 4. Adequacy and disruption assessment

Assessment of conceptual adequacy requires first determining which functions AUTHORSHIP should perform in AI(-enabled) translation contexts, and the extent to which these functions can be performed adequately when translation involves AI. For example, should AI-translated texts be protected by copyright to incentivise AI development and ensure quality translation services? Should protection focus on the original work, with translations treated as unprotected derivatives?

Should rights be allocated based on degree of human intervention in the translation process?

Moreover, when publishers commission translations, they need clarity about what rights they are acquiring and from whom. When original authors license translation rights, they need frameworks that specify what control they retain and how they benefit. Human translators need concepts that enable them to claim fair compensation for their labour. To what extent are such functions desirable? Is it desirable that AI developers capture value from translations, even when source texts were created by human authors whose works were used to train the systems? Is it desirable that original authors control all translation outputs regardless of the creative labour involved in producing them? Is it desirable that human translators maintain protections that distinguish them from AI on the grounds that human creative labour has distinctive value that merits recognition, even if this creates market inefficiencies? If so, under which conditions?

Again, from a socio-cultural perspective, is it desirable to recognise AI-generated translations as authored works, given that this might dilute the concept's capacity to mark distinctively human achievement? Is it desirable to deny recognition to AI(-assisted) translations, even when they perform important functions of enabling cross-cultural communication and making literature accessible? Is it desirable to maintain strong distinctions between human and AI(-assisted) translation, or should we move toward more graduated conceptions that recognise contributions along a spectrum?

Ultimately, AUTHORSHIP has to perform multiple, potentially conflicting functions, and different stakeholders will reasonably prioritise these functions differently. For example, some stakeholders will likely choose to prioritise the functions of recognising human creative labour and maintaining cultural frameworks that value human contribution. Others may prioritise functions of legal clarity, market efficiency, and maximising access to translated literature through cost-effective AI(-assisted) translation. Still others may seek concepts that can accommodate hybrid practices where human translators use AI tools, requiring frameworks that attribute contributions fairly across mixed workflows. It is also possible that no single concept of AUTHORSHIP may perform all these functions adequately.

It seems plausible that the assessment of the adequacy of the notion of AUTHORSHIP would reveal disruption across at least two dimensions. On the one hand, a conceptual gap may emerge from AI(-

assisted) translation—understood as non-human, non-conscious linguistic transformation that produces outputs functionally comparable to human-authored translations—representing a phenomenon that existing AUTHORSHIP concepts do not aim to capture. While some jurisdictions may have concepts akin to "computer-generated works", these are often underspecified for translation contexts, leaving uncertainty about rights allocation (Lacruz Mantecón, 2023). Moreover, the concept of "derivative work" captures that translations are based on source texts, but it presupposes human authorship of the derivative, creating ambiguity when translation is AI-generated. To the extent that no existing concept clearly performs all the functions required to govern AI translation comprehensively across legal, commercial, and socio-cultural domains, a conceptual gap may be present.

On the other hand, there may also be a conceptual conflict. Indeed, AI(-assisted) translations share functional features with human translations, yet they lack features central to traditional authorship concepts: intentionality, lived cultural understanding, conscious creative judgment. To the extent that the requirement of human creative agency is a constitutive principle of the concept of AUTHORSHIP, this would conflict with the empirical fact that AI systems produce (or aid in the production of) translations that perform roles previously requiring human authorship. This creates classificatory uncertainty: should AI translations be understood as authored works (stretching the concept to include non-human production) or as mechanical outputs (denying recognition to non-entirely-human production)? More fundamentally, if present, this disruption would reveal tensions in how AUTHORSHIP relates to translation more generally. When human translators argued for authorial recognition, they claimed that translation involves creativity comparable to original composition. If we now accept that AI can translate (or meaningfully aid in translation) without creativity, consciousness, or intentionality, does this undermine the basis for recognising human translators as authors? Or does it clarify that translation authorship was never really about these human capacities but about functional contributions to textual production?

Venuti (1995) has argued that translators' invisibility stemmed from "an individualistic concept of original authorship" that subordinated derivative work. AI translation creates a new version of this tension: if AI can be a translator-author, the concept is stretched beyond human agency; if AI cannot be a translator-author, we must either maintain that translation generally lacks authorship (contradicting gains

human translators achieved) or distinguish human from AI translation in ways that preserve recognition for human labour while denying it to functionally similar AI outputs. Again, this raises new issues when asked to disentangle the varied ways in which human translators may choose to avail themselves of AI systems to assist them in their work, as it is plausible to expect that AI tools used as grammar checks will not raise concerns of the same magnitude as AI tools which complete translations autonomously, with minimal input from the human translator.

These disruptions spans individual concept through conceptual clusters. At the individual concept level, AUTHORSHIP may face both gaps (lacking clear coverage for AI translation) and conflicts (if its constitutive principles clash with AI translation practices). At the cluster level, related concepts are implicated. If AI(-assisted) translation counts as authorship, it would be plausible to expect that neighbouring concepts such as CREATIVITY and ORIGINALITY would be similarly challenged. Conversely, if AI(-assisted) translation does not count as authorship, how do we distinguish human translation (which increasingly uses AI assistance) from AI translation (which may involve human prompting and editing)? The concepts of DERIVATIVE WORK, INTELLECTUAL PROPERTY, and TRANSLATION itself are all strained by uncertainty about how AUTHORSHIP applies to AI-generated outputs.

The severity of the disruption of AUTHORSHIP appears moderate to high. First, AI translation is rapidly becoming ubiquitous. Second, the stakes are substantial: legal frameworks depend on clear rights allocation to structure international publishing markets, human translators' livelihoods depend on concepts that recognise their contributions as distinct from automated processes, socio-cultural practices depends on concepts that mark which forms of labour merit recognition, and whether AI translation qualifies as authorship may have an impact on the volume of works that get translated and thus exposed to a broader audience that may not have had the resources to access the original text. Third, provisional solutions appear unstable: treating all AI translation as unprotected while recognising human translation as authored may create enforcement problems when outputs are indistinguishable and when human translators use different degrees of AI assistance. Conversely, treating both as authored collapses distinctions stakeholders value and may eliminate protections human translators struggled to achieve. Fourth, as Lauber-Rönsberg and Hetmank (2019) observe in related contexts, distinguishing human from AI outputs "is hardly possible without further information on the creative process", creating practical

barriers to implementing any conceptual distinction that depends on provenance.

Stage 5. Conceptual engineering interventions

Given the diagnosis of both gaps and conflicts at moderate-to-high severity, several intervention strategies are available. Preservation, retaining current AUTHORSHIP concepts that tie authorship to human agency, would maintain the concept's capacity to recognise human creative labour and mark it as culturally valuable. This preserves functions related to distinguishing human from automated production and maintaining incentives for human translators.

Re-engineering the concept, revising AUTHORSHIP to include AI translation as a form of authorship, would address the conceptual gap by providing clear legal coverage and rights allocation for AI-generated translations (Lacruz Mantecón, 2023). This could treat AI developers, users who prompt translation, or even AI systems themselves as authors, enabling the concept to perform legal and economic coordination functions. However, a similar re-engineering intervention would likely undermine the recognition of distinctively human creative labour. If AI systems can be authors of translations despite lacking human agency, consciousness, or intentionality, the concept's capacity to mark human achievement as special is diminished. Human translators who fought for recognition on grounds that translation involves creativity and skill would find their claims weakened: if AI can be an author-translator, then authorship apparently does not require the human capacities they emphasised. This intervention enables some functions (legal clarity, comprehensive coverage) while sacrificing others (distinctive recognition

of human labour, marking human creativity as singularly culturally valuable).¹⁰⁸

A *de novo* engineering response, introducing new concepts alongside traditional AUTHORSHIP, offers a middle path. This might amount to creating concepts like AI-GENERATED TRANSLATION or ALGORITHMIC DERIVATIVE WORK to provide legal coverage for AI outputs without extending AUTHORSHIP to non-human translators. In this case, AUTHORSHIP would continue to perform traditional functions for human-created translations, while new concepts would perform analogous

¹⁰⁸ A further consideration deserves emphasis here. Re-engineering AUTHORSHIP to encompass AI(-enabled) outputs would alter the concept across all the domains in which it operates, with consequences for all authors and authorial practices, not merely those involving translation. To the extent that AUTHORSHIP is a generic concept that structures creative practices across a variety of domains, including literature, visual arts, music, and academic research (van de Poel, 2025), a revision that extends AUTHORSHIP to cover AI(-enabled) outputs in the context of translation may similarly impact authorship in other contexts in which the concept is used. This raises a general methodological point about conceptual engineering of generic concepts: when a concept operates across a wide range of domains and practices, interventions motivated by issues in one domain of application may have unintended (negative) consequences across other domains. This should prompt the investigation of which practices would be affected by any proposed conceptual engineering strategy without assuming that any intervention that would solve the disruption in the domain currently targeted by the conceptual engineering project at hand will automatically justify proceeding with the implementation of such strategy. For example, it may be preferable to opt for a sub-optimal intervention strategy that can mitigate the disruption in the context under investigation while having fewer far-reaching effects on related uses of the concept in neighbouring domains. This point also has implications for the discussion about concepts, conceptions, and sub-concepts in section §6.2.1.1. One advantage of treating LEGAL AUTHORSHIP and SOCIO-CULTURAL AUTHORSHIP as distinct sub-concepts or conceptions, rather than as a single unified concept, is precisely that it limits the scope of intervention, such that e.g., re-engineering LEGAL AUTHORSHIP to include or exclude AI(-enabled) outputs need not entail re-engineering SOCIO-CULTURAL AUTHORSHIP in the same way. In this sense, when a concept is disrupted in one domain, domain-specific intervention through sub-concepts or conceptions may be preferable rather than retaining the unified concept across domains to avoid the collateral effects of conceptual engineering interventions to spread to uses of the original concept in domains where no disruption may be present. While this may not always be preferable, considering whether to intervene on a unified concept, a conception, or a sub-concept, may affect both how disruption is diagnosed and how contained or far-reaching any intervention will be.

functions (rights allocation, legal clarity) for AI translations (Sawerwain & Lim, forthcoming). This would preserve AUTHORSHIP's capacity to mark human labour as distinct while enabling comprehensive governance across different production modes. The original author might retain copyright over the source text, with AI translations treated as a new category of derivative work that grants limited rights to AI developers or users. Human translators' work would remain recognised as authored, maintaining the cultural and legal status they achieved historically, while AI translations would occupy a separate conceptual space.

The implementation challenge runs across all intervention strategies. For example, how do we distinguish human from AI translation in practice when human translators increasingly use AI tools? Consider a human translator who might use AI to generate initial drafts of specific excerpts, then revise them extensively: should this count as human or AI translation? The concept must be operationalisable: that is, criteria should be provided for what counts as "sufficient human contribution" to qualify as human authorship rather than AI generation. This might involve thresholds (e.g., more than 50% of sentences substantially revised by a human translator qualifies as human authorship), disclosure requirements (translators must declare AI use, with categorisation based on self-reporting), or institutional verification (publishers or professional associations may certify translations as human-authored based on process documentation). Each operationalisation involves trade-offs between precision and practicality.

Another version of *de novo* engineering would introduce graded concepts recognising degrees or types of contribution. Rather than binary human/AI authorship, concepts like PRIMARY AUTHOR-TRANSLATOR (for original human translation), COLLABORATIVE TRANSLATION (for substantial human-AI co-production), and AI-ASSISTED DERIVATIVE WORK (for minimally-edited AI outputs) could enable more fine-grained functional performance. This draws on Love's (2002) framework distinguishing precursory, executive, declarative, and revisionary authorship: a human translator who prompts AI, curates outputs, and edits results performs different authorial functions than AI systems executing linguistic transformation. Graded concepts would enable recognition of diverse contributions while maintaining distinctions stakeholders value. However, this would also multiply the conceptual complexity and may create coordination problems if stakeholders disagree about which category applies to particular cases.

A fourth option may primarily involve the implementation of a suitable infrastructure. This may involve AUTHORSHIP retaining its traditional human-centred meaning, and robust systems being developed to track translated works' provenance, enabling the verification of author contribution and the attribution of authorship rights in mixed workflows. This might involve technical solutions: AI systems could embed metadata in translations indicating what source text was used, what model generated the translation, what prompts were given, and what human editing occurred, with blockchain or similar technologies creating verifiable records of the translation process.¹⁰⁹ Moreover, professional certification systems could verify that human translators performed sufficient creative labour to warrant authorial recognition. However, technical and institutional infrastructure may be costly to develop, difficult to enforce across jurisdictions, and vulnerable to rapid technological change that may render specific solutions obsolete.

Stage 6. Conceptual implementation

Because the example being examined is ongoing, the evaluation of the extent to which the conceptual engineering intervention managed to meet the adequacy requirements must be approached prospectively. Such assessment would need to address the feasibility of the implementation of the conceptual engineering interventions. For example, legal adequacy requires that property rights are allocated clearly enough to enable orderly market transactions without creating excessive uncertainty or coordination costs. Can courts, copyright offices, and legal practitioners apply the concepts reliably? Can contracts specify rights transfers using available concepts? Are the concepts enforceable across jurisdictions with

¹⁰⁹ For example, one promising direction may involve watermarking or creating metadata standards that preserve provenance information throughout the publication chain. AI translation systems could embed information about source texts, models used, and processing steps directly in translated texts as metadata or through steganographic techniques. This would enable verification without requiring post-hoc detection, addressing Lauber-Rönsberg and Hetmank's (2019) concern that human and AI outputs are difficult to distinguish. However, such technical solutions would face challenges: watermarks may be removable, metadata may be stripped or falsified, and standards require coordination across a wide range of actors, including competing AI developers and publishing platforms. Ultimately, implementation feasibility depends on both technical capabilities and institutional cooperation.

different legal traditions? If the intervention depends on distinguishing human from AI translation, implementation must provide verification mechanisms that function reliably in practice; otherwise, even well-designed concepts may fail to perform their functions adequately. In the commercial domain, feasible implementation involves not just legal infrastructure but market conventions, professional standards, and business practices. If concepts are too complex or ambiguous, markets may develop informal workarounds that undermine intended functional performance. For instance, publishers might systematically misclassify AI translations as human-authored to avoid legal uncertainty, defeating the purpose of conceptual distinctions. Can publishers negotiate rights clearly with original authors, human translators, and AI developers? Does the framework enable market segmentation (e.g., premium human translations vs. budget AI translations) if stakeholders demand it? In the broader socio-cultural context, the adequacy of (re-)engineered concepts of authorship may instead be tracked in terms of how well such concepts align with the landscape in which they are used, as well as with evolving norms and values. Ultimately, if concepts fail to gain uptake in cultural practice, for example, if critics and readers ignore new conceptual distinctions or apply them inconsistently, functional adequacy may be undermined regardless of conceptual clarity.

Stage 7. Evaluation and monitoring

Lastly, ongoing monitoring is essential because AI translation capabilities continue evolving rapidly. A conceptual engineering intervention that restores conceptual adequacy for the current socio-technical landscape may face new disruptions as technology advances and the socio-technical landscape evolves. For example, if AI translation quality improves to the point where it consistently surpasses human translation on metrics such as readers' preference, cultural attitudes about the importance of human authorship might shift, potentially requiring a new conceptual engineering intervention. Moreover, if AI systems develop capabilities for understanding or creativity that are currently deemed uniquely human, authorship concepts linking authorship to human agency might face renewed pressure. Lastly, and perhaps more likely, if human translators increasingly work in hybrid modes, e.g., using AI for drafts, lexical suggestions, or quality checking, the boundary between human and AI translation may blur further, requiring more nuanced concepts or different implementation approaches.

The recursive structure of this approach is particularly important for this case. As the suspected disruption of AUTHORSHIP for AI translation is heavily technology-induced and the socio-technical landscape is in a state of constant flux, a one-off conceptual engineering response is unlikely to fully resolve all inadequacies and restore conceptual adequacy in a way that can sustain the test of time; rather, gradual solutions may be necessary. In this sense, this approach makes this iterative process manageable, allowing conceptual engineers to re-assess conceptual adequacy and re-evaluate the extent to which existing concepts may require further conceptual engineering interventions as the socio-technical landscape continues evolving.

* * * * *

Ultimately, the cycle provides a structured process that makes explicit what information is needed to determine standards for conceptual adequacy, and assess concepts against such standards (stages 1-4), which intervention options are available to conceptual engineers (stage 5), and how the success of conceptual engineering interventions might be assessed (stages 6-7).

6.4 Conclusion

This closing chapter synthesised the contributions from the conceptual engineering in the philosophy of technology literature and from previous chapters of this dissertation to answer the central research question that motivated this investigation: how should we (re)engineer concepts that have been disrupted in relation to socially disruptive technologies? In the chapter, I illustrated how to navigate conceptual disruption with conceptual engineering. I did so in two moves.

First, I synthesised the theoretical insights that underpin such an activity. That is, concepts are adequate to the extent that they fulfil their functions. When they fail to do so, a conceptual disruption occurs. Philosophers of technology are increasingly investigating how to respond to technology-induced conceptual disruptions with conceptual engineering. In doing so, attention is being placed not only on restoring conceptual adequacy in terms of ensuring that concepts fulfil the functions we wish them to serve, but also on attending to the affordances of concepts, i.e., the other uses that concepts enable or restrict given their

users and the contexts of use. In this sense, the literature provides us with a diagnostic category (conceptual disruption), which indicates when concepts' functioning has been interrupted, alongside a taxonomy of types of disruption, levels at which the disruption occurs, and urgency for intervention. It also provides us with an approach (conceptual engineering) to evaluate when conceptual disruption is present, and (if appropriate) respond to such a disruption through the (re-)design and implementation of concepts. It also provides us with an evaluative lens (functionalism, supplemented by affordances) through which to conduct adequacy judgements, and allowing for the assessment of how concepts support or hinder specific outcomes.

Second, I offered an operationalisation of such insights into a cyclical, iterative cycle, composed of seven stages: suspicion of disruption, context and stakeholder specification, functional and affordance mapping, adequacy and disruption assessment, conceptual engineering intervention, conceptual implementation, and evaluation and monitoring. Lastly, I availed myself of the examples of historical and current (suspected) disruptions of the concept of AUTHORSHIP with regard to human and AI-assisted translation to illustrate how this approach may be operationalised in practice. While this is a limited set from which to draw general conclusions about the effectiveness of this approach for navigating conceptual disruption with conceptual engineering, the case studies suggest that the cycle can be successfully applied to both historical and speculative suspicions of disruption.

Quando finisce un'opera? [...] E cosa succede quando l'opera viene data al pubblico, e quindi esposta non solo all'interpretazione ma anche alla relazione fisica con il tempo e lo spazio? Quante volte gli artisti avranno voglia di rapire le loro tele e aggiustare quella virgola che a distanza di anni proprio non va più bene?

When is an artwork finished? [...] And what happens when the work is given to the public, and thus exposed not only to interpretation but also to the physical relationship with time and space? How many times will artists want to abduct their canvases and fix that comma that just doesn't fit anymore?

Daniela Collu, Un minuto d'arte

Chapter 7

CONCLUSION

*7.1 Summary and key findings — 7.2 Limitations
— 7.3 Prospects for future research*

7.1 Summary and key findings

In this dissertation, I set out to investigate how to navigate technology-induced conceptual disruption (i.e., the interruption in the functional adequacy of concepts) with conceptual engineering (i.e., the evaluation, design, and implementation of concepts). This investigation has shown that this question cannot be adequately addressed by either discipline in isolation, but requires bringing the philosophy of technology and conceptual engineering into sustained dialogue. In doing so, it has introduced methodological and substantive innovations to both disciplines.

On the one hand, it has argued that conceptual engineering offers philosophers of technology the vocabulary and theoretical grounding to more explicitly and intentionally engage with conceptual work, a practice that philosophers of technology have long been conducting implicitly. The dissertation has built on this move from the philosophy of technology to conceptual engineering to address and conceptualise the phenomenon of technology-induced conceptual disruption, which philosophers of technology are increasingly interested in exploring. In doing so, it has equipped the philosophy of technology with a richer and sharper vocabulary for navigating conceptual disruption, and it has positioned conceptual engineering as a central approach for systematically identifying, assessing, and addressing conceptual disruption.

Specifically, chapter 2 (*What is conceptual disruption?*) provided the diagnostic groundwork of the dissertation, by bringing clarity to the phenomenon of conceptual disruption. Against existing understandings of conceptual disruption emphasising classificatory uncertainty or challenges to meaning, the chapter suggested that conceptual disruption should instead be understood as an interruption in the functional adequacy of concepts, i.e., in terms of the failure of concepts to adequately fulfil their functional roles. Moreover, the chapter proposed a taxonomy of the phenomenon of conceptual disruption along three dimensions: (i) *levels* at which the disruption may occur, i.e., concepts, conceptual cluster, conceptual scheme; (ii) *types* of disruption, i.e., conceptual gaps and conceptual conflicts; and (iii) *severity* of the disruption, extending from mild to severe. The chapter also emphasised the aptness of conceptual engineering in responding to conceptual disruptions, and suggested that different kinds of conceptual engineering responses best correspond to different types of disruption, such as re-

engineering for conceptual conflicts and de novo engineering for conceptual gaps.

The phenomenon of conceptual disruption and its relation to conceptual engineering was further developed in chapter 5 (*Conceptual resilience*), which examined the phenomenon of resilience vis-à-vis conceptual disruption and conceptual engineering. The focus on resilience shifted the attention from purely remedial conceptual engineering interventions to cure instances of conceptual disruption to preventive approaches to avoid disruptions from manifesting in the first place. The chapter introduced the CONCEPTUAL RESILIENCE as the capacity of concepts to maintain functional adequacy under stress, arguing that this capacity is not binary but graded. Drawing on interdisciplinary literatures, the chapter developed two complementary heuristics: Conceptual Resilience as Immutability (CRI), emphasising through stability, and Conceptual Resilience as Adaptation (CRA), emphasising resilience through flexible adjustment to changing contexts. The chapter then addressed the question of desirability, arguing that resilience is sometimes advantageous, particularly for thin concepts at the individual level, but not always desirable. By incorporating resilience into conceptual engineering, the chapter clarified how preventative conceptual engineering strategies can complement remedial ones, by equipping engineers to (re-)engineer concepts in ways that minimise future vulnerability to disruption. Ultimately, the chapter showed that, when navigating technology-induced conceptual disruption, conceptual engineering should not always be reactive. Rather, whenever deemed desirable, concepts may be (re-)designed for resilience to better endure disturbances.

The dissertation has also argued that the philosophy of technology can, in turn, advance conceptual engineering, by exploring the analogy between concepts and artefacts. Specifically, it has taken seriously the analogy between concepts and artefacts—often mentioned in the conceptual engineering literature, but rarely developed in dialogue with the rich philosophy of technology scholarship on socio-technical artefacts—and examined what that analogy yields when rigorously pursued. In particular, it has argued that the philosophy of technology possesses a rich set of tools and approaches developed for the evaluation, design, development, and deployment of socio-technical artefacts. Such insights and tools can be fruitfully applied to conceptual engineering projects given the salient similarities between concepts and socio-

technical artefacts, as well as between the similar normative aims of conceptual engineering and the philosophy of technology.

In particular, chapter 3 (*Concepts as artefacts*) expanded on the intuition, widespread in conceptual engineering, that concepts are saliently similar to artefacts, by establishing a link between the conceptual engineering literature on concepts and the philosophy of technology literature on socio-technical artefacts. The chapter argued that exploring the similarities between concepts and (the study of) socio-technical artefacts can provide three sources of untapped potential for further conceptual engineering exploration.

First, it was argued that conceptual engineering should not limit itself to the study of the functions of concepts (i.e., what concepts do or should do), but would benefit greatly from the investigation of the affordances of concepts (i.e., what concepts enable and constrain).

Second, conceptual engineering should consider (re-)engineered concepts as value-laden. For example, concepts reflect (some of) the values of conceptual engineers and of the broader socio-techno-cultural landscape in which the conceptual engineering process takes place. Third and last, acknowledging that the investigation of artefacts in the philosophy of technology is often operationalised through structured methodologies, it was argued that conceptual engineering would benefit from exploring similar systematic approaches to the evaluation, design, and implementation of concepts.

This provided the methodological underpinning for the analysis of conceptual disruption and for responsible conceptual engineering. This enabled structuring the subsequent investigation in a manner that resembled how ethicists of technology approach the responsible evaluation, design, development, and deployment of socio-technical artefacts. For example, concepts are deemed adequate or inadequate not in a vacuum, but within the context in which they are embedded and with respect to the users that they affect. This was addressed in chapter 4 (*Conceptual affordances*), which introduced CONCEPTUAL AFFORDANCES, understood as the potential actions, uses, or thoughts that a concept enables or constrains within a given context and relative to specific stakeholders. The chapter argued that functional approaches to conceptual engineering can and should be enriched by grounding functionalism in a theory of affordances, thus broadening the assessment of conceptual adequacy to include the range of actions, uses, and thoughts that a concept *enables* or *constrains* across contexts and with respect to specific stakeholders, beyond what the aims that it *should* serve.

Ultimately, the chapter proposed that affordance-informed functionalism provides both theoretical and ethical benefits, e.g., as it allows conceptual engineers to account for and treat differences in affordances as a normatively significant criterion when multiple adequate concepts compete.

Moreover, building on the analogy between concepts and artefacts, as well as on the salient similarities between conceptual engineering and the philosophy of technology, the dissertation encouraged the adoption of variations of existing approaches developed in the ethics of technology for conceptual engineering purposes. Chapter 6 (*Navigating conceptual disruption with conceptual engineering*) offered an example of how conceptual engineers may tap into philosophy of technology approaches to navigate conceptual disruption. The chapter synthesised and operationalised the contributions of the previous chapters by addressing the practical task of how to translate theoretical insight into an operationalisable approach. First, the chapter synthesised the dimensions that should be of interest to conceptual engineering when navigating conceptual disruption. Second, it tied them together into a step-by-step cyclical and iterative approach aiming to guide philosophers of technology and conceptual engineering in their conceptual engineering projects, from the moment of the suspicion of disruption through to conceptual engineering intervention and monitoring. In short, the chapter operationalised how to determine when, and if so, how, we should (re-)engineer our concepts in response to technology-induced conceptual disruption. This approach may be used by philosophers of technology and conceptual engineers alike to navigate conceptual disruption with conceptual engineering.

Ultimately, this investigation contributed to the advancement of both conceptual engineering and the philosophy of technology. Moreover, the dissertation illustrated that bringing the philosophy of technology and conceptual engineering into sustained dialogue not only provides the necessary insights to adequately navigate technology-induced conceptual disruption with conceptual engineering, but also generates significant mutual benefits for both disciplines.

7.2 Limitations

Several limitations circumscribe the scope of this investigation, while also pointing toward productive avenues for further inquiry. First, this dissertation has focused primarily on technology-induced conceptual

disruption. However, conceptual disruption can also arise from non-technological sources, such as scientific discoveries, social movements, legal developments, or cultural evolution. As illustrated in the case studies of chapter 6, the affordances-informed functional approach and cyclical methodology developed here are in principle applicable to these contexts. However, this dissertation has not systematically explored whether diagnostic criteria or intervention strategies require adjustment for non-technological disruptions. This opens questions about whether technology-induced disruption exhibits distinctive features, such as rapid onset, ongoing instability, or material embedding in technological systems, that warrant specialised frameworks, or whether the tools developed here generalise straightforwardly across different drivers of conceptual change.

Second, the dissertation adopts a deliberately ecumenical conceptualisation of concepts as a functional kind, remaining neutral about their precise metaphysical nature to maintain compatibility with diverse theoretical commitments. While Chapter 4 illustrates how the affordances framework operates across competing metaphysical views, this broad approach leaves certain questions about conceptual change mechanisms underspecified. How exactly do engineered concepts achieve uptake? What cognitive or social processes enable or resist revision? At what point does modification create a new concept rather than revising an existing one? These questions also relate to the discussion surrounding conceptual resilience. The methodological character of this dissertation justified bracketing questions surrounding which concept of concept one ought to adopt in conceptual engineering, as well as how one ought to determine which functions concepts should fulfil and what should count as adequate functional fulfilment. However, a more substantive engagement with such questions more would arguably contribute to refining and enhancing both the understanding of the phenomena of conceptual adequacy and conceptual disruption, as well as conceptual engineering approaches to navigate conceptual disruption.

One more limitation concerns the predominant theoretical and methodological contribution of this dissertation. The operationalised framework in Chapter 6 provided structured guidance for navigating conceptual disruption, but necessarily involved significant context-dependent judgments at each stage. It specified what information ought to be gathered and which considerations should inform decisions, but was not intended to mechanically generate answers about when

disruption is present or which interventions are appropriate. In this sense, the framework's utility depends greatly on the skill and ethical sensitivity of practitioners applying it, regarding which the dissertation provided limited guidance. For example, the investigation did not address how to structure collaborative deliberation when applying the framework with diverse stakeholders or how to navigate power dynamics when interests conflict.

7.3 Prospects for future research

This dissertation has sought to clarify when, and if so how, we should (re-)engineer our concepts in response to technology-induced conceptual disruption. While the chapters collectively offer an account of conceptual disruption, conceptual resilience, and affordance-informed functionalism, as well as an operationalised approach to conceptual engineering in practice, the work presented here is far from exhaustive. On the contrary, it is best understood as opening a set of pathways for further exploration. The conceptual tools developed in this dissertation provide fertile ground for further elaboration and critical reassessment. Moreover, the broader field of philosophy of technology, particularly in relation to socially disruptive technologies, offers a constantly evolving landscape in which new applications, refinements, and challenges will inevitably arise. In what follows, I highlight a number of promising directions for future research that emerge from the present inquiry.

One pressing avenue for future research concerns the ethical dimensions of conceptual engineering itself. While conceptual engineering is a normative project, explicit reflection on the ethics of intervening in our conceptual repertoire remains underdeveloped (Crisp, 2025; Löhr, 2024; Veluwenkamp, 2025). Yet, as this dissertation has shown, conceptual interventions are rarely neutral: they are normatively charged, value-laden, and capable of shaping how technologies are understood and governed. Future work should therefore develop systematic accounts of responsible conceptual engineering, e.g., focusing on the values that guide decisions about when and how to (re-)engineer concepts. This may involve addressing questions about paternalism, the legitimacy of nudging through conceptual change, or the ethical stakes of engineering concepts in contexts where values are contested. The development of guidelines for responsible conceptual engineering in the philosophy of technology would contribute to ensuring that conceptual work does not merely react to technological change, but actively

contributes to shaping more responsible socio-technical futures. This also raises questions of how to integrate conceptual engineering into responsible innovation frameworks (Veluwenkamp & van den Hoven, 2023; Löhr & Veluwenkamp, 2025).

Another promising research direction lies in comparative approaches to conceptual engineering. While this dissertation has emphasised the value of viewing concepts as artefacts, further progress could be made by drawing parallels with conceptual engineering practices in neighbouring domains. For instance, moral and political philosophy, the philosophy of science, and applied ethics all engage, implicitly or explicitly, in conceptual engineering projects. Comparative studies can highlight methodological commonalities and divergences, shedding light on the strengths and limitations of different approaches to conceptual evaluation, (re-)design, and implementation. More concretely, comparative conceptual engineering could provide new resources for dealing with the challenges posed by socially disruptive technologies, particularly when moral and social concepts are at stake. By taking inspiration from evaluative frameworks, heuristics, and methodological tools developed in different domains, conceptual engineering can enhance its methodological robustness. Such work would also encourage interdisciplinary dialogue, broadening the reach of conceptual engineering beyond philosophy of technology and embedding it within a larger ecosystem of conceptual inquiry.

Moreover, closely tied to this inquiry is the need to explore the interplay between conceptual disruption and social disruption (Anderson et al, 2026). As socially disruptive technologies alter entrenched norms, institutions, and practices, they often generate corresponding pressures on our conceptual landscape (Hopster, 2021; 2024; Löhr, 2023a; 2023b; Marchiori & Scharp, 2024; Veluwenkamp et al, 2024). Investigating the relationship between these two forms of disruptiveness promises to enrich our understanding of both. On the one hand, conceptual disruptiveness may serve as an indicator of social disruptiveness, insofar as failures of concepts to adequately fulfil their functions often signal broader tensions within the socio-technical landscape. On the other hand, attention to social disruptiveness may illuminate which conceptual domains are most likely to require engineering in the future. Future research could thus develop a framework to systematically assess the interconnections between conceptual and social disruptiveness, ultimately strengthening the contribution of conceptual engineering to the ethics and governance of technology. Such work would help establish

conceptual engineering as a key tool not only for philosophical analysis but also for anticipatory governance and policy-making.

Finally, future work should explore the implications of conceptual engineering for the governance of technology more broadly. The cyclical and iterative approach proposed in this dissertation provides a starting point, but much remains to be done in translating theoretical insights into practically operationalisable guidelines. One promising avenue is to integrate conceptual engineering into anticipatory governance practices, where emerging technologies are assessed not only in terms of risks and benefits but also in terms of their potential to destabilise conceptual, normative, and institutional structures (Guston, 2010; Michelson, 2016; Sarewitz, 2011). In this regard, conceptual engineering could complement existing regulatory approaches, such as the European Union's risk-based model for AI, by offering tools to identify and address the conceptual challenges raised by socially disruptive technologies (Marchiori & al, 2025). Future research should further elaborate how conceptual engineering can function as a conceptual pillar of technology governance, e.g., ensuring that governance regimes and policy-making are not undermined by inadequate or outdated concepts. This would represent a significant step toward embedding conceptual engineering within broader interdisciplinary and societal efforts to guide technological development responsibly.

REFERENCES

- Achterhuis, H. (Ed.). (2001). *American Philosophy of Technology: The Empirical Turn*. Indiana University Press.
- Adams, D. (1979). *The Hitchhiker's Guide to the Galaxy*. London: Gollancz.
- Adams, D. (1995). *The Restaurant at the End of the Universe*. Ballantine Books.
- Alexander, D. E. (2013). Resilience and disaster risk reduction: an etymological journey. *Natural Hazards and Earth System Sciences*, 13(11), 2707-2716.
- Al-Razgan, M. S., Al-Khalifa, H. S., Al-Shahrani, M. D., & AlAjmi, H. H. (2012). Touch-based mobile phone interface guidelines and design recommendations for elderly people: A survey of the literature. In *International Conference on Neural Information Processing* (pp. 568-574). Berlin, Heidelberg: Springer.
- Alaca, I. V. (2017). Materiality in picturebooks. In B. Kümmerling-Meibauer (Ed.), *The Routledge Companion to Picturebooks* (pp. 59-68). Routledge.
- Allen, C. R., & Holling, C. S. (Eds.). (2008). *Discontinuities in ecosystems and other complex systems*. Columbia University Press.
- Ames, R.T. (2021) *Human Becomings: Theorizing Persons for Confucian Role Ethics*. State University of New York Press.
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973-989.
- Anderson, R. T. (2013). Marriage: What it is, why it matters, and the consequences of redefining it. *Heritage Foundation Backgrounder*, 2775, 1-12.
- Anderson, J., Hopster, J., & Lundgren, B. (2026). Defining Socially Disruptive Technologies and Reframing the Ethical Challenges They Pose. *Technology in Society*, 103216.
- Archer, B. (1979). Design as a discipline. *Design Studies* 1(1), 17-20.
- Arnold, S., Kuroishi, M., Karashima, R., Adachi, T., & Yamazaki, K. (2023). Recognising Affordances in Predicted Futures to Plan

- With Consideration of Non-Canonical Affordance Effects. *IEEE Robotics and Automation Letters*, 8(3), 1455-1462.
- Ashmore, R. (2004). Word and Gesture: On Xuan-School Hermeneutics of the Analects. *Philosophy East and West*, 54 (4), 458-488. <https://www.jstor.org/stable/4148015>
- Babushkina, D. (2020). Robots to blame. In M. Nørskov, J. Seibt, and O.S. Quick (Eds.), *Culturally Sustainable Social Robotics: Proceedings of Robophilosophy*, 305-315. <https://doi.org/10.3233/FAIA200927>
- Baier, K., & Rescher, N. (1969). Values and the Future: The Impact of Technological Change on American Values.
- Bajohr, H. (2024). Writing at a distance: Notes on authorship and artificial intelligence. *German Studies Review*, 47(2), 315-337.
- Baker, R. (2013). *Before Bioethics*. Oxford University Press.
- Bærentsen, K. B., & Trettvik, J. (2002). An activity theory approach to affordance. *Proceedings of the Second Nordic Conference on Human-Computer Interaction*, 51-60.
- Belleri, D. (2021). Downplaying the change of subject objection to conceptual engineering. *Inquiry*. <https://doi.org/10.1080/0020174X.2021.1908161>
- Belleri, D. (2023). Conceptual Engineering Between Representational Skepticism and Complacency: Is There a Third Way? *Topoi*, 42(4), 1051-1062. <https://doi.org/10.1007/s11245-023-09924-3>
- Bennett, K. (2020). Authorship and (self-) translation in academic writing: Towards a genetic approach. *Genetic Translation Studies: Conflict and Collaboration in Liminal Spaces*, 179-196.
- Bennett, J., & Bennett, D. (2012). The Impact of Unintended Affordances on User Experience. *Journal of Usability Studies*, 7(2), 45-56.
- Bennett, M. & Hacker, P. (2008). *History of Cognitive Neuroscience*. Oxford: Wiley-Blackwell.
- Berensmeyer, I., Buelens, G., & Demoor, M. (2012). Authorship as Cultural Performance. *New Perspectives in Authorship Studies*. ZAA 60(1), 5-29.

- Berkes, F., & Folke, C. (Eds.). (1998). *Linking social and ecological systems: Management practices and social mechanisms for building resilience*. Cambridge University Press.
- Bernat, J. L. (1998). A defense of the whole-brain concept of death. *Hastings Center Report*, 28(2), 14-23.
- Bernat, J. L. (2002). The biophilosophical basis of whole-brain death. *Social Philosophy and Policy*, 19(2), 324-342.
- Bernat, J. L. (2017). Brain death and the definition of death. *Neuroethics: Anticipating the future*, 336-65.
- Bennett, C. J. (2011). In defense of privacy: The concept and the regime. *Surveillance & Society*, 8(4), 485-496.
- Bimo, S. (2024). A brief history of authorship. *The Routledge Handbook of AI and Literature*, 49.
- Blackburn, S. (1992). Through thick and thin. *Proceedings of the Aristotelian Society*, 66, 285-299.
- Boeglin, J. (2015). The costs of self-driving cars: reconciling freedom and privacy with tort liability in autonomous vehicle regulation. *Yale Journal of Law & Technology*, 17, 171.
- Boenink, M., Swierstra, T., & Stermerding, D. (2010). Anticipating the interaction between technology and morality: A scenario study of experimenting with humans in bionanotechnology. *Studies in Ethics, Law, and Technology*, 4(2), 1-38. <https://doi.org/10.2202/1941-6008.1098>
- Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist*, 59(1), 20-28. <https://doi.org/10.1037/0003-066X.59.1.20>
- Borghini, A. M. (2021). Affordances, context and sociality. *Synthese*, 199(5), 12485-12515. <https://doi.org/10.1007/s11229-018-02044-1>
- Boroditsky, L. (2006). Linguistic relativity. *Encyclopedia of Cognitive Science*.
- Bosnell, R., & Madder, H. (2011). Concepts of brain death. *Surgery (Oxford)*, 29(7), 289-294.

- Brandom, R. (2009). *Articulating reasons: An introduction to inferentialism*. Harvard University Press.
- Brey, P. (2010). Philosophy of technology after the empirical turn. *Techne: Research in Philosophy & Technology*, 14(1).
- Brey, P. A. (2012). Anticipatory ethics for emerging technologies. *NanoEthics*, 6(1), 1–13.
- Brey, P. (2016). Constructive philosophy of technology and responsible innovation. In M. Franssen, P. E. Vermaas, E. Kroes, & A. W. Meijers (Eds.), *Philosophy of Technology After the Empirical Turn* (pp. 127-143). Cham: Springer International Publishing.
- Brey, P. (2018). The strategic role of technology in a good society. *Technology in Society*, 52, 39-45.
- Briguglio, L., Cordina, G., Farrugia, N., & Vella, S. (2009). Economic vulnerability and resilience: Concepts and measurements. *Oxford Development Studies*, 37(3), 229-247.
- Brożek, B., & Janik, B. (2019). Can artificial intelligences be moral agents? *New Ideas in Psychology*, 54, 101-106. <https://doi.org/10.1016/j.newideapsych.2018.12.002>
- Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., ... & Von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra*, 19(4), 733-752.
- Burgess, A., & Plunkett, D. (2013). Conceptual ethics I. *Philosophy Compass*, 8(12), 1091-1101.
- Burgess, A., Cappelen, H., & Plunkett, D. (2020). *Conceptual Engineering and Conceptual Ethics*. Oxford University Press.
- Burgess, A., & Plunkett, D. (2020). On the relation between conceptual engineering and conceptual ethics. *Ratio*, 33(4), 281-294.
- Burke, B. (1992). Searching for exoplanets. *Targets for Space-Based Interferometry*, European Space Agency, 81-83.
- Caldeira, S., & Timmins, F. (2016). Resilience: Synthesis of concept analyses and contribution to nursing classifications. *International Nursing Review*, 63(2), 191-199.

- Cappelen, H. (2018). *Fixing Language: An Essay on Conceptual Engineering*. Oxford University Press.
- Cappelen, H., & Plunkett, D. (2020). A guided tour of conceptual engineering and conceptual ethics. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 1–26). Oxford University Press. <https://doi.org/10.1093/oso/9780198801856.003.0001>
- Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.
- Cartlidge, E. (2012). Loose cable may unravel faster-than-light result. *Science*, 335(6072), 1027. <https://doi.org/10.1126/science.335.6072.1027>
- Casati, R., & Varzi, A. (1994). *Holes and other superficialities*. Cambridge, MA/London: The MIT Press.
- Čerka, P., Grigienė, J., & Širbikytė, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, 31(3), 376-389.
- Chalmers, D. J. (2020). What is conceptual engineering and what should it be? *Inquiry*. <https://doi.org/10.1080/0020174X.2020.1817141>
- Chalmers, D. J. (2011). Verbal disputes. *Philosophical Review*, 120(4), 515-566.
- Chatzipanteli, A., & Gorozidis, G. S. (2024). Developing young children's physical literacy through picturebooks. *Early Childhood Education Journal*, 52(7), 1681-1692.
- Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, 15(2), 181-195.
- Chemero, A., Klein, C., & Cordeiro, W. (2003). Events as changes in the layout of affordances. *Ecological Psychology*, 15(1), 19-28.
- Chen, Y. (2025). How Does Gene Editing Technology Reshape the Theoretical Cognitive Framework of Disease Mechanisms. *Computers and Education letters*, 2(02), 48-64.
- Coca-Vila, I. (2018). Self-driving cars in dilemmatic situations: An approach based on the theory of justification in criminal law. *Criminal Law and Philosophy*, 12(1), 59-82.

- Coeckelbergh, M. (2016). Responsibility and the moral phenomenology of using self-driving cars. *Applied Artificial Intelligence*, 30(8), 748-757.
- Coeckelbergh, M., & Gunkel, D. J. (2025). *Communicative AI: A critical introduction to large Language models*. John Wiley & Sons.
- Cohen, B., & Murphy, G. L. (1984). Models of concepts. *Cognitive science*, 8(1), 27-58.
- Coll, S. (2014). Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. *Information, Communication & Society*, 17(10), 1250-1263.
- Comporti, M. (2009). *Fatti illeciti: Le responsabilità oggettive*. Giuffrè Editore.
- Condit, N. (2022). Regulating heritable human genome editing: Drawing the line between legitimate and controversial use. *European Journal of Health Law*, 29(3-5), 435-457.
- Confucius. (2003). *Analects: With Sections from Traditional Commentaries*. (Slingerland, E., Trans.) Hackett.
- Conrey, J. B. (2003). The Riemann Hypothesis. *Notices of the AMS*, 50(3), 341-353.
- Cosentino, E. (2021). Artifacts and affordances. *Synthese*, 198(Suppl17), 4007-4026. <https://doi.org/10.1007/s11229-019-02297-4>
- Cretney, R. (2014). Resilience for whom? Emerging critical geographies of socio-ecological resilience. *Geography Compass*, 8(9), 627-640.
- Crisp, R. (2025). Towards an ethics of conceptual engineering. *Inquiry*, 68(2), 755-768.
- Crootof, R., & Ard, B. J. (2020). Structuring techlaw. *Harvard Journal of Law & Technology*, 34, 347.
- D'Ambrosio, P. (2019). Wang Bi's Commentary on the *Analects*: A Confucian-Daoist Critique of Effable Morality. *Philosophy East and West*, 69 (2), 357-375. <https://doi.org/10.1353/pew.2019.0030>
- D'Ambrosio, P. (2016). Guo Xiang on Self-so Knowledge. *Asian Philosophy*, 26 (2), 119-132. <https://doi.org/10.1080/09552367.2016.1163774>

- Danaher, J., & Hopster, J. (2022). The normative significance of future moral revolutions. *Futures*, 144, 103046.
- Danaher, J., & Sætra, H. S. (2022). Technology and moral change: the transformation of truth and trust. *Ethics and Information Technology*, 24(3), 35.
- Danaher, J., & Sætra, H. S. (2023). Mechanisms of techno-moral change: A taxonomy and overview. *Ethical Theory and Moral Practice*, 26(5), 763-784.
- da Fonseca, A. T., Vaz de Sequeira, E., & Barreto Xavier, L. (2023). Liability for AI Driven Systems. In *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (pp. 299-317). Cham: Springer International Publishing.
- Dancy, J. (1993). *Moral Reasons*. Blackwell
- De Georgia, M. A. (2014). History of brain death as death: 1968 to the present. *Journal of Critical Care*, 29(4), 673-678.
- de Terte, I., Stephens, C., & Huddleston, L. (2014). The development of a three-part model of psychological resilience. *Stress and Health*, 30(5), 416-424. <https://doi.org/10.1002/smi.2625>
- DeCew, J. (2018). Privacy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition). Available at <https://plato.stanford.edu/archives/spr2018/entries/privacy> (last accessed March 31, 2023).
- Dembroff, R. (2024). Intersection is not identity, or how to identify overlapping systems of injustice. In R. Chang and A. Srinivasan (Eds.), *Conversations in Philosophy, Law, and Politics* (pp. 383-400). Oxford University Press.
- Deola, S., Guerrouahen, B. S., Sidahmed, H., Al-Mohannadi, A., Elnaggar, M., Elsadig, R., ... & Cugno, C. (2018). Tailoring cells for clinical needs: Meeting report from the Advanced Therapy in Healthcare symposium (October 28-29 2017, Doha, Qatar). *Journal of translational medicine*, 16(1), 276.
- Deutsch, M. (2020). Trivializing conceptual engineering. *Inquiry*, 1-16. <https://doi.org/10.1080/0020174X.2020.1853343>

- Díaz-León, E. (2025). *The Metaphysics of Gender*. Cambridge University Press.
- Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., Génova, G., et al. (2018). Ethics by design: Necessity or curse?. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 60-66).
- Dummett, M. (1993). *Seas of Language*. Oxford: Oxford University Press.
- Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press
- Eklund, M. (2002). Inconsistent languages. *Philosophy and Phenomenological Research*, 64(2), 251-275. <https://doi.org/10.1111/j.1933-1592.2002.tb00001.x>
- Eriksen, C. (2020). *Moral Change*. Springer International Publishing.
- Feng, P., & Feenberg, A. (2008). Thinking about design: Critical theory of technology and the design process. In *Philosophy and Design: From Engineering to Architecture* (pp. 105-118). Dordrecht: Springer Netherlands.
- Fineman, M. A. (2006). The meaning of marriage. *Marriage proposals: Questioning a legal status*, 29, 43-57.
- Fineman, M. A. (2001). Why marriage? *Virginia Journal of Social Policy & the Law*, 9, 239.
- Floridi, L. (2025). Distant Writing: Literary Production in the Age of Artificial Intelligence. *Minds and Machines*, 35(3), 1-26.
- Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press
- Floridi, L., & Strait, A. (2021). Ethical foresight analysis: What it is and why it is needed? *The 2020 Yearbook of the Digital Ethics Lab*, 173–194.
- Folke, C., Carpenter, S. R., Walker, B., Scheffer, M., Chapin, T., & Rockström, J. (2010). Resilience thinking: Integrating resilience, adaptability and transformability. *Ecology and Society*, 15(4)
- Franssen, M., & Koller, S. (2016). Philosophy of technology as a serious branch of philosophy: The empirical turn as a starting point. In

- Philosophy of Technology after the Empirical Turn* (pp. 31-61). Cham: Springer.
- Franssen, M., Vermaas, P. E., Kroes, P., & Meijers, A. W. (Eds.). (2016). *Philosophy of Technology after the Empirical Turn* (Vol. 23). Cham: Springer International Publishing.
- Fraser, C. (2020). Metaphysics and Agency in Guo Xiang's Commentary on the *Zhuangzi*. In Chai, D. (Ed.) *Dao Companion to Xuanxue* 玄學 (Neo-Daoism). (pp. 343-366). Springer.
- Friedewald, M., & Pohoryles, R. J. (Eds.). (2016). *Privacy and Security in the Digital Age: Privacy in the Age of Super-Technologies*. Routledge.
- Friedman, B. (1996). Value-sensitive design. *Interactions*, 3(6), 16–23.
- Friedman, B., & Kahn Jr, P. H. (2007). Human values, ethics, and design. In *The Human-Computer Interaction Handbook* (pp. 1267–1292). CRC Press.
- Fuchs, C. (2011). Towards an alternative concept of privacy. *Journal of Information, Communication and Ethics in Society*, 9(4), 220-237.
- Gallie, W. B. (1955). Essentially contested concepts. *Proceedings of the Aristotelian Society*, 56, 167–198.
- Garzelli, B. (2011). Translating cultures: mediation, authorship and the role of the translator. In " *The Translator as Author. Perspectives on Literary Translation*" (pp. 175-180).
- Gaver, W. W. (1991). Technology affordances. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 79-84.
- Gert, J. (2016). Thick moral concepts and intuitionism. *Ethics*, 126(3), 721-747. <https://doi.org/10.1086/684711>
- Gervais, D. J. (2025). Artificial Intellectual Property. *Chicago-Kent Law Review*, 100, 265.
- Gibbard, A. (2003). Reasons thin and thick. *The Journal of Philosophy*, 100(6), 288-304.
- Gibbard, A., & Blackburn, S. (1992). Morality and thick concepts. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 66, 267-299.

- Gibson, E. J. (1982). The concept of affordances in development: The renaissance of functionalism. In W. A. Collins (Ed.), *The Concept of Development: The Minnesota Symposia on Child Psychology* (Vol. 15) (pp. 55-81). New Jersey: Lawrence Erlbaum Associates.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Gibson, J. J. (1977a). The theory of affordances. In R. Shaw (Ed.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Erlbaum.
- Gibson, J. (1977b). Perceiving, Acting, and Knowing: Toward an Ecological Psychology. In Shaw, R. & Bransford, J. (Eds.). *The Theory of Affordances*. Hillsdale, NJ: Lawrence Erlbaum.
- Gillon, R. (1990). Death. *Journal of Medical Ethics*, 16(1), 3.
- Glăveanu, V.P. (2020). Affordance. In: *The Palgrave Encyclopedia of the Possible*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-98390-5_10-2
- Gless, S., Silverman, E., & Weigend, T. (2016). If robots cause harm, who is to blame? Self-driving cars and criminal liability. *New Criminal Law Review*, 19(3), 412-436.
- Goetze, T. S. (2022). Mind the gap: autonomous systems, the responsibility gap, and moral entanglement. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 390-400).
- Gonçalves, V. P., de Almeida Neris, V. P., Seraphini, S., Dias, T. C., Pessin, G., Johnson, T., & Ueyama, J. (2017). Providing adaptive smartphone interfaces targeted at elderly people: An approach that takes into account diversity among the elderly. *Universal Access in the Information Society*, 16(1), 129-149.
- Gosseries, A. (2008). How (un)fair is intellectual property?. In *Intellectual Property and Theories of Justice* (pp. 3-26). London: Palgrave Macmillan UK.
- Grant, J., & Fox, F. (1992). Understanding the Role of the Designer in Society. *Journal of Art & Design Education*, 11(1), 77-87.

- Grove, K. (2018). Resilience as Essentially Contested Concept. In K. Grove, *Resilience* (1st ed., pp. 30–64). Routledge. <https://doi.org/10.4324/9781315661407-2>
- Grunwald, A. (2000). Against over-estimating the role of ethics in technology development. *Science and Engineering Ethics*, 6, 181–196.
- Guerit, J. M. (2004). The concept of brain death. *Brain death and disorders of consciousness*, 15-21.
- Gunderson, L. H. (2000). Ecological resilience in theory and application. *Annual Review of Ecology and Systematics*, 31(1), 425-439.
- Guston, D. H. (2010). The anticipatory governance of emerging technologies. *Applied Science and Convergence Technology*, 19(6), 432-441.
- Hage, J. (2017). Theoretical foundations for the responsibility of autonomous agents. *Artificial Intelligence and Law*, 25(3), 255-271.
- Hamel, G., & Välikangas, L. (2003). The quest for resilience. *Harvard Business Review*, 81(9), 52-63
- Hansson, S. O. (2006). Defining technical function. *Studies In History and Philosophy of Science Part A*, 37(1), 19-22.
- Hardy, G. H., & Wright, E. M. (1960). *An Introduction to the Theory of Numbers* (4th ed.). Oxford University Press.
- Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs*, 34(1), 31–55.
- Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
- Haslanger, S. (2018). Social Explanation: Structures, Stories, and Ontology. A Reply to Díaz León, Saul, and Sterken. *Disputatio: International Journal of Philosophy*, 10(50).
- Haslanger, S. (2020a). Going on, not in the same way. In A. Burgess, H. Cappelen, D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 230–260).
- Haslanger, S. (2020b). How not to change the subject. In *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability*. Oxford

University Press. <https://doi.org/10.1093/oso/9780198803331.003.0013>

- He, T. (2019). The sentimental fools and the fictitious authors: rethinking the copyright issues of AI-generated contents in China. *Asia Pacific Law Review*, 27(2), 218-238.
- Heras-Escribano, M. (2019). *The philosophy of affordances*. Springer.
- Hermann, J. (2022). Socially disruptive technologies and moral certainty. In *Philosophical Perspectives on Moral Certainty* (pp. 19-34). Routledge.
- Hermann, J. (2019). The dynamics of moral progress. *Ratio*, 32(4), 300-311.
- Hettinger, E. C. (1989). Justifying intellectual property. *Philosophy & Public Affairs*, 31-52.
- Hickman, L. A. (2001). *Philosophical tools for technological culture. Putting pragmatism to work*. Indiana University Press.
- Hindriks, F., & Veluwenkamp, H. (2023). The risks of autonomous machines: from responsibility gaps to control gaps. *Synthese*, 201(1), 21.
- Holland, A. (2012). Ecological Balance. In D. Callahan & P. Singer (Eds.), *Encyclopedia of Applied Ethics: Vol. 2* (pp. 10-13).
- Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4(1), 1-23.)
- Holling, C. S. (1996). Engineering resilience versus ecological resilience. In P. C. Schulze (Ed.), *Engineering within Ecological Constraints* (pp. 31-44). National Academy Press.
- Holmes, K. (2018). *Mismatch: How Inclusion Shapes Design*. MIT Press.
- Hopster, J. (2021). What are socially disruptive technologies? *Technology in Society*, 67, 101750. <https://doi.org/10.1016/j.techsoc.2021.101750>
- Hopster, J. (2024a). Is ARTIFICIËLE INTELLIGENTIE een defectief concept?. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 116(4), 337-351.
- Hopster, J. K. G. (2024b). Socially disruptive technologies and epistemic injustice. *Ethics and Information Technology*, 26(1), 14.

- Hopster, J. (2026). Human Disenhancement, Socially Disruptive Technologies, and Legitimate Expectations. *AJOB neuroscience*, 17(1), 65-67.
- Hopster, J., Brey, P., Klenk, M. B. O. T., Löhr, G., Marchiori, S., Lundgren, B., & Scharp, K. (2023). Conceptual disruption and the ethics of technology. In I. van de Poel, L. Frank, J. Hermann, J. Hopster, D. Lenzi, S. Nyholm, B. Taebi, & E. Ziliotti (Eds.), *Ethics of Socially Disruptive Technologies: An Introduction* (pp. 141–162). Open Book Publishers. <https://doi.org/10.11647/obp.0366.06>
- Hopster, J. K. G., Gerola, A., Hofbauer, B., Löhr, G., Rijssenbeek, J., & Korenhof, P. (2023). Who owns NATURE? Conceptual appropriation in discourses on climate and biotechnologies. *Environmental Values*. <https://doi.org/10.1177/09632719231196535>
- Hopster, J., & Löhr, G. (2023). Conceptual Engineering and Philosophy of Technology: Amelioration or Adaptation? *Philosophy & Technology*, 36(4), 70. <https://doi.org/10.1007/s13347-023-00670-3>
- Hopster, J. K., & Maas, M. M. (2024). The technology triad: disruptive AI, regulatory gaps and value change. *AI and Ethics*, 4(4), 1051-1069.
- Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A review of definitions and measures of system resilience. *Reliability Engineering & System Safety*, 145, 47-61
- Houkes, W., & Vermaas, P. E. (2010). *Technical functions: On the use and design of artefacts* (Vol. 1). Berlin: Springer.
- Huang, B., Chen, C., & Shu, K. (2025). Authorship attribution in the era of llms: Problems, methodologies, and challenges. *ACM SIGKDD Explorations Newsletter*, 26(2), 21-43.
- Iancu, I., & Iancu, B. (2020). Designing mobile technology for elderly. A theoretical overview. *Technological Forecasting and Social Change*, 155, 119977.
- International Astronomical Union (IAU) General Assembly. (2006). Definition of a planet in the Solar System. *IAU Resolution B5*. Busan, South Korea. Available at https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf (last accessed March 31, 2023)

- Ihde, D. (1995). *Postphenomenology: Essays in the Postmodern Context*. Northwestern University Press.
- Insanguine Mingarro, F. A. (2023). Gene Therapy and Germline Cells Research. In *Handbook of Bioethical Decisions. Volume I: Decisions at the Bench* (pp. 79-99). Cham: Springer International Publishing.
- Isaac, M. G. (2020). How to Conceptually Engineer Conceptual Engineering? *Inquiry: An Interdisciplinary Journal of Philosophy*, 1–24. <https://doi.org/10.1080/0020174x.2020.1719881>
- Isaac, M. G. (2023). Which Concept of Concept for Conceptual Engineering? *Erkenntnis: An International Journal of Scientific Philosophy*, 88(5), 2145–2169. <https://doi.org/10.1007/s10670-021-00447-0>
- Ivy, B. P. U., Mandiwa, P., & Kumar, M. (2012). A modified RSA cryptosystem based on 'n' prime numbers. *International Journal Of Engineering And Computer Science*, 1(2), 63-66.
- Jacobs, M. (1999). Sustainable development as a contested concept. In A. Dobson (Ed.), *Fairness and Futurity* (pp. 21–45). Oxford University Press.
- Jameson, G. J. O. (2003). *The prime number theorem*. Cambridge University Press.
- Jansen, H. (2019). I'm a translator and I'm proud: How literary translators view authors and authorship. *Perspectives*, 27(5), 675-688.
- Jenkins, K. (2016). Amelioration and inclusion: Gender identity and the concept of woman. *Ethics*, 126(2), 394–421.
- Johnson, D. G. (2007). Ethics and technology 'in the making': An essay on the challenge of nanoethics. *NanoEthics*, 1, 21–30.
- Jones, M. L. (2018). Does technology drive law? The dilemma of technological exceptionalism in cyberlaw. *University of Illinois Journal of Law, Technology & Policy*, 249.
- Jorem, S. (2022). The good, the bad and the insignificant. Assessing Concept functions for conceptual Engineering. *Synthese*, 200(2). <https://doi.org/10.1007/s11229-022-03548-7>

- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, 86(1), 65–79. Kirchin, S. (2013). *Thick Evaluation*. Oxford University Press.
- Klenk, M. (2021). How Do Technological Artefacts Embody Moral Values?. *Philosophy & Technology*, 34, 525-544. <https://doi.org/10.1007/s13347-020-00401-y>
- Klenk, M., O'Neill, E., Arora, C., Blunden, C., Eriksen, C., Frank, L., & Hopster, J. (2022). Recent work on moral revolutions. *Analysis*, 82(2), 354-366.
- Koch, S. (2021a). Engineering what? On concepts in conceptual engineering. *Synthese*, 199(1), 1955-1975.
- Koch, S. (2021b). Why Conceptual Engineers Should Not Worry About Topics. *Erkenntnis*, 88(5), 1–21. <https://doi.org/10.1007/s10670-021-00446-1>
- Köhler, S., & Veluwenkamp, H. (2024). Conceptual Engineering: For What Matters. *Mind*, 133(530), 400–427. <https://doi.org/10.1093/mind/fzad064>
- Kölbel, M. (2023). Varieties of conceptual analysis. *Analytic Philosophy*, 64(1), 20–38. <https://doi.org/10.1111/phib.12249>
- Korein, J., & Machado, C. (2004). Brain death: updating a valid concept for 2004. *Brain death and disorders of consciousness*, 1-14.
- Kocurek, A. W. (2022). What topic continuity problem? *Inquiry (United Kingdom)*. <https://doi.org/10.1080/0020174X.2022.2126884>
- Koutamanis, A. (2023). Technologies, Inbetweenness and Affordances. *Global Philosophy*, 33, 5. <https://doi.org/10.1007/s10516-023-09668-0>
- Kroes, P., & Meijers, A. (2006). The dual nature of technical artefacts. *Studies in History and Philosophy of Science*, 37(1), 1-4.
- Kümmerling-Meibauer, B., & Meibauer, J. (2013). Towards a cognitive theory of picturebooks. *International Research in Children's Literature*, 6(2), 143-160.

- Kümmerling-Meibauer, B., & Meibauer, J. (2017a). Early-concept books and concept books. In B. Kümmerling-Meibauer (Ed.), *The Routledge Companion to Picturebooks* (pp. 149-157). Routledge.
- Kümmerling-Meibauer, B., & Meibauer, J. (2017b). Picturebooks and cognitive studies. In B. Kümmerling-Meibauer (Ed.), *The Routledge Companion to Picturebooks* (pp. 391-400). Routledge.
- Kurniawan, S. (2007). Mobile phone design for older persons. *Interactions*, 14(4), 24-25.
- Lacruz Mantecón, M.L. (2023). Authorship and Rights Ownership in the Machine Translation Era. In H. Moniz, and C. Parra Escartín, C. (Eds.), *Towards Responsible Machine Translation. Machine Translation: Technologies and Applications, Vol 4*. Springer, Cham. https://doi.org/10.1007/978-3-031-14689-3_5
- Lakoff, G. (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 63-100). Cambridge University Press.
- Lalumera, E. (2009). *Cosa Sono i Concetti*. Editore Laterza.
- Lalumera, E. (2010). Concepts are a functional kind. *Behavioral and Brain Sciences*, 33(2-3), 217.
- Lalumera, E. (2014). On the explanatory value of the concept-conception distinction. *Rivista italiana di filosofia del linguaggio*, 8(3), 73-81.
- Lamb, D. (2020). *Death, brain death and ethics*. Routledge.
- Lambert, M. D. (2017). Picturebooks and page layout. In B. Kümmerling-Meibauer (Ed.), *The Routledge Companion to Picturebooks* (pp. 28-37). Routledge.
- Lane, B. (2025). Socially Disruptive Technologies, Moral Progress, and Rule Following. *Philosophy & Technology*, 38(2), 64.
- Lang, M. (2022). Conceptions of literary authorship in modern literary theories: history, issues, approaches. *Frontiers of Narrative Studies*, 8(1), 78-100.
- Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Harvard University Press.

- Latour, B. (2007). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.
- Lauber-Rönsberg, A., & Hetmank, S. (2019). The concept of authorship and inventorship under pressure: Does artificial intelligence shift paradigms?. *Journal of Intellectual Property Law & Practice*, 0(0), 1-10.
- Lee, T. K. (2020). Translation and copyright: towards a distributed view of originality and authorship. *The Translator*, 26(3), 241–256. <https://doi.org/10.1080/13556509.2020.1836770>
- Leichenko, R. (2011). Climate change and urban resilience. *Current Opinion in Environmental Sustainability*, 3(3), 164-168.
- Lengnick-Hall, C. A., Beck, T. E., & Lengnick-Hall, M. L. (2011). Developing a capacity for organizational resilience through strategic human resource management. *Human resource management review*, 21(3), 243-255.
- Lin, C. J., Hsieh, T. L., & Shiang, W. J. (2009). Exploring the interface design of mobile phone for the elderly. In *International Conference on Human Centered Design* (pp. 476-481). Berlin, Heidelberg: Springer.
- Liu, P., Yang, R., & Xu, Z. (2019). How safe is safe enough for self-driving vehicles?. *Risk analysis*, 39(2), 315-325.
- Lohmann, M. F. (2016). Liability issues concerning self-driving vehicles. *European Journal of Risk Regulation*, 7(2), 335-340.
- Löhr, G. (2022). Linguistic Interventions and the Ethics of Conceptual Disruption. *Ethical Theory and Moral Practice*, 25(5), 835–849. <https://doi.org/10.1007/s10677-022-10321-9>
- Löhr, G. (2023a). Conceptual disruption and 21st century technologies: A framework. *Technology in Society*, 74, 102327. <https://doi.org/10.1016/j.techsoc.2023.102327>
- Löhr, G. (2023b). If conceptual engineering is a new method in the ethics of AI, what method is it exactly? *AI Ethics*. <https://doi.org/10.1007/s43681-023-00295-4>

- Löhr, G. (2024). What's the relation between conceptual ethics and ethics? An instrumentalist defense of conceptual engineering. In *Conceptual engineering* (pp. 161-180). Brill.
- Löhr, G., & Michel, C. (2023). Conceptual engineering, predictive processing, and a new implementation problem. *Mind & Language*, 1-19. <https://doi.org/10.1111/mila.12471>
- Löhr, G., & Veluwenkamp, H. (2025). Rethinking philosophical methodology: Conceptual engineering meets Value Sensitive Design. *Metaphilosophy*, 56(3-4), 328-343.
- Lozano-Robledo, Á. (2019). *Number theory and geometry: An introduction to arithmetic geometry* (Vol. 35). American Mathematical Society.
- Lucivero, F. (2016). *Ethical Assessments of Emerging Technologies*. Cham: Springer.
- Lundgren, B. (2024). Undisruptable or stable concepts: Can we design concepts that can avoid conceptual disruption, normative critique, and counterexamples? *Ethics and Information Technology*, 26(2), 33. <https://doi.org/10.1007/s10676-024-09767-5>
- Lundgren, B., & Möller, N. (2019). Defining Information Security. *Science & Engineering Ethics*, 25, 419–441. <https://doi.org/10.1007/s11948-017-9992-1>
- Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development*, 71(3), 543-562.
- Machado, C. (2000). Is the concept of brain death secure?. *Major Problems in Neurology*, 36, 193-212.
- Machery, E. (2009). *Doing without Concepts*. Oxford University Press.
- Mackenzie, C., Stoljar, N. (Eds.) (2000). *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford University Press.
- Maier, J. R., & Fadel, G. M. (2007). Identifying affordances. In *DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007* (pp. 841-842).

- Malkov, S. M. (2024). "Translation difficulties": The problem of authorship of translated texts. *Philosophical Letters. Russian and European Dialogue*, 7(1), 209-223.
- Manyena, S. B. (2006). The concept of resilience revisited. *Disasters*, 30(4), 434-450.
- Marçal, K. (2021). *Mother of invention: How good ideas get ignored in an economy built for men*. Abrams.
- Marchiori, S. (2025). Conceptual affordances: (How) should they inform conceptual engineering?. *Synthese*, 206(2), 1-23. <https://doi.org/tudelft.idm.oclc.org/10.1007/s11229-025-05161-w>
- Marchiori, S., Hopster, J. K., Puzio, A., Riemsdijk, M. B. V., Kraaijeveld, S. R., Lundgren, B., Viehoff, J., & Frank, L. E. (2025). A social disruptiveness-based approach to AI governance: complementing the risk-based approach of the AI Act. *Science and engineering ethics*, 31(5), 25.
- Marchiori, S., & Scharp, K. (2024). What is conceptual disruption?. *Ethics and Information Technology*, 26(1), 18.
- Marchiori, S., & Sta. Maria, J. (2026). Conceptualising conceptual resilience. A comparative approach. *Philosophy & Technology* 39(1), 45. <https://doi.org/10.1007/s13347-026-01045-0>
- Margolis, E., & Laurence, S. (2023). *Concepts*. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition). Retrieved from <https://plato.stanford.edu/archives/fall2023/entries/concepts>
- Margolis, E., & Laurence, S. (1999). *Concepts: Core Readings*. Cambridge, MA: MIT Press.
- Margolis, E., & Laurence, S. (2007). The ontology of concepts. Abstract objects or mental representations? *Noûs*, 41(4), 561-593.
- Margolis, E., & Laurence, S. (2023). *Concepts*. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition). Available at <https://plato.stanford.edu/archives/fall2023/entries/concepts>
- Marques, T. (2025). Representing or shaping reality? What class can teach about woman. In *New Perspectives on Conceptual Engineering-*

Volume 2: Across Philosophy (pp. 95-113). Cham: Springer Nature Switzerland.

- Martens, M. (2012). Issues of access and usability in designing digital resources for children. *Library & Information Science Research*, 34(3), 159-168.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175-183.
- Maynard, J. (2019). The twin prime conjecture. *Japanese Journal of Mathematics*, 14(2), 175-206.
- Mazzi, F. (2024). Authorship in artificial intelligence-generated works: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection. *The Journal of World Intellectual Property*, 27(3), 410-427.
- McDonald, L. (2023). AI systems and liability: An assessment of the applicability of strict liability & A case for limited legal personhood for AI. *St Andrews Law Journal*, 3(1), 5-21.
- McGrenere, J., & Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Graphics interface* (pp. 179-186).
- McMahan, J. (2006). An alternative to brain death. *Journal of Law, Medicine & Ethics*, 34(1), 44-48.
- Meerow, S., Newell, J. P., & Stults, M. (2016). Defining urban resilience: A review. *Landscape and Urban Planning*, 147, 38-49.
- Menegazzi, D., & Nunes, C. C. (2025). From pages to places: cartographic literacy through children's book design. *Blucher Design Proceedings*, 13, 697-713.
- Michelson, E. (2016). Assessing the societal implications of emerging technologies: Anticipatory governance in practice. Routledge.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press.
- Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, 56(2), 288-302.

- Millikan, R. (2000). *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
- Miracchi, L. (2021). A case for integrative epistemology. *Synthese*, 198(12), 12021–12039. <https://doi.org/10.1007/s11229-020-02848-0>
- Miranda Vilchis, R. (2022). Improving conceptual engineering by differentiating the functions of concepts. *Logos and Episteme*, 13(4), 359–379. <https://doi.org/10.5840/logos-episteme202213431>
- Mitcham, C. (1994). *Thinking through technology: The path between engineering and philosophy*. University of Chicago Press.
- Moore, G. E. (1959 [1903]). *Principia ethica*. Cambridge University Press.
- Mufwene, S. S. (2013). Language as technology. Some questions that evolutionary linguistics should address. In T. Lohndal & J. Benjamins (Eds.), *In Search of Universal Grammar: From Norse to Zoque* (Vol. 202). John Benjamins Publishing.
- Mumford, E. (1987). Sociotechnical systems design: Evolving theory and practice. *Computers and Democracy*, 59–76.
- Nado, J. (2021). Conceptual engineering, truth, and efficacy. *Synthese*, 198, 1507–1527. <https://doi.org/10.1007/s11229-019-02096-x>
- Nair-Collins, M. (2010). Death, brain death, and the limits of science: Why the whole-brain concept of death is a flawed public policy. *Journal of Law, Medicine & Ethics*, 38(3), 667–683.
- Nathanson, M. B. (1996). *Additive Number Theory* (Vol. 164). New York: Springer.
- Nickel, P. J. (2020). Disruptive innovation and moral uncertainty. *NanoEthics*, 14(3), 259–269. <https://doi.org/10.1007/s11569-020-00375-3>
- Nickel, P. J. (2025). A Pluralistic Model of Technology-Driven Value Change. In *Der Sog des Neuen und der Schock des Alten* (pp. 39–58). Nomos Verlagsgesellschaft.
- Nickel, P. J., Kudina, O., & van de Poel, I. (2022). Moral uncertainty in technomoral change: Bridging the explanatory gap. *Perspectives on Science*, 30(2), 260–283.

- Nimmer, D. (2001). Authorship and Originality. *Houston Law Review*, 38, 1.
- Nissenbaum, H. (2018). Respecting context to protect privacy: Why meaning matters. *Science and Engineering Ethics*, 24(3), 831-852.
- Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32-48.
- Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
- Nordberg, A. (2021). Defining human enhancement: Towards a foundational conceptual tool for enhancement law. *Journal of Law, Information & Science*, 26, 23.
- Norman, D. A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.
- Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3), 38-43.
- Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.
- Nyholm, S. (2018a). The ethics of crashes with self-driving cars: A roadmap, I. *Philosophy Compass*, 13(7), e12507.
- Nyholm, S. (2018b). The ethics of crashes with self-driving cars: A roadmap, II. *Philosophy Compass*, 13(7), e12506.
- Nyholm, S. (2023). Minding the gap(s): Different kinds of responsibility gaps related to autonomous vehicles and how to fill them. In *Connected and automated vehicles: integrating engineering and ethics* (pp. 1-18). Cham: Springer Nature Switzerland.
- Nyholm, S. (2026). Social Robots: Culturally Sustainable or Socially Disruptive? The Case of Humanoid Robots. In *Social Robots and Cultural Sustainability* (pp. 179-203). Cham: Springer Nature Switzerland.
- O'Neill, M. E. (2009). The genuine Sieve of Eratosthenes. *Journal of Functional Programming*, 19(1), 95-106.
- Pagano, A., Perini, C., Cunha, E., & Pagano, A. (2024). Authorship attribution in translated texts: a stylometric approach to

- translator style. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, Vol. 2 (pp. 109-117).
- Palm, E., & Hansson, S. O. (2006). The case for ethical technology assessment (eTA). *Technological Forecasting and Social Change*, 73(5), 543-558.
- Patel, S. S., Rogers, M. B., Amlôt, R., & Rubin, G. J. (2017). What Do We Mean by 'Community Resilience'? A Systematic Literature Review of How It Is Defined in the Literature. *PLoS currents*, 9, ecurrents.dis.db775aff25efc5ac4f0660ad9c9f7db2.
- Pattison, M., & Stedmon, A. W. (2006). Inclusive design and human factors: Designing mobile phones for older users. *PsychNology Journal*, 4(3), 267-284.
- Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.
- Petrovčič, A., Taipale, S., Rogelj, A., & Dolničar, V. (2018). Design of mobile phones for older adults: An empirical analysis of design guidelines and checklists for feature phones and smartphones. *International Journal of Human-Computer Interaction*, 34(3), 251-264.
- Pimm, S. L. (1984). The complexity and stability of ecosystems. *Nature*, 307(5949), 321-326.
- Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, 14(3), 399-441.
- Plato. (2012) Euthyphro. In Reeve, C.D., (Ed.) *A Plato Reader*. (pp. 1-20). Hackett.
- Pleasant, N. (2018). The structure of moral revolutions. *Social Theory and Practice*, 567-592.
- Plunkett, D. (2015). Which concepts should we use? Metalinguistic negotiations and the methodology of philosophy. *Inquiry*, 58(7-8), 828-874.
- Plunkett, D., & Burgess, A. (2013). Conceptual ethics II. *Philosophy Compass*, 8(12), 1102-1110.

- Plunkett, D., & Sundell, T. (2023). Varieties of metalinguistic negotiation. *Topoi*, 42(4), 983-999.
- Plunkett, D., & Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*, 13(23), 1-37.
- Potter, N. 1980. *What is a Designer*. Reading, UK: Hyphen Press.
- Prinzing, M. (2018). The Revisionist's Rubric: Conceptual Engineering and the Discontinuity Objection. *Inquiry: An Interdisciplinary Journal of Philosophy*, 61(8), 854–880. <https://doi.org/10.1080/0020174x.2017.1385522>
- Putnam, H. (1965). A philosopher looks at quantum mechanics. In R. G. Colodny (Ed.), *Beyond the Edge of Certainty* (pp. 75-101). Englewood Cliffs, NJ: Prentice-Hall.
- Queloz, M. (2019). The points of concepts: Their types, tensions, and connections. *Canadian Journal of Philosophy*, 49(8), 1122–1145.
- Queloz, M. (2022). Function-based conceptual engineering and the authority problem. *Mind*, 131(524), 1247-1278.
- Ralph, P., & Wand, Y. (2009). A proposal for a formal definition of the design concept. *Design requirements engineering: A ten-year perspective*, 14, 103-136.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9z6v>
- Raz, J. (1994). Moral change and social relativism. *Social Philosophy and Policy*, 11(1), 139-158.
- Reed, D. J., & Monk, A. (2006). Design for inclusion. In *Designing Accessible Technology* (pp. 53-63). London: Springer London.
- Reese, E., & Riordan, J. (2017). Picturebooks and developmental psychology. In B. Kümmerling-Meibauer (Ed.), *The Routledge Companion to Picturebooks* (pp. 381-390). Routledge.
- Rey, G. (2010). Concepts versus conceptions (again). *Behavioral and Brain Sciences*, 33(2-3), 221-222.

- Riggs, J. (2021). Deflating the functional turn in conceptual engineering. *Synthese*, 199(3–4), 11555–11586. . <https://doi.org/10.1007/s11229-021-03302-5>
- Rip, A. (2000). There's no turn like the empirical turn in the philosophy of technology. In *The Empirical Turn in the Philosophy of Technology* (pp. 3-19). Emerald.
- Rogacz, D. (2021). Operating with Names: Operational Definitions in the *Analects* and Beyond. *Dao*, 21, 19-35. <https://doi.org/10.1007/s11712-021-09813-9>
- Rosenberger, R., & Verbeek, P. P. (2015). A field guide to postphenomenology. *Postphenomenological Investigations: Essays on Human-Technology Relations*, 9-41.
- Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In *Cognitive development and acquisition of language* (pp. 111-144). Academic Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, 104(3), 192.
- Rosch, E. (1978 [2024]). Principles of categorization. In *Cognition and categorization* (pp. 27-48). Routledge.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4), 573-605.
- Rovira, E., McLaughlin, A. C., Pak, R., & High, L. (2019). Looking for age differences in self-driving vehicles: examining the effects of automation reliability, driving risk, and physical impairment on trust. *Frontiers in Psychology*, 10, 800.
- Rowe, F., Jeanneret Medina, M., Journé, B., Coëtard, E., & Myers, M. (2024). Understanding responsibility under uncertainty: A critical and scoping review of autonomous driving systems. *Journal of Information Technology*, 39(3), 587-615.
- Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, 57(3), 316-331.
- Saidak, F. (2006). On Goldbach's conjecture for integer polynomials. *The American Mathematical Monthly*, 113(6), 541-545.

- Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & technology*, 34(4), 1057-1084.
- Sapir, E. (1929). The status of linguistics as a science. *Language*, 207–214.
- Sarbey, B. (2016). Definitions of death: brain death and what matters in a person. *Journal of Law and the Biosciences*, 3(3), 743-752.
- Sarewitz, D. (2011). Anticipatory governance of emerging technologies. In *The growing gap between emerging technologies and legal-ethical oversight: The pacing problem* (pp. 95-105). Dordrecht: Springer Netherlands.
- Sauer, H. (2023). *Moral Teleology: A Theory of Progress*. Taylor & Francis.
- Sauer, H., Blunden, C., Eriksen, C., & Rehren, P. (2021). Moral progress: Recent developments. *Philosophy Compass*, 16(10), e12769.
- Saul, J. (2006). II—Jennifer Saul: Gender and Race. In *Aristotelian Society Supplementary Volume* (Vol. 80, No. 1, pp. 119-143). Oxford, UK: Oxford University Press.
- Sawerwain, S., & Lim, D. (Forthcoming). The case for AI authorship in copyright law. *Law, Innovation and Technology*. <https://doi.org/10.2139/ssrn.5108423>
- Sebastián, M. Á. (2021). First-person representations and responsible agency in AI. *Synthese*, 199 (3-4), 7061-7079. <https://doi.org/10.1007/s11229-021-03105-8>
- Selberg, A. (1949). An elementary proof of Dirichlet's theorem about primes in an arithmetic progression. *Annals of Mathematics*, 50(2), 297-304.
- Sellars, W. (1953). Inference and meaning. *Mind*, 62(247), 313-338.
- Scharp, K. (2013). *Replacing Truth*. Oxford: Oxford University Press.
- Schellekens, M. (2018). No-fault compensation schemes for self-driving vehicles. *Law, Innovation and Technology*, 10(2), 314-333.
- Schlag, C. (2012). The new privacy battle: How the expanding use of drones continues to erode our concept of privacy and privacy rights. *Pittsburgh Journal of Technology Law & Policy*, 13, i.

- Schleidgen, S., Dederer, H. G., Sgodda, S., Cravcisin, S., Lüneburg, L., Cantz, T., & Heinemann, T. (2020). Human germline editing in the era of CRISPR-Cas: risk and uncertainty, inter-generational responsibility, therapeutic legitimacy. *BMC Medical Ethics*, 21(1), 87.
- Shariff, A., Bonnefon, J. F., & Rahwan, I. (2021). How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. *Transportation research part C: emerging technologies*, 126, 103069.
- Shemanske, T. R. (2017). *Modern Cryptography and Elliptic Curves* (Vol. 83). American Mathematical Society.
- Shep, S. J. (2015). Digital materiality. In S. Schreibman, R. Siemens, and J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 322-330). Wiley. <https://doi.org/10.1002/9781118680605.ch22>
- Shneiderman, B. (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (2nd ed.). Addison-Wesley.
- Simion, M. (2018a). Epistemic Trouble for Engineering ‘Woman’. *Logos & Episteme*, 9(1), 91–98.
- Simion, M. (2018b). The ‘should’ in conceptual engineering. *Inquiry*, 61(8), 914–928.
- Simion, M., & Kelp, C. (2020). Conceptual innovation, function first. *Noûs*, 54(4), 985–1002.
- Sohn, L. B. (1973). The Impact of Technological Changes on International Law. *Washington & Lee Law Review*, 30, 1.
- Solove, D. J. (2010). *Understanding Privacy*. Harvard University Press
- Soltanzadeh, S. (2015). Humanist and Nonhumanist Aspects of Technologies as Problem Solving Physical Instruments. *Philosophy and Technology*, 28(1), 139–156. <https://doi.org/10.1007/s13347-013-0145-4>
- Soltanzadeh, S. (2016). Questioning Two Assumptions in the Metaphysics of Technological Objects. *Philosophy and Technology*, 29(2), 127–135. <https://doi.org/10.1007/s13347-015-0198-7>

- Soltanzadeh, S. (2019). A Practically Useful Metaphysics of Technology. *Techné*, 23(2), 232–250. <https://doi.org/10.5840/techne2019924103>
- Soltanzadeh, S. (2021). *Problem-Solving Technologies. A User-Friendly Philosophy*. Rowman & Littlefield.
- Sommaggio, P., & Marchiori, S. (2018). Break the chains: A new way to consider machines' moral problems. *Biolaw Journal*, 5(3), 241-257. <https://doi.org/10.15168/2284-4503-339>
- Stahl, B. C. (2011). IT for a better future: how to integrate ethics, politics and innovation. *Journal of Information, Communication and Ethics in Society*, 9(3), 140-156.
- Stahl, B. C., & Flick, C. (2011). ETICA workshop on computer ethics: Exploring normative issues. *Privacy and Identity Management for Life: 6th IFIP WG, Helsingborg, Sweden, August 2-6, 2010, Revised Selected Papers 6*, 64–77. https://doi.org/10.1007/978-3-642-20769-3_6
- Stahl, B. C., Heersmink, R., Goujon, P., Flick, C., Van Den Hoven, J., Wakunuma, K., ... & Rader, M. (2012). Identifying the ethics of emerging information and communication technologies: an essay on issues, concepts and method. In *Ethical Impact of Technological Advancements and Applications in Society* (pp. 61-79). IGI Global Scientific Publishing.
- Stahlberg, D., Braun, F., Irmen, L., & Szczesny, S. (2011). Representation of the sexes in language. In *Social communication* (pp. 163–187). Psychology Press.
- Stilgoe, J. (2021). How can we know a self-driving car is safe?. *Ethics and Information Technology*, 23(4), 635-647.
- Suh, N. P. (2021). What is Design?. In *Design Engineering and Science* (pp. 35-69). Cham: Springer International Publishing.
- Sullins, J. P. (2009). Artificial moral agency in technoethics. In R. Luppigini & R. Adell (Eds.), *Handbook of Research on Technoethics* (pp. 205-221). IGI Global. <https://doi.org/10.4018/978-1-60566-022-6.ch014>

- Summers, C. (2017). Understanding Translated Authorship. In *Examining Text and Authorship in Translation: What Remains of Christa Wolf?* (pp. 23-57). Cham: Springer International Publishing.
- Swierstra, T. (2015). Identifying the normative challenges posed by technology's 'soft' impacts. *Etikk i praksis-Nordic Journal of Applied Ethics*(1), 5-20.
- Swierstra, T., Stemerding, D., & Boenink, M. (2009). Exploring technomoral change: the case of the obesity pill. In *Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technology Developments*. (pp. 119-138). Dordrecht: Springer Netherlands.
- Swierstra, T., & Waelbers, K. (2012). Designing a good life: A matrix for the technological mediation of morality. *Science and Engineering Ethics*, 18, 157–172.
- Tahir Gürçağlar, Ş. (2010). Scouting the borders of translation: Pseudotranslation, concealed translations and authorship in twentieth-century Turkey. *Translation Studies*, 3(2), 172-187.
- Tao, T. (2011). Structure and randomness in the prime numbers. In *An Invitation to Mathematics: From Competitions to Research* (pp. 1-7). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Tavani, H. T. (2017). Balancing intellectual property rights and the intellectual commons: A Lockean analysis. In *Computer Ethics* (pp. 85-94). Routledge.
- Thagard, P. (1993). *Conceptual Revolutions*. Princeton University Press.
- Thomasson, A. L. (2017). Metaphysical disputes and metalinguistic negotiation. *Analytic Philosophy*, 58(1), 1-28.
- Thomasson, A. (2020) A pragmatic method for conceptual ethics. In Cappelen, H., D. Plunkett, and A. Burgess (Eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 435–458). Oxford: Oxford University Press.
- Thomasson, A. (2021). Conceptual engineering: When do we need it? How can we do it? *Inquiry*. <https://doi.org/10.1080/0020174X.2021.2000118>

- Thomasson, A. (2022). How should we think about linguistic function?. *Inquiry*. <https://doi.org/10.1080/0020174X.2022.2074886>
- Tollon, F. (2022). Artifacts and affordances: from designed properties to possibilities for action. *AI & Society*, 37, 239-248. <https://doi.org/10.1007/s00146-021-01155-7>
- Tromp, N., Hekkert, P., & Verbeek, P.-P. (2011). Design for socially responsible behavior: A classification of influence based on intended user experience. *Design Issues*, 27(3), 3–19.
- Turoff, M., & Linstone, H. (1975). *The Delphi method: Techniques and applications*. Addison-Wesley Publication Company.
- Uluğ, E., & Bayraktar, A. (2014). Determination of the required features of children's picture books. *Journal of Education and Future*, (6), 25-42.
- Umbrello, S. (2024). *Technology ethics: Responsible innovation and design strategies*. John Wiley & Sons.
- van de Poel, I. (2020). Core Values and Value Conflicts in Cybersecurity: Beyond Privacy Versus Security. In: M. Christen, B. Gordijn, and M. Loi (Eds.), *The Ethics of Cybersecurity. The International Library of Ethics, Law and Technology, Volume 21* (pp. 45-71). Cham: Springer. https://doi.org/10.1007/978-3-030-29053-5_3
- van de Poel, I. (2022). Socially Disruptive Technologies, Contextual Integrity, and Conservatism About Moral Change. *Philosophy & technology*, 35(3), 82.
- van de Poel, I., Frank, L., Hermann, J., Hopster, J., Lenzi, D., Nyholm, S., Taebi, B., & Ziliotti, E. (Eds.). (2023). *Ethics of Socially Disruptive Technologies: An Introduction*. Open Book Publishers. <https://doi.org/10.11647/OBP.0366>
- van de Poel, I., & Kudina, O. (2022). Understanding Technology-Induced Value Change: a Pragmatist Proposal. *Philosophy & Technology*, 35(2), 40.
- Väyrynen, P. (2013). *The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics*. Oxford University Press.
- Vellinga, N. E. (2017). From the testing to the deployment of self-driving cars: Legal challenges to policymakers on the road ahead. *Computer Law & Security Review*, 33(6), 847-863.

- Veluwenkamp, H. (2025). Impactful conceptual engineering: Designing technological artefacts ethically. *Ethical Theory and Moral Practice*, 28(1), 155-170.
- Veluwenkamp, H., Capasso, M., Maas, J., & Marin, L. (2022). Technology as driver for morally motivated conceptual engineering. *Philosophy & Technology*, 35(3), 71.
- Veluwenkamp, H., Hopster, J., Köhler, S., & Löhr, G. (2024). Socially disruptive technologies and conceptual engineering. *Ethics and Information Technology*, 26(4), 65.
- Veluwenkamp, H., & van den Hoven, J. (2023). Design for values and conceptual engineering. *Ethics and Information Technology*, 25(1). <https://doi.org/10.1007/s10676-022-09675-6>
- Venuti, L. (1995). Translation, authorship, copyright. *The Translator*, 1(1), 1-24. <https://doi.org/10.1080/13556509.1995.10798947>
- Verbeek, P.-P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press.
- Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation. *Science, Technology, & Human Values*, 31(3), 361–380.
- Verbeek, P. P. (2008). Morality in design: Design ethics and the morality of technological artifacts. In *Philosophy and design: From engineering to architecture* (pp. 91-103). Dordrecht: Springer Netherlands.
- Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. University of Chicago Press.
- Verbeek, P. P. (2012). Expanding mediation theory. *Foundations of Science*, 17(4), 391-395.
- Verbeek, P.-P. (2015). Beyond interaction: A short introduction to mediation theory. *Interactions*, 22(3), 26–31.
- Verbeek, P. P. (2022). The empirical turn. *The Oxford Handbook of Philosophy of Technology*, 35-54.
- Vermaas, P. E., & Houkes, W. (2003). Ascribing functions to technical artefacts: A challenge to etiological accounts of functions. *The British Journal for the Philosophy of Science*, 54(2), 261-289.

- Vermaas, P. E., & Houkes, W. (2006). Technical functions: a drawbridge between the intentional and structural natures of technical artefacts. *Studies in History and Philosophy of Science Part A*, 37(1), 5-18.
- Vermaas, P. E. (2016). An Engineering Turn in Conceptual Analysis. *Philosophy of Engineering and Technology*, 23, 269–282. https://doi.org/10.1007/978-3-319-33717-3_15
- Vissers, C. A., Pires, L. F., Quartel, D. A. C., & Van Sinderen, M. (2016). *Architectural Design*. Springer International Publishing.
- Vogus, T. J., & Sutcliffe, K. M. (2007). Organizational resilience: Towards a theory and research agenda. In *2007 IEEE international conference on systems, man and cybernetics* (pp. 3418-3422). IEEE.
- Von Schomberg, R. (2013). A vision of responsible research and innovation. In R. Owen, M. Heintz and J. Bessant (Eds.), *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society* (pp. 51-74). London: John Wiley.
- Vyas, D., Chisalita, C. M., & Van Der Veer, G. C. (2006). Affordance in interaction. *Proceedings of the 13th European Conference on Cognitive Ergonomics: Trust and Control in Complex Socio-Technical Systems*, 92–99.
- Waelbers, K., & Swierstra, T. (2014). The family of the future: how technologies can lead to moral change. In *Responsible Innovation 1: Innovative Solutions for Global Issues* (pp. 219-236). Dordrecht: Springer Netherlands.
- Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, adaptability and transformability in social–ecological systems. *Ecology and society*, 9(2), 5.
- Walker, B., & Salt, D. (2012). *Resilience thinking: Sustaining ecosystems and people in a changing world*. Island Press.
- Walker, S. (2023). *Design for Resilience*. MIT Press.
- Waltz, M., Walker, R. L., Flatt, M. A., MacKay, D., Conley, J. M., Juengst, E. T., & Cadigan, R. J. (2024). Challenging the boundaries between treatment, prevention, and enhancement in human genome editing. *The CRISPR Journal*, 7(4), 180-187.

- Warren, B., & Brandeis, L. (1890). The Right to Privacy. *Harvard Law Review*, 4(5), 193-220.
- Wasmuth Jr, C. E. (1969). The concept of death. *Ohio State Law Journal*, 30, 32.
- Watiktinnakorn, C., Seesai, J., & Kerdvibulvech, C. (2023). Blurring the lines: how AI is redefining artistic ownership and copyright. *Discover Artificial Intelligence*, 3(1), 37.
- Wendehorst, C. (2020). Strict liability for AI and other emerging technologies. *Journal of European Tort Law*, 11(2), 150-180.
- Whorf, B. L. (2012). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.
- Williams, B. (1985). *Ethics and the Limits of Philosophy*. Harvard University Press
- Wilson, R., & Gerber, L. (2012). *Creative constraints: translation and authorship*. Monash University Publishing.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19), 7780–7785.
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. Routledge & Kegan Paul.
- Wittgenstein, L. (1953). *The Philosophical Investigations*. Oxford: Blackwell.
- Woodfield, A. (1991). Conceptions. *Mind*, 100(4), 547-572.
- Xiao X. (2008). Ancient Greek and Chinese Patterns of Definition: A Comparative Study. *Intercultural Communication Studies*, 7 (2), 61-77.
- Yan, J., & Du, J. (2020). Multiple authorship of translated literary works: A study of some Chinese novels in American publishing industry. *Translation Review*, 106(1), 15-34.
- Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton University Press

- Yusof, M. F. M., Romli, N., & Yusof, M. F. M. (2014). Design for elderly friendly: Mobile phone application and design that suitable for elderly. *International Journal of Computer Applications*, 95(3).
- Zeller, B. (2000). On translation and authorship. *Meta*, 45(1), 134-139.
- Zeng, Y., Gordiichuk, P., Ichihara, T., Zhang, G., Sandoz-Rosado, E., Wetzel, E. D., ... & Strano, M. S. (2022). Irreversible synthesis of an ultrastrong twodimensional polymeric material. *Nature*, 602(7895), 91-95.
- Zhou, X., & Tang, X. (2011). Research and implementation of RSA algorithm for encryption and decryption. In *Proceedings of 2011 6th International Forum on Strategic Technology* (Vol. 2, pp. 1118-1121). IEEE.

About the author

Samuela Marchiori completed her PhD in conceptual engineering in the philosophy of technology at Delft University of Technology, as part of the gravitation programme Ethics of Socially Disruptive Technologies (ESDiT), where she was a Research Fellow between 2022 and 2026. During her time at TU Delft, she was a visiting scholar at LMU Munich and at the University of Edinburgh, received (with B. Lundgren and M. Vanžura) the 2023 SPT Early Career Award, and was named one of the Rising Stars in AI Ethics of 2025. Prior to joining TU Delft, she acquired graduate degrees in law (JD/LLM 2020, University of Trento) and philosophy of technology (MSc 2022, University of Twente), and was a student researcher at Columbia University (2019). Her research lies at the intersection of the legal and moral philosophy of technology, with a focus on the development of methodologies for the ethical governance of new and emerging technologies, particularly AI. Her work has been published in *Ethics and Information Technology*, the *Journal of Medical Ethics*, *Science and Engineering Ethics*, *Inquiry*, *Philosophy and Technology*, and *Synthese*.

List of publications

- Marchiori, S. (2026). Concepts as artefacts. Three insights that conceptual engineering can draw from the philosophy of technology. *Inquiry*, 1-25.
- Marchiori, S., & Sta. Maria, J. (2026). Conceptualising conceptual resilience. A comparative approach. *Philosophy & Technology*, 39(45).
- Marchiori, S., Hopster, J. K., Puzio, A., Riemsdijk, M. B. V., Kraaijeveld, S. R., Lundgren, B., Viehoff, J., & Frank, L. E. (2025). A social disruptiveness-based approach to AI governance: complementing the risk-based approach of the AI Act. *Science and Engineering Ethics*, 31(5), 25.
- Marchiori, S. (2025). Conceptual affordances: (How) should they inform conceptual engineering?. *Synthese*, 206(2), 88.
- Rueda, J., Segers, S., Hopster, J., Kudlek, K., Liedo, B., Marchiori, S., & Danaher, J. (2025). Anticipatory gaps challenge the public governance of heritable human genome editing. *Journal of Medical Ethics*, 51(5).
- Marchiori, S., & Scharp, K. (2024). What is conceptual disruption?. *Ethics and Information Technology*, 26(1), 18.
- Hopster, J., Brey, P., Klenk, M., Löhr, G., Marchiori, S., Lundgren, B., & Scharp, K. (2023). Conceptual disruption and the ethics of technology. In I. van de Poel, L. E. Frank, J. Hermann, J. Hopster, D. Lenzi, S. Nyholm, B. Taebi, and E. Ziliotti (Eds.), *Ethics of Socially Disruptive Technologies: An Introduction* (pp. 141-162). Open Book Publishers.
- Marchiori, S. (2022). MaxSAT meets Alexy: A computational approach to solve trolley problem-like scenarios in the legal sphere. *Journal of Ethics and Legal Technologies*, 4(2), 105-116.
- Sommaggio, P., & Marchiori, S. (2020). Health data ethics in the time of covid-19: A legal perspective. *Humanidades & Tecnologia em Revista*, 2020(1), 95-101.

- Sommaggio, P., & Marchiori, S. (2020). Moral dilemmas in the AI era: A new approach. *Journal of Ethics and Legal Technologies*, 2(1), 89-102.
- Sommaggio, P., & Marchiori, S. (2020). Tragic choices in the time of pandemics. *BioLaw Journal*, 2020(SI 1), 453-458.
- Sommaggio, P., & Marchiori, S. (2018). Break the chains: a new way to consider machines' moral problems. *BioLaw Journal*, 2018(3), 241-257.

This page intentionally contains only this sentence.

While a rich scholarship has examined how technology impacts our practices, norms, values, and beliefs, considerably less explicit attention has been devoted to investigating how technology impacts our concepts, i.e., the notions through which we understand, reason about, and navigate our world. For example, the emergence of autonomous weapons systems has put pressure on the concept CONTROL adopted in international humanitarian law, which proved inadequate in establishing clear lines of responsibility for uses of force in contexts in which control over weapons was no longer exercised directly by a human agent, but distributed among human and non-human entities. Similarly, the advent of autonomous vehicles has raised pressing questions about (among others) liability regimes, safety standards, and the differential impact of this technology across different social groups. Many of such questions are tied to the concepts that underlie them.

This dissertation investigates how to navigate technology-induced conceptual disruption (i.e., the interruption in the functional adequacy of concepts) with conceptual engineering (i.e., the evaluation, design, and implementation of concepts). The dissertation is a sustained illustration of the claim that conceptual engineering and the philosophy of technology—two disciplines which have so far proceeded on largely parallel tracks despite engaging with related problems, operating on similarly conceived objects of inquiry, and sharing broadly comparable normative aims—are better positioned to address their respective research in dialogue rather than in isolation. This has led to contributions to both conceptual engineering—illustrating how insights and tools developed for the study of socio-technical artefacts can enrich the theoretical foundations and operationalisation of conceptual engineering—and the philosophy of technology—showing how the vocabulary of conceptual engineering can equip philosophers of technology with sharper tools for identifying, evaluating, and responding to conceptual disruptions.

