PASS THE BALL

Designing accountability into the socio-technical context of algorithmic systems

Master Thesis Design for Interaction & Science Communication

Laurien Albeda

Master Thesis

by Laurien Johanna Albeda

in partial fulfillment of the
requirements for the degree of

MSc Design for Interaction
MSc Science Communication

at Delft University of Technology

Thesis Defence: September 20, 2018

Supervisory team

Industrial Design Engineering
Prof. dr. Gerd Kortuem
Dr. ir. Nynke Tromp

Science Communication
Dr. Maarten van der Sanden
Dr. ir. Steven Flipse
Prof. dr. Marc de Vries

PASS THE BALL

Designing accountability into the socio-technical context of algorithmic systems

Preface

With this Master Thesis, my combined MSc programmes of Design for Interaction and Science Communication come to a close. In the nine month project that it describes, I tried to fully integrate these two disciplines for the first time.

To explain the value of combining these studies, I often talk in extremes. I would say Design for Interaction is about zooming in, understanding the context of people's behaviour and using a creative process to come up with solutions to improve their well-being. Science Communication is about analysing the broader, societal context, considering the various stakeholders involved, and using theory and a strong rationale to improve communication and collaboration.

Naturally, it isn't as black and white as that. Elements of these approaches can be found in both disciplines. Moreover, there is a clear similarity between the two: both fields play out at the crossroads of humans and technology. The topic of algorithms, their place in our society and what our future interactions with smart systems will look like therefore seemed a challenging but appealing subject for a final, interdisciplinary project.

It was with the help of tools and skills that both Masters have given me, that I could unravel this complex topic and come to find that this vague concept 'algorithmic accountability' was really about passing the ball to others instead of dribbling to the basket on your own. But more on that later.

At times, this project has felt like quite a hike. Luckily, *you never walk alone*...

First of all, I'd like to thank my supervisory team for coaching me from start to finish and always challenging me to improve. Maarten, thanks for all the inspiring and uplifting talks. Steven, for the constructive feedback and your confidence in a good result. Gerd, for the good discussions and of course the intriguing research topic. Nynke, for your sharp observations and remarks.

Second, thanks to my interviewees and research participants for giving up your time for my research. Thanks as well to Vincent, Marijn and Joris for acting out the design fiction scenarios.

Next, a shout-out to the fellow graduate students with whom I could 'spar' or just relieve stress whenever necessary. Josephine and Linda, our Monday morning sessions helped me start up my project. Jelmer, our discussions always helped me see my results in a different light, and Vera, it was great to have you as a buddy for the last miles. Also, thanks to Eline, Ruben, Charlot, Maaike, Pelin, Zoë, Sarah and Marijke for helping me along the way.

For all the moral support, I'd like to thank the other students in the SEC graduation room, my friends, and my family. Thank you, Inger, for literally walking with me on our trainings, which took my mind off the project for a day. And finally, thank you Sander, for marching beside me all along, always ready to offer advice or a pep talk.

Now, it is time to turn towards the subject at hand. After an introductory part, I will guide you through my four research cycles, ending with a closing part that presents my conclusions and discussions. May it be a thought-provoking journey!

Laurien Albeda
Delft, September 2018

Summary

Research aim
Algorithms are all around us. And while they may facilitate many of our daily tasks, they also assert a certain power over us. This algorithmic governance has led to authors to discuss how we, in turn can govern these algorithms. There are some challenges here, because of the opaque and automated nature of algorithms. Also, they should be considered within the socio-technical context in which they are produced and employed. Accountability of algorithms is increasingly seen as an important requirement for a fair and controlled working of algorithms. The aim of this research was to build towards a framework that supports designing accountability into the socio-technical context in which algorithms exist. This means that both the technical attributes of algorithms and the social aspects of the people that interact with them should be considered. The research question was defined as: *How can algorithmic accountability be implemented by design?*

Approach
This research followed a case study approach, combining the findings from a literature review with the insights from the Vehicle to Grid case study to answer the research question. The research process can be described in four cycles.

Process
In the first cycle, algorithmic accountability was explored through a design excursion and then defined in a theoretical framework based on a literature review.

This showed that algorithmic accountability is highly dependent on the context. The theoretical framework proposed five aspects that are important in achieving accountability: a clearly defined responsibility, explainability of the working, debatability of the algorithmic decisions, auditability of the outcomes and sanctionability when the account that is given is insufficient.

With this knowledge of algorithmic accountability, the case study of the Vehicle to Grid project by Alliander, Newmotion and Enervalis was addressed in the second cycle. In a Vehicle to Grid (V2G) system electric vehicles are regarded as driving batteries: when linked to the energy grid, EVs can help in balancing the energy supply and demand, by using the car battery to either store energy surplus or draw energy in case of a shortage of energy.

Through interviews with representatives of the stakeholders, barriers and opportunities for achieving algorithmic accountability were identified. After analysing the relations between these barriers and opportunities in a causal loop diagram, a focus on the interest of the user in the working of the V2G system was chosen. By hightening this interest, the user would be expected to engage more with the working of the system, and this would encourage the implementation of explanations and debating opportunities.

The problems that might arise from the automated decision-making by the V2G unit were also investigated, based on the literature and based on improvisations in acting out the V2G unit and a car driver. An individual problem that might arise is that of an EV driver arriving at his car and finding it uncharged - with no course of action to take. On a larger scale, the V2G system may develop to disadvantage certain groups of users, without them realising this or being able to intervene.

In the third cycle, the user perspective was explored in a contextmapping session with EV drivers. In analysing the results, the value of freedom to travel anywhere, anytime and the need for control over smart technology were found to be relevant for the V2G context. Undesired developments of the algorithmic system based on this value and need were captured in a dystopian vision and then turned into a utopian vision. This was the stepping stone to formulating a design challenge: *'designing an interaction between the user and the V2G unit that optimises the system while providing relevant transparency to the user'*

In the fourth and final cycle, ideas were generated for this design challenge. An interaction vision was found that captured the interaction that the design should lead to: for the user, it should feel like the ball is being passed towards him during a basketball match. It's a clear sign of trust and you feel you can influence the game. With this in mind, the final design was created.

The design was represented in an interface for the car, but contained solutions that went beyond this. These accountability mechanisms were related to the five aspects of algorithmic accountability. For instance, explanations about the charging level are offered to the user by showing the conditions for the supply and demand of energy. Also, the user can review his charging sessions and debate the working of the system by sending in one or multiple records of these sessions.

Conclusion

In a reflection, the design process was reviewed to find out what were effective methods and what parts of the process might be adapted in designing for algorithmic accountability. This provided starting points for the conclusion.

Based on the insights from research and design activities in this research, I conclude d that the following guidelines will structure the process of designing accountability mechanisms:

(1) Understand the context in which the algorithm is created and employed;
(2) predict unwanted developments or effects of the system;
(3) design context-specific accountability mechanisms based on these findings. There are four aspects to consider in designing these mechanisms: how to explain the decisions of the algorithm to the user, how to offer a course of action to debate the algorithmic outcomes, through what data and by whom auditing should take place and how to sanction the responsible parties when the account of the system's working doesn't suffice.

Recommendations for further research

To further develop these guidelines, I propose treating them as a first iteration . By applying them in other design case studies, they can be further extended and detailed. In particular, more insights are required for how the accountability mechanisms might be further implemented and how the aspects of auditing and sanctioning might be established as a part of algorithmic accoutnability.

Contents

Setting the stage

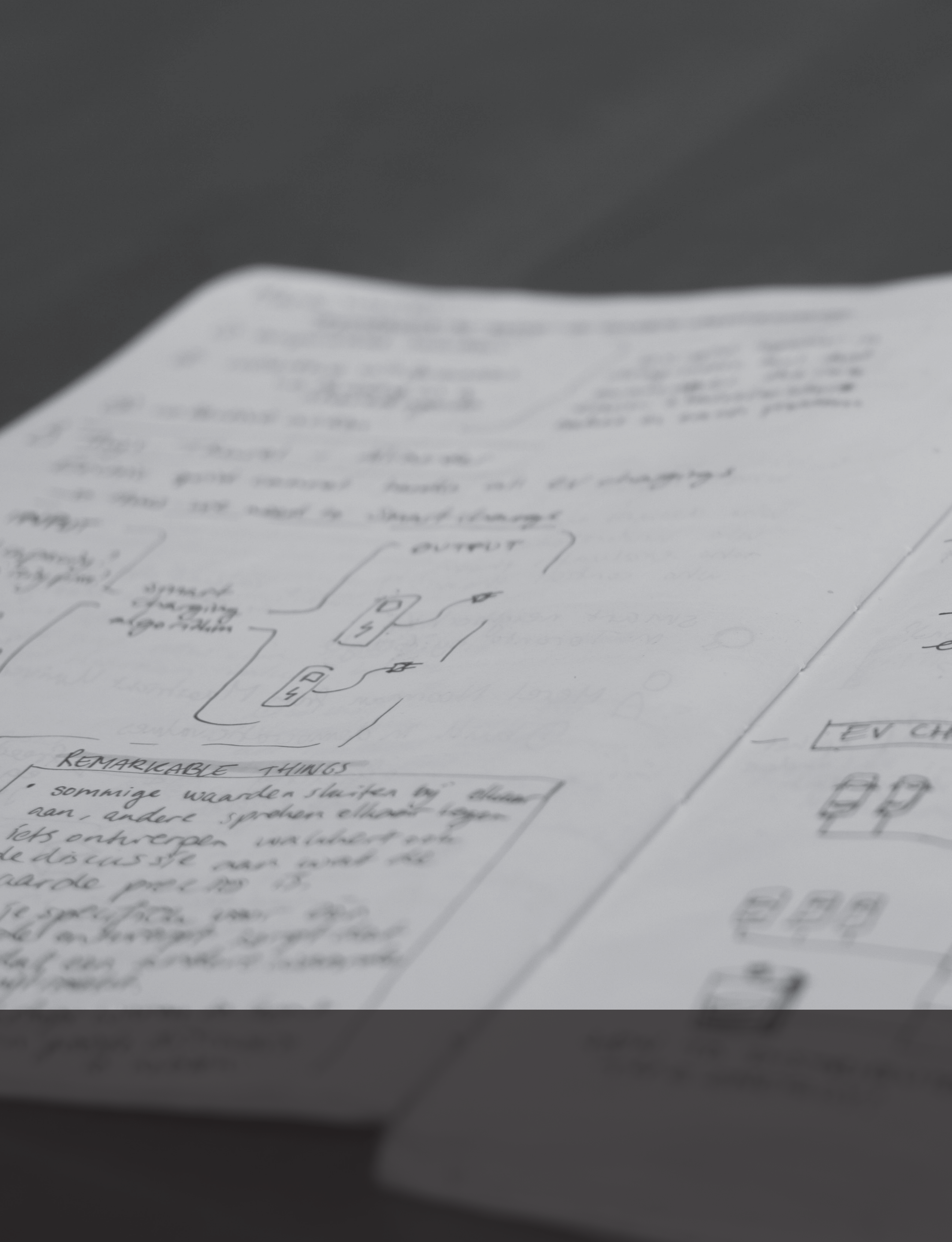
I. Exploring algorithmic accountability

II. Deep dive into V2G

III. Framing the design challenge

IV. Solution finding

The final act


OUTPUT
smart charging algorithm
REMARKABLE THINGS
• sommige waarden sluiten bij elkaar aan, andere spreken elkaar tegen

Setting the stage

This first part of the thesis sets the stage on which the rest of the project can unfold. The topic and research aim are introduced, and the research approach is outlined.

1. Introduction

Algorithms. A term that you might hear in the news, sometimes decribing their great achievements, sometimes warning for their artificial smartness. But what is an algorithm exactly? And what is all the fuss about? This introduction will try to answer these questions (1.1) as it works towards the aim (1.2) and research questions (1.3) of my thesis. The relevance of this research to the fields of Design for Interaction and Science Communication will be discussed as well (1.4).

1.1. Background

1.1.1. The rise of algorithms

Algorithms are increasingly integrated in our daily lives. Throughout your day, algorithms decide what you see in your Facebook News Feed, which restaurants will be shown first when you Google for a place to eat, and what songs you listen to on Spotify. When you apply for a job, your CV might even be automatically screened by algorithms (Žliobaitė, 2017).

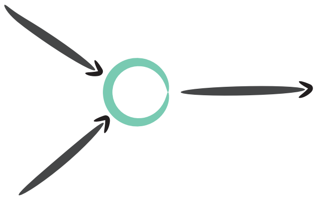
Simply put, an algorithm is a series of steps that follow up on each other to produce a certain output. In most cases, these series are put together in a whole network of algorithms (Kitchin, 2017). Bbox 1 on page 19 further illustrates this. Algorithms can make our lives easier, because they can tackle complex tasks involving large amounts of data. For companies, the automation can lead to a significant reduction in costs and to opportunities for new services (Kitchin, 2017).

1.1.2. Algorithmic flaws

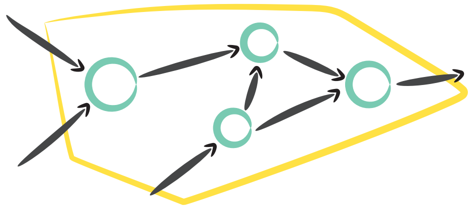
While algorithms and machine learning have shown impressive results, performing tasks that a human brain couldn't possibly achieve, they should not be regarded as flawless or neutral. After all, the programmers that develop them are only humans. Besides this, the real-life data that is used to train the algorithms will also contain real-life biases. Algorithms that prioritise, classify, associate and filter data therefore sometimes result in unfair differential treatment and discrimination (Vedder & Naudts, 2017). For example, an algorithm that is set to decide the most suitable candidate for a CEO, will find few records of successful female CEOs in its input data, and might therefore decide that 'being a man' is an accurate predictor of 'being a successful CEO'.

Zarsky (2016) summarises the main concerns of algorithmic decision-making as being the fairness of the decisions and the efficiency. In the case of the CEO, this algorithm would be both unfair to female candidates and might also lead to a bad choice for a CEO, undermining the added value of an algorithmic decision-making process.

Because of the flaws of algorithms, questions an algorithm

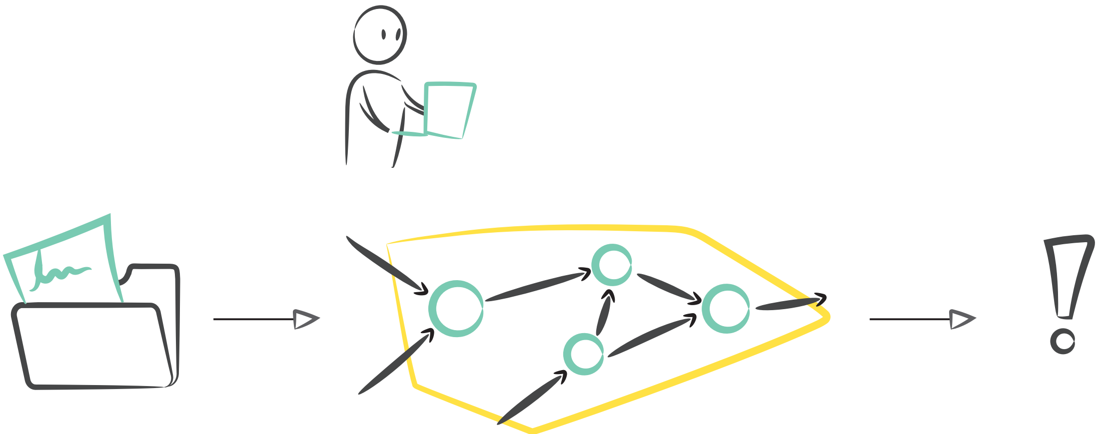


a network of algorithms



algorithmic decision-making

a programmer develops the network of algorithms



data input... ...through the algorithms... ...leads to a decision

Often when an 'algorithm' is discussed, a network of algorithms is meant. This network is in fact a model that uses data to form a certain decision. A programmer has decided which input is relevant (which 'proxies') to come to a correct decision. For example, Amazon's algorithms calculate a product's overall rating by taking into account proxies such as the individual rating, the age of the review and the helpfulness votes that other customers have given. (Eslami, Vaccaro, Karahalios, & Hamilton, 2017)

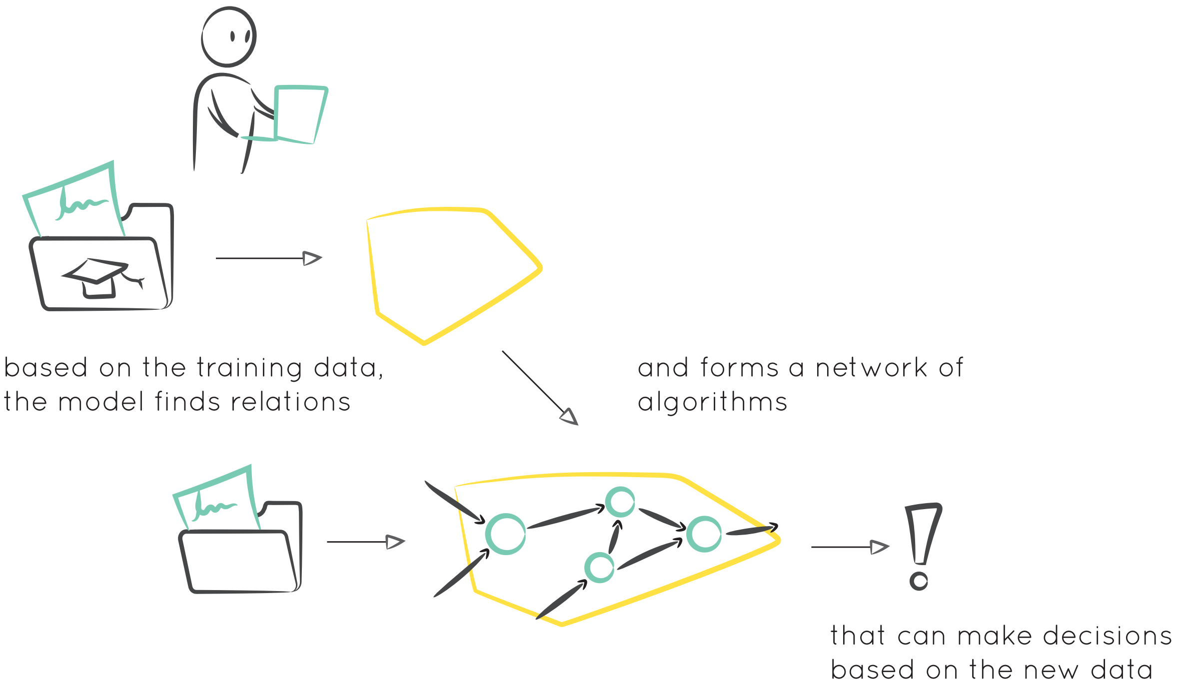
box 1 | the basics of algorithmic decision-making

have been raised in literature about how we can govern the algorithms that in turn govern us (Kitchin, 2017; Saurwein, Just, & Latzer, 2015; Danaher, 2016). This should help to prevent the risks and reap the benefits. Unfortunately, this is no easy feat.

1.1.3. Complicating characteristics of algorithms

First of all, the characteristics of algorithms complicate matters. According to Zarsky, the main attributes that exacerbate these concerns are the opacity and the automation of algorithmic processes; an algorithm uses non-transparent

algorithmic decision-making based on machine-learning
a programmer develops the model for a network of algorithms and provides training data.



An algorithmic model can also be instructed to create its own rules by finding patterns in data. In this form of algorithmic decision-making that is known as machine learning, a software developer only provides the outlines of a model, deciding what output it should optimise and providing it with training data. The algorithmic network then forms itself by analysing the training data for relevant variables that should be taken into account to provide the correct output. This happens through an extensive process of trial-and-error. This can be seen, for instance, in visual recognition. After 'seeing' many pictures that contain a dog, a computer will find patterns that predict whether something is a dog or a cat (Sullivan, 2015).

measures in its calculations and automatically analyses the data it collects. As Vedder & Naudts (2017) describe it, algorithms 'seem to function behind an impermeable veil, not allowing of attempts to analyse the decision-making and its outcomes' (p.2). Therefore, an algorithm is often perceived as a black box. Its decisions are hard to trace back and explain, and so it is hard to judge the extent to which they are fair and accurate.

Secondly, an algorithm can't be seen as a separate entity. It is produced in a trial and error process, constantly being edited and extended, based on the various purposes and restrictions of the political, legal, economic and cultural circumstances in which it is created. Moreover, once an algorithmic system is employed in its context of use, it keeps on developing. Kitchin (2017) refers to this as a 'process of domestication', in which the user integrates it in his life and adapts it to his purposes. For example, the algorithm that selects and orders the Facebook posts to be viewed in a user's timeline, adapts based on the way that the user responds to these posts and how he interacts with his Facebook friends (Kitchin, 2017).

1.1.4. The importance of algorithmic accountability

To confine the power that algorithms and the organisations behind them have, algorithmic decision-making should be accountable. Similar to how a government should be accountable to its citizens regarding the use of its power, algorithmic accountability is required if we are to accept the authority of algorithms (Diakopoulos, 2016).

To explain algorithmic accountability, it helps to first define algorithmic transparency, because there is a slight overlap but also a clear difference between the two. Transparency requires an explanation of the purposes of algorithms and the logic behind the outcomes of algorithmic decision-making. For accountability, reasons for the actions of algorithms are important as well, but it further requires oversight mechanisms and correcting the working of algorithms where necessary (Zouave & Marquenie, 2017). In this way, algorithmic accountability offers a way of governing the working of algorithms once they are active.

While academics as well as politicians and regulators have stressed the need for creators of algorithmic systems to disclose information about its working, the talk of algorithmic accountability is mostly still quite abstract. In the General Data Protection Regulation (GDPR), introduced in May 2018, Article 15.1 describes the right for individuals to ask for 'an account of the 'logic' behind automated decisions made about them' (Binns, 2017, p. 5) (see also Box 2). But how exactly this should happen is not yet specified (Vedder & Naudts, 2017).

In the recently published report by the

box 2 | GDPR and algorithmic accountability

The introduction of the General Data Protection Regulation was not easily missed. This new regulation meant that companies had to update their Data Protection, and inform their data subjects of this; hence the many e-mails with privacy policy updates.

While mainly discussing regulations about data controlling, the GDPR also had some implications for algorithmic or automated decision-making:

- Within an algorithmic context, the data controlling party is to be held responsible and accountable for the algorithmic decision-making.
- If a type of processing is likely to have a high risk of violating the rights and freedoms of natural persons involved, a data controller must carry out an impact assessment of the intended processing operations on the protection of personal data.
- The data subject shall have the right not to be subject to a decision that is fully automated and significantly affects him or her. Although not mentioned explicitly, the exercise of the right not to be subject to an automated decision requires strong accountability mechanisms towards the decision subject.
- The GDPR requires data processing activities to be transparent, which entails that any information and communication concerning the processing of personal data should be easily accessible and easy to understand for individuals.

(Vedder & Naudts, 2017)

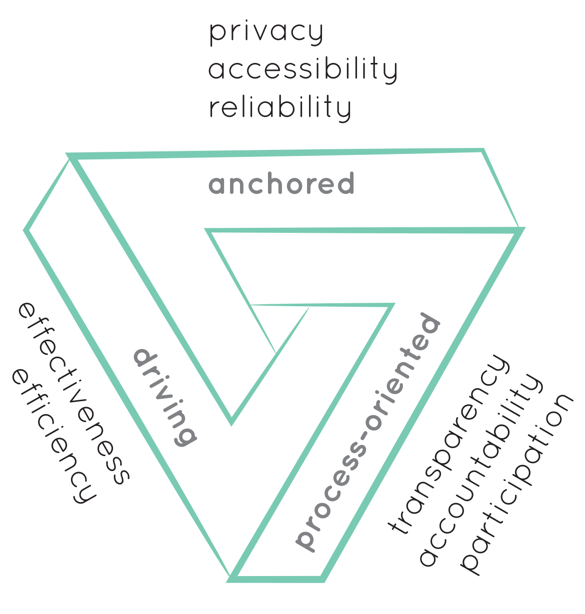


figure 1 | driving, anchored and process-oriented forces in a digitalised society (adapted from Hollander et al., 2017)

“Planning office for the Living Environment” (Planbureau voor de Leefomgeving, PBL), accountability is mentioned as a process-oriented value (figure 1). Along with transparency and participation it should provide more insight into choices that are made and should lead to more measurable decisions in a digitalised society (Hollander, Vonk, Snellen, & Huitzing, 2017). Again, the importance of accountability is stressed but questions about what this accountability should look like remain.

1.2. Research aim

This research explores how algorithmic accountability can be achieved in an algorithmic context.

As Ananny & Crawford (2016) indicate, the way in which accountability can be established will be different for each socio-technical context. What I take from this, is that to design algorithmic accountability into the system, a holistic approach is required that includes all aspects of the system. Since this would make any implementation of algorithmic accountability context-specific, a case has been chosen as an object of study throughout this research. This case study will be briefly introduced in the upcoming section and be further elaborated in paragraph 6.1 on page 46.

The ultimate aim of the research is to take the first steps towards a framework that structures the challenge of implementing accountability, providing possible starting points for designing solutions.

1.2.1. Case study: Vehicle to Grid

The Vehicle2Grid project was initiated by Alliander, an energy grid manager in the Netherlands. (City-zen, n.d.) In 2014, the project was set up to investigate whether the Vehicle2Grid concept could

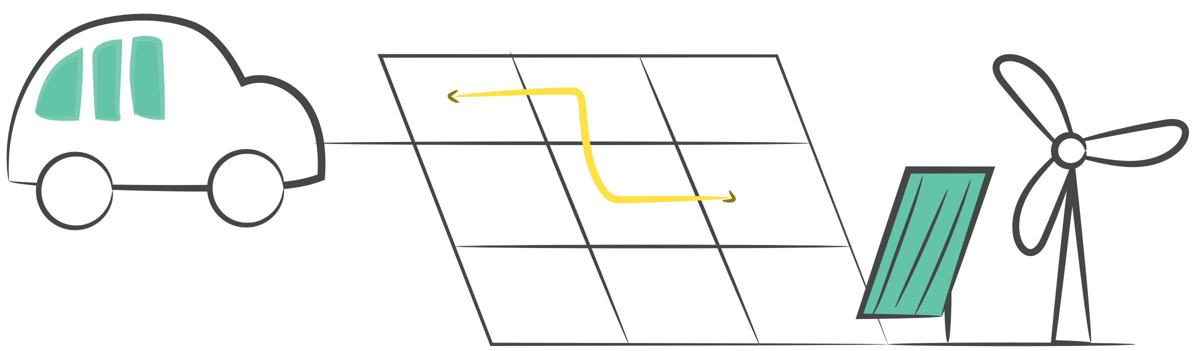


figure 2 | The V2G concept: balancing the grid by using the car batteries as a buffer help in decreasing congestion in the energy grid. In February 2018, a pilot with four participants was kicked off.

In a Vehicle to Grid (V2G) system electric vehicles are regarded as driving batteries: when linked to the energy grid, EVs can help in balancing the energy supply and demand. This can be achieved by using 'smart charging' – postponing the charging of a car, dividing it into bits, or reducing the charging speed – and by using the car battery to either store energy surplus or draw energy in case of a shortage of energy. In the future, electric drivers who use this technology will have the opportunity to receive a fee for their contribution back to the grid.

The V2G concept has until now mainly been researched from a technological point of view. This leaves the question of how users will interact with the technology and how their behaviour will impact the working of the system. In what way might algorithmic accountability 'watch over' these developments to ensure a fair and accurate working?

1.3. Research questions

The following research question was defined to achieve the research aim:

How can algorithmic accountability be implemented by design?

With implementing 'by design', I mean to say that the accountability is a part of how the algorithm operates in its context. The main research question is answered by first addressing the following subquestions:

SQ1. What is algorithmic accountability?

To answer this question, the concept was first studied through a design excursion. Next a review of the literature was executed to gather the current knowledge on algorithmic accountability and define the concept in a framework.

Research question
How can algorithmic accountability be implemented by design?

SQ2. What is the socio-technical context of the Vehicle to Grid case study?

What is required to govern algorithms according to Ananny & Crawford (2016), is to understand the relations between human and non-human agents in the algorithmic system. The context of the Vehicle to Grid case study was therefore investigated through interviews with it stakeholders.

SQ3. What are barriers and opportunities to achieving algorithmic accountability in the V2G algorithmic system?

Based on the framework set up in answer to SQ1, the barriers and opportunities to achieving algorithmic accountability within the V2G case study were investigated.

SQ4. What might an accountable design of the V2G system look like?

Finally, by formulating a design challenge and developing a design solution for the V2G system, a practical implementation of the framework was attempted. With the insights gathered from this design process and the answers to the other questions, the research question will be answered in the conclusion of this thesis.

1.4. Relevance to Science Communication and Design for Interaction

This research integrates the fields of Science Communication (SC) and Design for Interaction (DfI).

The increased use of algorithmic decision-making and other forms of artificial intelligence in all parts of our society will impact many domains, including the DfI field. It is important for designers to adapt their design methodology where necessary, in order to integrate the technology in a user-centred and responsible way. This research will extend the current knowledge on how to do this.

The SC field investigates ways to form the link between society on the one hand and technology and science on the other, by developing decision supportive tools for professionals, stakeholders, to make decisions upon the process of linking science, technology and society. Responsible innovation is also a relevant topic in Science Communication. The Responsible Research and Innovation (RRI) field investigates ways of governing the process of research and innovation, with the objective of "democratically including, early on, all parties concerned in anticipating and discerning how research and innovation can or may benefit society." (Burget, Bardone, & Pedaste, 2017). The results of this research should provide insights for a support tool for practitioners developing algorithmic decision-making networks.

Algorithms are all around us. And while they may facilitate many of our daily tasks, they also assert a certain power over us. This algorithmic governance has led to authors to discuss how we, in turn can govern these algorithms. There are some challenges here, because of the opaque and automated nature of algorithms. Also, they should be considered within the socio-technical context in which they are produced and employed. Accountability of algorithms is increasingly seen as an important requirement for a fair and controlled working of algorithms. The aim of this research is to build towards a framework that supports designing accountability into the socio-technical context in which algorithms exist. This means that both the technical attributes of algorithms and the social aspects of the people that interact with them should be considered.

TO SUMMARISE...

2. Approach

Now that the background and aim of this research have been elaborated, this chapter will present how I plan to arrive at a conclusion that answers the research question (2.1, 2.2). The approach and research process that followed from it also brought about the structure of this report (2.3).

2.1. A case study approach

As has been introduced, algorithmic decision-making should be viewed within the socio-technical system that it is a part of. By definition, these systems are complex. They constantly change through the interactions between humans and the interactions with the algorithm. Such complex systems can't simply be designed, as Snowden (2002) argues. They need to be understood by the interactions that take place within the system. By zooming in like this and probing the complex space, we can start to understand how the system works and how it might be changed. (Snowden, 2002)

Therefore, a case study was chosen in which to examine these relations and interactions: the V2G case study presented in the previous chapter. Within this case study, the challenge of designing algorithmic accountability into the system will be investigated through a design process. According to Edelson (2002), design in research might lead to three useful outcomes: (a) design offers opportunities to gain insights in context that can add to existing theory, (b) design research yields practical findings that can be directly applied, and (c) design research engages researchers in the direct improvement of practices, without being bound by market considerations. Until now, the research that has been done into algorithmic decision-making and its governance is theoretical and philosophical by nature, arguing that algorithmic accountability could be required for fair and accurate algorithms. The research field now requires more practical frameworks and methods. This research focuses on how algorithmic accountability might be achieved and provides examples of how this might work.

Hence, this research follows a case study approach applying design-based research. With the knowledge that arises from this process, the subquestions and research question can be answered. This is shown in figure 3.

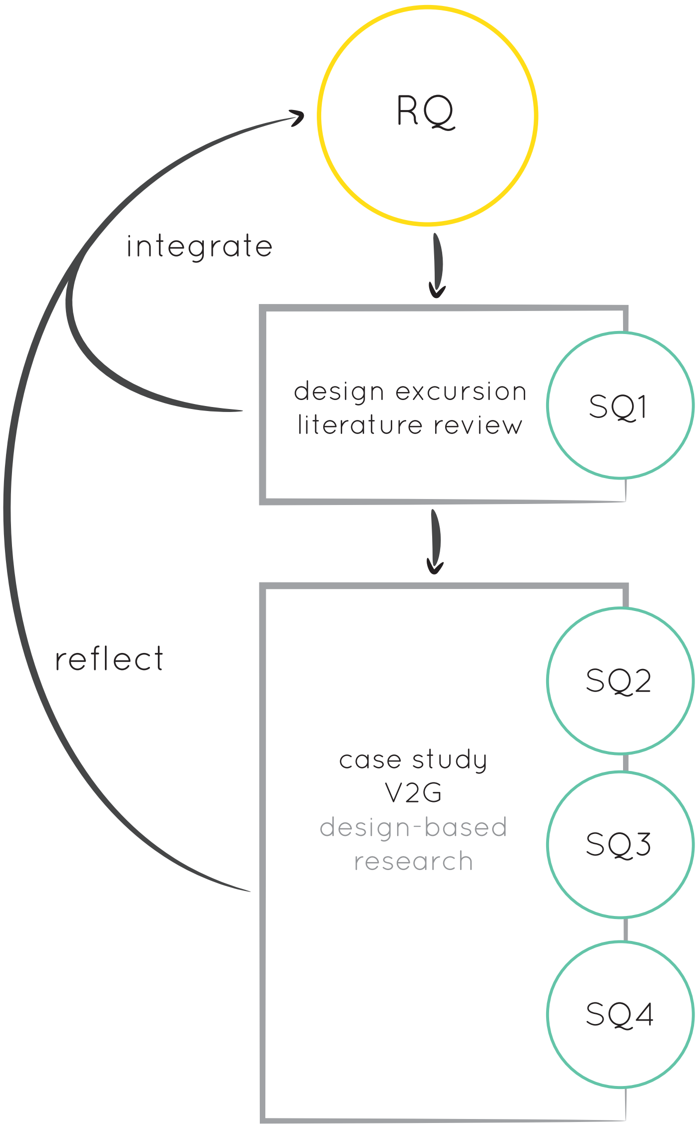


figure 3 | answering the research question and subquestions through the chosen research approach

2.2. Case suitability

The V2G project was considered a suitable case study for this research, because it's still not fully developed but has advanced to a stage where a pilot is being run with a few car drivers in Amsterdam. This means there are less constrictions and stakeholders are more open to changes than in a developed and commercial setting. And yet, there is a concrete project to discuss.

Moreover, the V2G case is a concept that can represent many similar efforts to implement smart system that manage our energy more efficiently and enable a shift from fossil fuels to renewable energy. There is a great urgency to apply such innovative solutions, but the social and ethical challenges that arise from these innovations need to be addressed as well.

Therefore, the V2G project is a relevant and suitable case for this research.

2.3. Report structure

The design process is messy and requires many iterations. This report however, is linear and my objective is to present my findings in a comprehensible way. Looking back, my process can be viewed as having taken four cycles of research, design and concluding, closing off with

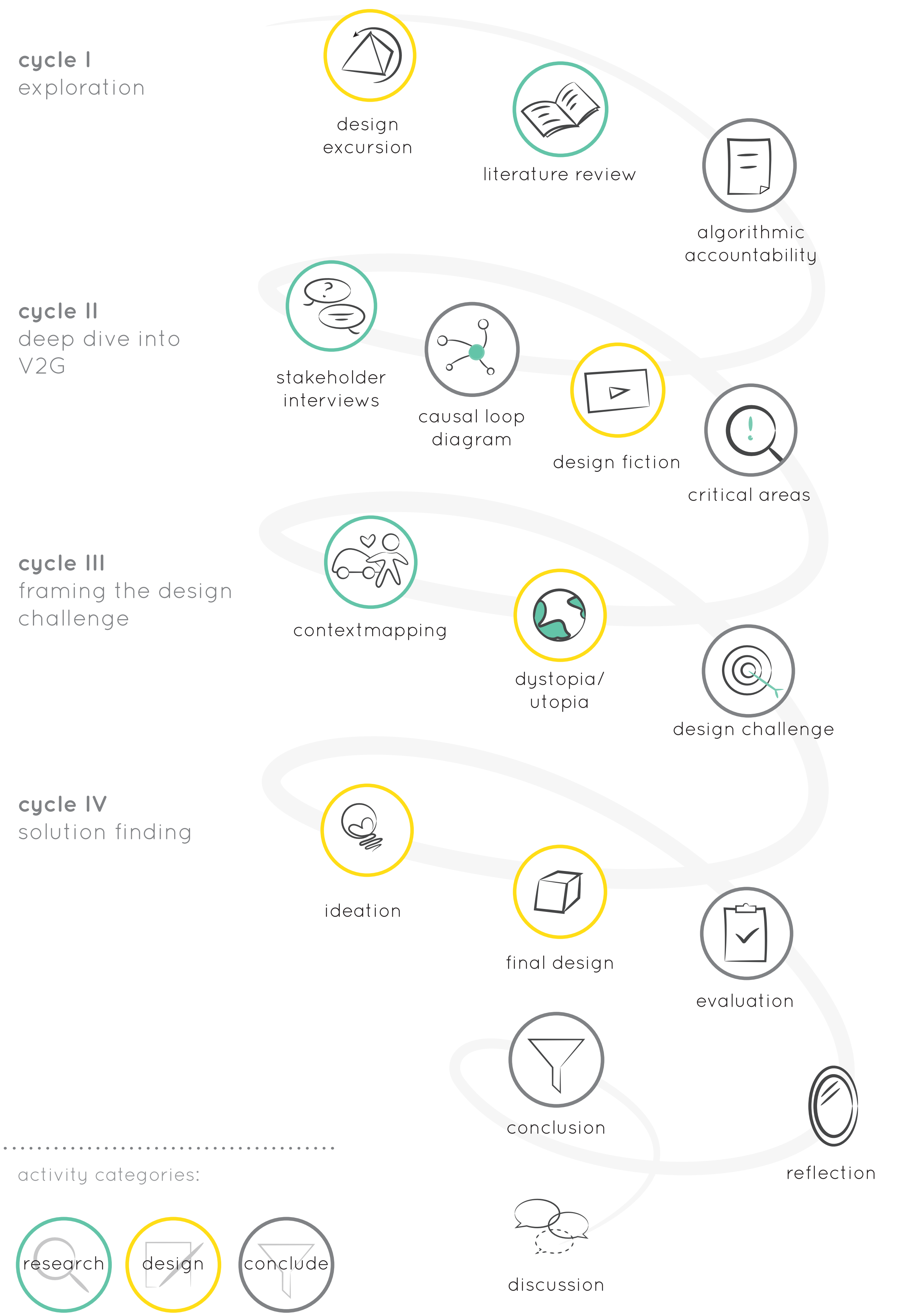


figure 4 | an overview of the four cycles within the research process a reflection, a conclusion and a discussion of the implications of the research. The research and design methods that were used throughout the process have been visualised in figure 4. Each cycle and its methods will be briefly described here. Throughout the report, these methods and their limitations will be elaborated more extensively.

2.3.1. Cycle 1: exploring accountability

In the first cycle, the concept of algorithmic accountability was explored and defined through design excursions and a literature research. The design excursions consisted of two short cases, that sought to define the meaning of governance of algorithms and algorithmic accountability in a practical setting. Next, algorithmic accountability was defined in a framework by researching both the concept of accountability and the meaning that the research field of algorithmic decision-making has given it. Based on these insights, the first subquestion was answered.

2.3.2. Cycle 2: deep dive into the V2G case study

The second cycle dived into a specific context: the Vehicle to Grid case study. The stakeholders of the project were interviewed, in order to get an understanding of the working of the system and of the barriers and opportunities to achieving algorithmic accountability. The findings from the interview were brought together in a causal loop diagram, in which a focus was chosen within the socio-technical system. The design fiction that was then created helped to imagine what might give users in the system a reason to ask for an account, and what might prevent them from doing so. Another result of this phase was a formulation of the problems that might arise as the V2G algorithms start interacting with their context.

2.3.3. Cycle 3: framing the design challenge

Once a focus was chosen to find a way to engage the user in the implementation of accountability, a contextmapping study was set up to understand the context of electric vehicle (EV) drivers, to find out what needs, concerns and dreams a design might 'hook onto'. A design challenge was set up based on the findings up to this stage and an ideation phase followed.

2.3.4. Cycle 4: solution finding

In the final phase, ideas were generated for the design challenge set in the third cycle. An interaction vision captured the essence of how the design should achieve its challenge. Design criteria were formulated as well as a set of assumptions that would scope the design challenge.
Then, a final design was developed that would incorporate the research and design findings throughout the cycles: the opportunities and barriers; the user's context; the possible problems that might arise. A brief literature research into Human-Computer Interaction design and accountability provided insight in the existing findings and questions on this topic. The final design was evaluated in an expert interview and by reviewing it theoretically based on the framework that was set up in cycle 1.

This research follows a case study approach, combining the findings from a literature review with the insights from the Vehicle to Grid case study to answer the research question. The four cycles that followed from this approach, form the main body of this report.

TO SUMMARISE...

3. Design excursion

4. Literature review

5. Algorithmic accountability

I

Exploring algorithmic accountability

With the stage set - the research topic and approach having been established - it is now time to turn towards the first cycle.Through a design excursion (chapter 3) and a literature review (chapter 4), the concept of algorithmic accountability is defined. The findings from both activities are brought together in chapter 5.

3. Design excursion

By journeying through a set of design excursions, an initial exploration of algorithmic governance and accountability was done (3.1, 3.2). As the aim of this research is to develop a design framework that structures the challenge of designing for accountability, these design studies were meant to get a better understanding of the problems a designer might face (3.3).

3.1. Method

Specifically, the design excursions revolved around ways in which people affected by algorithmic decision-making might be able to govern the system in some way. These excursions led through the design process of investigating the context and coming up with ideas, making the concepts discussed in the introduction more concrete. For each case, the following topics were discussed afterwards:

- The socio-technical context of the decision-making
- The opportunities and challenges to achieving accountability
- Implementing accountability by design

Since the first design study illustrates the findings more clearly, this design process will be elaborated in the next paragraph. A full description of both design cases can be found in Appendix A.

3.2. Design excursion: Smart lanterns in the market square

This first case was carried out by me alone, as an initial exploration into what designing for an artefact that includes algorithmic decision-making would mean. Designing for accountability was not explicitly mentioned in the design challenge but the system being fair and accountable was considered a criterium.

The problem statement for the case study was formulated as follows:

'How to design a system of lanterns that adjusts the lights when calamities are sensed on the market square to increase the social safety?'

This idea of 'smart lanterns' has been researched in various smart city projects, such as in Eindhoven (Smart Data City, n.d.). The effects of changing intensity and colour of street lanterns has been tested in research labs, but not yet in public settings. The supposed client for this case study was the municipality of Delft, and the target

figure 5 | a user scenario of the desired working of the smart lanterns

group was its citizens. To give the project a clear context, the city's market square was chosen as the environment in which the lanterns would be placed.

3.2.1. Method

In approaching the problem, I mainly used user scenarios (figure 5) to sketch out what interactions between the lanterns and users may be. This helped in understanding how the lanterns might provide a positive contribution to the city and how it might have a negative effect.

3.2.2. Results

The design process resulted in three concepts which will be briefly elaborated. The concept sketches are shown in figure 6.

1

In the first concept, situations are judged by volunteers. The details of a 'case' along with a photo are sent via an app. The volunteer judges to what extent the action that was undertaken was effective and fair. This way, a human is in-the-loop by taking over the learning component from the algorithm.

2

The second concept offers the possibility of a human-in-the-loop. On the market square itself, a display shows the measurements that the lanterns' sensors have made, the conclusions that were drawn from this, the action that was performed and the conclusions that were drawn on the effectiveness of the action. People can sit on a bench and review the various 'loops' and contest an unfair outcome.

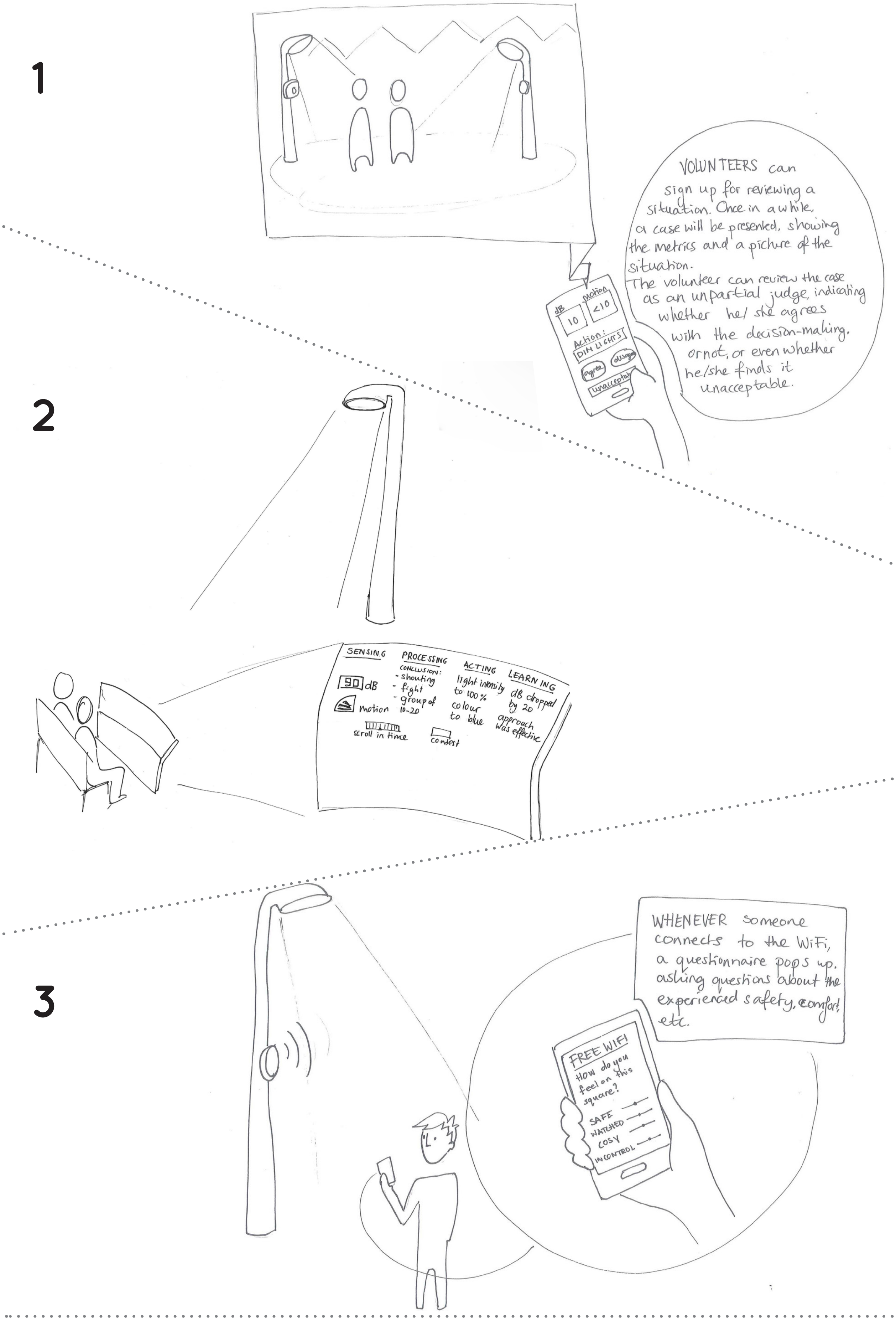


figure 6 | the three concepts sketches for the design excursion

The final concept provides a possibility for an affected citizen to hold the system accountable. Whenever people want to connect to the free WiFi that a lantern offers, questions are asked about the way the person experiences the environment and the possibility is offered to ask for the justification of a certain action that is performed by the lanterns.

3.3. Discussion & conclusion

After going through the design processes in the case studies, the three topics presented in the introduction of this chapter were considered.

The socio-technical context of algorithmic decision-making

In the first case study, many people might be involved in the algorithmic decision-making, either indirectly or directly, and either intentionally or unintentionally. The context is the city and its citizens, tourists and other visitors. The designs focused on creating touchpoints between those affected by the decision-making and the system, as they did not exist yet and the technology had a less obvious presence than in the second case study. An important stakeholder was also the municipality, and it would be important to understand the reasons for implementing these smart lanterns in the first place.

Opportunities and challenges to achieving accountability

In designing for accountability, it became clear in the first case study that it would first be necessary to anticipate what the system might have to account for. This is hard to predict of course and might be one of the challenges in designing for accountability. Working out user scenarios would then be an opportunity.

Another challenge that arises when reflecting on the concepts that were created in this case study, is the investment it asks from the user. Could this effort be minimised, or the motivation of the user be enhanced? The second case study also assumed users would be willing to invest ample time in planning the working of the system.

A challenge that was present in both cases was not only the complexity of algorithmic decision-making itself, but also of accountability. Designing with the aim of making the system accountable asked for a different mindset than designers might be accustomed to. There are many aspects to take into account: the user's needs, the collective goal that the algorithm is meant to fulfil, and how and when it might be necessary to have a human 'in-the-loop' of the automated decision-making.

Implementing accountability by design

Based on the explorations in these case studies, I would propose four considerations for implementing accountability by design. First of all, it's important to imagine – and preferably research – the context itself first, in order to anticipate what undesired situations might arise and how to prevent this. This requires some understanding of the needs and values of the user that might be obstructed by the algorithm.

Secondly, in the case studied the solutions asked a considerable time investment from the user, which might not be realistic. It would be desirable to design this in a way that the required intervention asks minimum resources. Thirdly, accountability might be achieved by providing the user with insight into the system, or with a way of influencing its working. And lastly, the touchpoints by which the user becomes aware of the algorithmic decision-making should be designed into the system in a natural and user-centred way.

A first stab at designing for accountability showed how complex this can be. The purpose of the design, the possible consequences of the algorithmic decision-making and the interactions with the user all need to be considered.

TO SUMMARISE...

4. Literature review

Having examined algorithmic accountability from a design perspective, I turned towards literature to define algorithmic accountability theoretically. A systematic literature review (4.1) was done to explore both the concept of accountability and of algorithmic accountability (4.2). The aim of the literature review was to come up with a framework by which algorithmic accountability can be investigated in context (4.3).

4.1. Method

To gather the relevant literature, the databases of Scopus and Web of Science were searched. The exact search method and selection of articles will be discussed separately for accountability and algorithmic accountability in the coming sections.

4.1.1. Accountability

First, a literature research was executed to gain an understanding of the general concept of accountability.

A time frame of 20 years was used here, because a concept evolves over time and old interpretations are less likely to still be relevant. Only when the word 'accountability' was mentioned in the title, the article was included, because accountability had to be the main focus of the article. The search query that was used was:

TITLE (("accountability" AND "concept"))

The abstracts of the resulting articles were screened for a focus on the discussion of the concept of accountability. This led to four articles that were studied.

4.1.2. Algorithmic accountability

Next, a more thorough literature review of the existing literature on algorithmic accountability was executed. No specific time frame was chosen here since the subject is very new and all contributions might be relevant. The search query that was used was:

TITLE-ABS-KEY ((("Algorithmic decision-making" OR "automated decision-making") AND ("accountability")) OR ("algorithmic accountability"))

Two articles were added that had been found during an exploratory search of the literature.

The abstracts were screened for: a discussion of algorithmic accountability in some way; offering either a definition or a way to achieve it. The remaining articles were read and judged on eligibility by reviewing the perspective from which algorithmic accountability was considered: articles that took a solely technical approach, focusing on the mathematical code, were excluded.

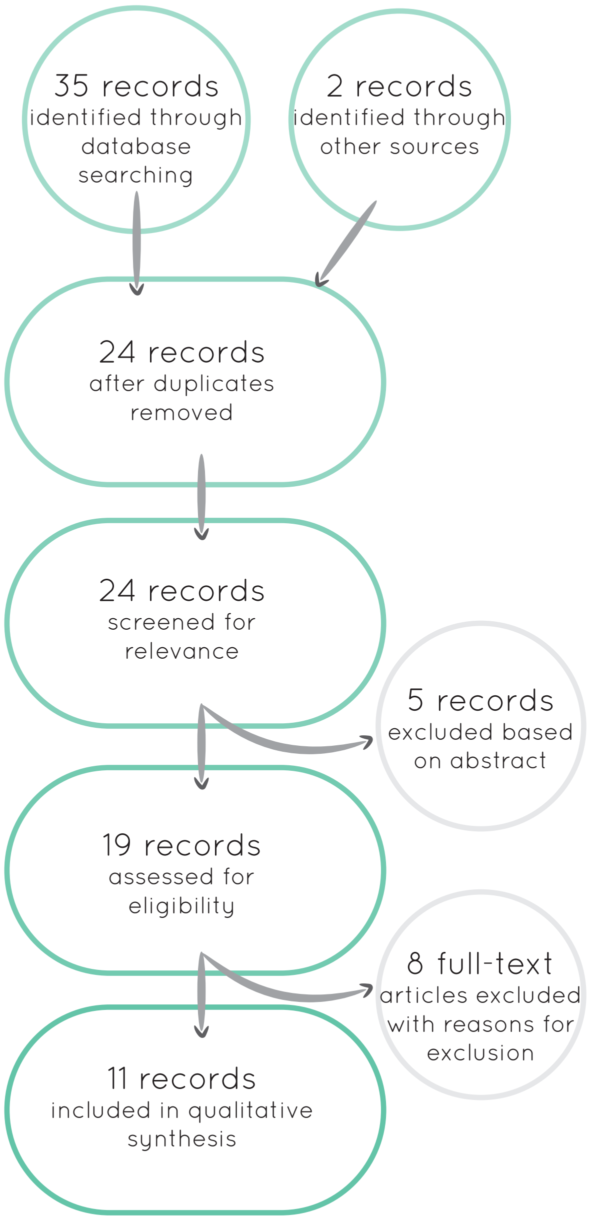


figure 7 | PRISMA diagram of the literature selection process

Moreover, some articles turned out not to mention algorithmic accountability at all, or only mentioned fleetingly. See figure 7 for a visual description of this selection process.

Besides these scientific articles, the book 'Weapons of Math Destruction' by Cathy O'Neil ((O'Neil, 2016)) was read as background to the subject of the risks of automated decision-making.

4.2. Results

4.2.1. The concept of accountability

To understand what algorithmic accountability exactly is, first I investigated the general concept of accountability.

The concept has been used in various domains, and in many ways, making it less defined and requiring clarification in each context (Mulgan, 2000)(Bovens, 2010), and perhaps even losing its core meaning (Lindberg, 2013). In response, these authors have tried to define the concept of accountability, so as to make it more valuable instead of making it an overarching concept that in the end says nothing at all. According to Mulgan (2000), the emphasis of accountability should be on "holding the powerful to account through political and legal channels of external scrutiny and sanctions." (p. 571).

Bovens (2006) defines accountability more generally, as a social relationship 'between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences.' (p.9). This definition of accountability clearly distinguishes different roles and actions that are a part of accountability

> Accountability:
> **" a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences "**
> Bovens (2006)

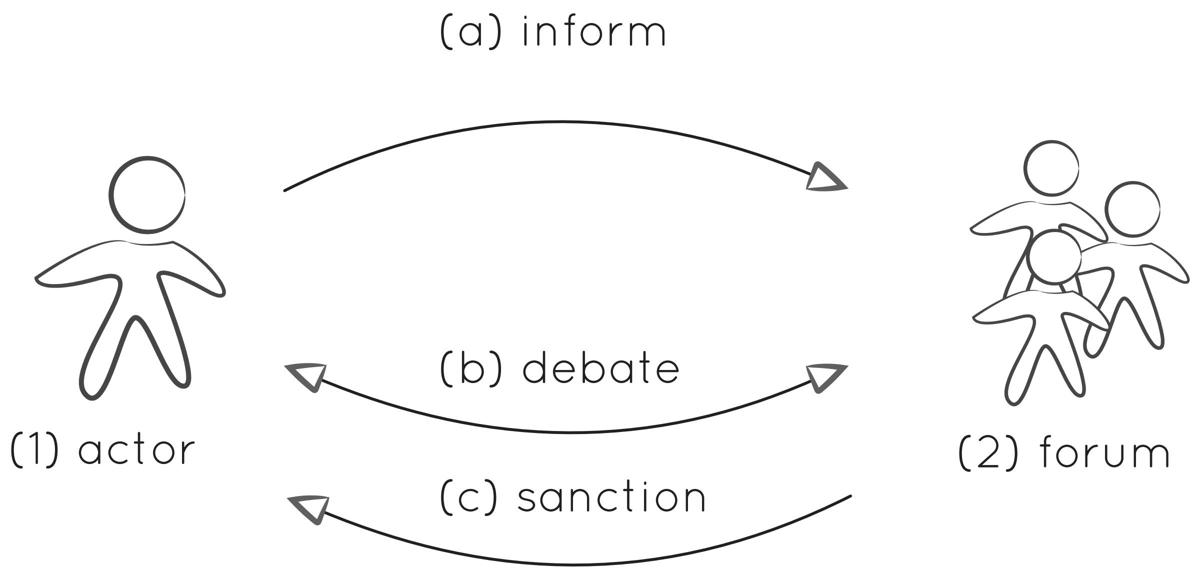


figure 8 | roles and actions in accountability

and therefore provides a suitable framework by which to analyse it. The roles are (1) an actor which performs certain actions and (2) a forum that can ask for an account of these actions. This can transpire in the following three steps (a) the actor informs the forum about his actions, (b) the forum interrogates the actor and can ask for further justification, (c) the forum passes judgement on the actor's conduct (figure 8)

This aligns with what Bovens later defines as 'accountability as a mechanism' and which he emphasises should be distinguished from 'accountability as a virtue', which is more like a set of standards for the evaluation of the behaviour of public actors (Bovens, 2010).

4.2.2. A framework for algorithmic accountability

Returning to the topic of algorithmic accountability, this can now be related to the framework presented in the previous section.

Algorithmic accountability is generally defined in terms of 'openness and transparency, at least theoretically enabling the algorithm, its authorship and consequences to be called to question by those subject to algorithmic decision-making processes' (Neyland, 2016). Diakopoulos is one of the first to draw attention to the importance of algorithmic accountability. He makes a call for it in the domain of journalism.

"In the case of the modern democratic state, citizens elect a government that provides social goods and exercises its power and control in a way that is moderated through norms and regulation. The government is legitimate only to the extent it is accountable to the citizenry. But algorithms are largely unregulated now, and they are indeed exercising power over individuals or policies in a way that in some cases (for example, hidden government watch lists) lacks any accountability whatsoever."

(Diakopoulos, 2016, p. 58.)

In other words, because the algorithms exert a certain control over citizens, they should also be accountable to society. Lepri, Oliver, et al. (2017) call this an information asymmetry between the government and big companies and the people on the other. Relating this to the framework presented in the previous section, the company and its algorithms would be the actor. The forum, then, might be the general public, the people affected by the algorithmic decision-making or a

“ algorithms are largely unregulated now, and they are indeed exercising power over individuals or policies in a way that in some cases lacks any accountability whatsoever. ”
Diakopoulos (2016)

third party that represents these affected people. Binns' explanation of algorithmic accountability comes closes to this, and it is this definition that I will follow:

“In the context of algorithmic decision-making, an accountable decision-maker must provide its decision-subjects with reasons and explanations for the design and operation of its automated decision-making system. The decision-subject can then judge whether this justification is adequate, and if not, the decision-maker may face some kind of sanction or be forced to retract or revise certain decisions.”

(Binns, 2017, p.2)

Concerning the components of accountability, five principles for developing accountable algorithms have been proposed by a community of researchers and practitioners concerned with Fairness, Accountability and Transparency in Machine Learning (FAT/ML). These are: responsibility, explainability, auditability, fairness and accuracy (Diakopoulos et al., 2017). In the following section I will investigate what these principles entail and what other authors have said on the topic, after which I will compare them with the three steps that Bovens describes.

Responsibility, means having an internal role for the person who is responsible when an algorithmic decision has undesired effects, and communicating clearly how to address this person or the company he/she represents. This should prevent responsible parties from hiding behind the algorithm. This is also described by Lepri, Oliver, et al. (2017), who say that algorithmic accountability should lead to clarity about who is responsible for the algorithmic decisions that are made, and by Broeders et al. (2017), who state that the data processor remains responsible for the quality of data used and the method with which it is processed.

Explainability as described by Diakopoulos et al. (2017) is ensuring that the decisions that are made and the data behind it can be explained in a way that is intelligible for end-users and other stakeholders. Providing explanations have been touched upon by many authors in the context of transparency. While the opaque nature of algorithmic decision-making can complicate this, post-hoc interpretations of algorithmic models might provide sufficient explanation without having to elaborate the exact process (Lepri et al., 2017). Zouave & Marquenie (2017) describe explanations of the aims, reasoning and processes behind algorithm as an important way of achieving transparency and for offering individuals an option to challenge the outcomes.

Auditability signifies the opportunity for third parties to understand and evaluate the behaviour of the algorithm by offering sufficient insight into the input data or outcomes of the algorithm. Literature offers various examples of how this auditing could play out. Broeders et al. (2017) imagine a specialised oversight authority that could monitor the way an algorithmic system is developed. Spielkamp (2017) elaborates the non-profit initiative 'AlgorithmWatch' that he has set up to audit algorithmic decision-making processes that impact human action and society. And in an experiment to create an accountable security camera Neyland (2016) sets up an ethics board that reviews his development process.

Accuracy requires identifying possible faulty or uncertain areas of error in the algorithm and its data set to assess possible worst case scenarios. *Fairness* is defined as preventing discriminatory or unjust impacts of the algorithm. These principles are similar to the goals of fairness and efficiency as described by Zarsky (2016), which were discussed in the introduction of this report. I consider this a further clarification of what a creator is responsible for, but not as a part of the accountability mechanism.

Now comparing these principles to the framework discussed in the previous section,

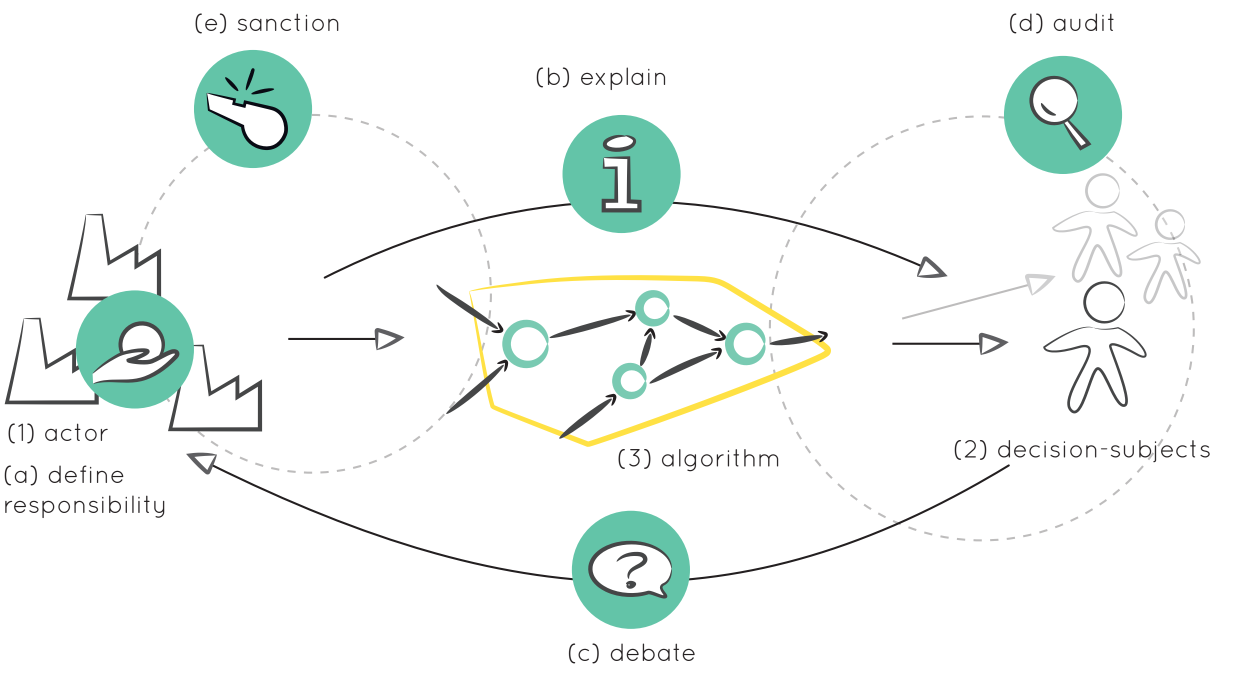


figure 9 | combining Diakopoulos and Bovens' views in a theoretical framework.

To achieve algorithmic accountability, the actors (1), need to clearly define their responsibility (a) for the algorith (3), which exercises power over the decision-subjects/users of the system (2). An explanation (b) for the algorithm's working should be provided, in response to which the decision-subject should have an opportunity to debate (c). Auditing (d) the outcomes for all decision- subjects should be possible as well. If the algorithms decision can't be adequately justified, it should be possible to impose sanctions on the actor(s).

the principle of 'explainability' can be viewed as similar to Bovens' action of informing the forum. The stage in which the forum can interrogate the responsible actor and impose sanctions, however, has not been included in the principles proposed by Diakopoulos et al. (2017). This could be due to the fact that the principles describe the desired behaviour of actors, and in that sense aim for accountability as a virtue (Bovens, 2010).

However, both Ananny & Crawford (2016) and Neyland (2016) underline the importance of viewing algorithmic accountability in the context of the whole socio-technical system and examining the power of an algorithm in the set of relations that 'bring an algorithm into being'. Therefore, the framework shown in figure 10 was developed, which combines the principles of Diakopoulos et al. (2017) and the stages of Bovens' mechanism of accountability.

4.3. Conclusion

Based on a literature review on the concepts of accountability and algorithmic accountability, I defined five aspects that I expect to be relevant in achieving algorithmic accountability in a socio-technical system. These are as follows:

- Responsibility for the effects of an algorithmic system, and a clarity over who has the authority to step in when the algorithm has unwanted individual or societal effects.

- Explainability towards the people affected by algorithmic decision-making.

- Debatability of the information that is provided about the system or the possibility to interrogate the responsible party about outcomes of the system.

- Auditability of the algorithm and the input data by a third party.

- Sanctionability in case the account that is given for algorithmic decisions doesn't suffice.

Ffigure 9 visualises the relations betweens these aspects in a theoretical framework.

By combining Bovens' definition of accountability with the principles of algorithmic accountability as defined by Diakopoulos et al., a new framework arised that theoretically captures how algrithmic accountability can be implemented in a socio-technical context. The relevance of this framework for my research will be discussed in the final chapter of this cycle.

TO SUMMARISE...

5. Algorithmic accountability

cycle I conclusion

In this first cycle, I investigated what algorithmic accountability is, both by designing for it and by reviewing the current literature on the topic. I will conclude the cycle by answering the first subquestion.

The first subquestion was formulated as:

SQ1. What is algorithmic accountability?

Following Bovens conceptualisation of accountability and combining this with the insights I gained in the literature review, I define algorithmic accountability as a relationship between the decision-maker and the decision-subjects, in which the creator of algorithmic decision-making system provides reasons and explanations for its working, and the decision-subject can debate this justification. If it is inadequate, a sanction might be imposed on the decision-maker or he might be forced to change the functioning of the system. This is based on the definition of Binns (2017).

The literature review provided five aspects If we consider the complete socio-technical context, there are multiple aspects to this accountability: Responsibility, explainability, debatability, auditability and sanctionability. The theoretical definitions have been provided in subchapter 4.3. Reflecting on the design case in chapter 3 once more, provides some context for these aspects.

The design excursion and the insights that were gained from it, provide a context to understand these aspects, and how they might lead to algorithmic accountability. First of all, the user will be likely to have limited time available and probably won't understand the working of the whole system. So, regarding the aspect of explainability, this would need to be tailored to the user group and should have a clear focus. Second, concerning auditability, the design study illustrates that not simply everything can be audited. Careful considerations would need to be made about what might go wrong and what data would be required to uncover this. Third, the need for a touchpoint which was found in the design case can be related to debatability. For this it would be relevant to understand at what time a user might require information and what might be a suitable channel for asking this.

The last two aspects, responsibility and sanctionability, had not been considered yet in the design study. Apparently, they were less inherent to me as a designer. However, it did become clear that the needs and values of those affected by the algorithmic decision-making need to be clear, so these can be protected through accountability. This might be seen as the responsibility that the decision-maker should take.

In conclusion, what algorithmic accountability is exactly, is highly dependent on the context. The theoretical framework that was set up in chapter 4 led to an understanding of its various aspects. Armed with the insights from this first exploration, I dived into the case study of the Vehicle to Grid system.

TO SUMMARISE...

Photo by Tom Rumble on Unsplash

6. Stakeholder interviews

7. Causal loop diagram

8. Design fiction

9. Critical areas

II

Deep dive into V2G

With the concept of algorithmic accountability having been defined in a framework in the previous, this cycle turns towards the Vehicle-to-Grid project. This pilot, coordinated by grid manager Alliander, was chosen as a case study for examining algorithmic accountability in a specific context. The stakeholders of the project were interviewed (chapter 6), and the insights were brought together in a causal loop diagram (chapter 7). Next, a design fiction was created to examine possible effects of automation (chapter 8) and in a conclusion critical areas for establishing algorithmic accountability in the V2G case study were identified (chapter 9).

6. Stakeholder interviews

In this chapter, the Vehicle to Grid case will be further introduced (6.1) and the method for interviewing the project's stakeholders will be presented (6.2). These interviews provided an understanding of how and why the stakeholders work together, and of what the barriers and opportunities might be to achieving algorithmic accountability (6.3, 6.4).

6.1. The V2G project

The source for the information concerning the V2G project is the interview with the project manager of the project, unless indicated otherwise. A full transcript of this interview can be found in Appendix B.

The Vehicle to Grid project was initiated by the Dutch energy grid manager Alliander in 2014. To Alliander, V2G is an innovative concept that could help in the prevention of congestion in the energy grid. In the coming years, the energy load on the grid is expected to continue rising, especially as the electric vehicle becomes more dominant. The project was therefore set up to explore the potential of this concept.

The V2G project is one of multiple projects in the Cityzen research programme. This programme is built up out of projects that investigate in practice how a city might operate entirely on clean energy, investigating the technical as well as the economic and social barriers (City-zen, n.d.). In case of the V2G project, a technical challenge is the bi-directional charging of the car, while a social question is whether people will be prepared to hand in some flexibility regarding driving. (Amsterdam SmartCity, 2018).

6.1.1. Project stakeholders

In February 2018, a pilot with 4 participating EV drivers was kicked off. Alliander had joined forces with Enervalis and Newmotion to realise this pilot. Newmotion delivers and maintains the hardware, the actual V2G charging units and is in charge of invoicing the EV drivers for their energy usage. Enervalis manages the data flow and designs the intelligence of the system. These stakeholders and their task are presented visually in figure 11 on page 48.

6.1.2. The working of the V2G system

The working of the system within the pilot will be illustrated (figure 10) and elaborated here step by step.

1) After arrival, the user connects his car to the V2G

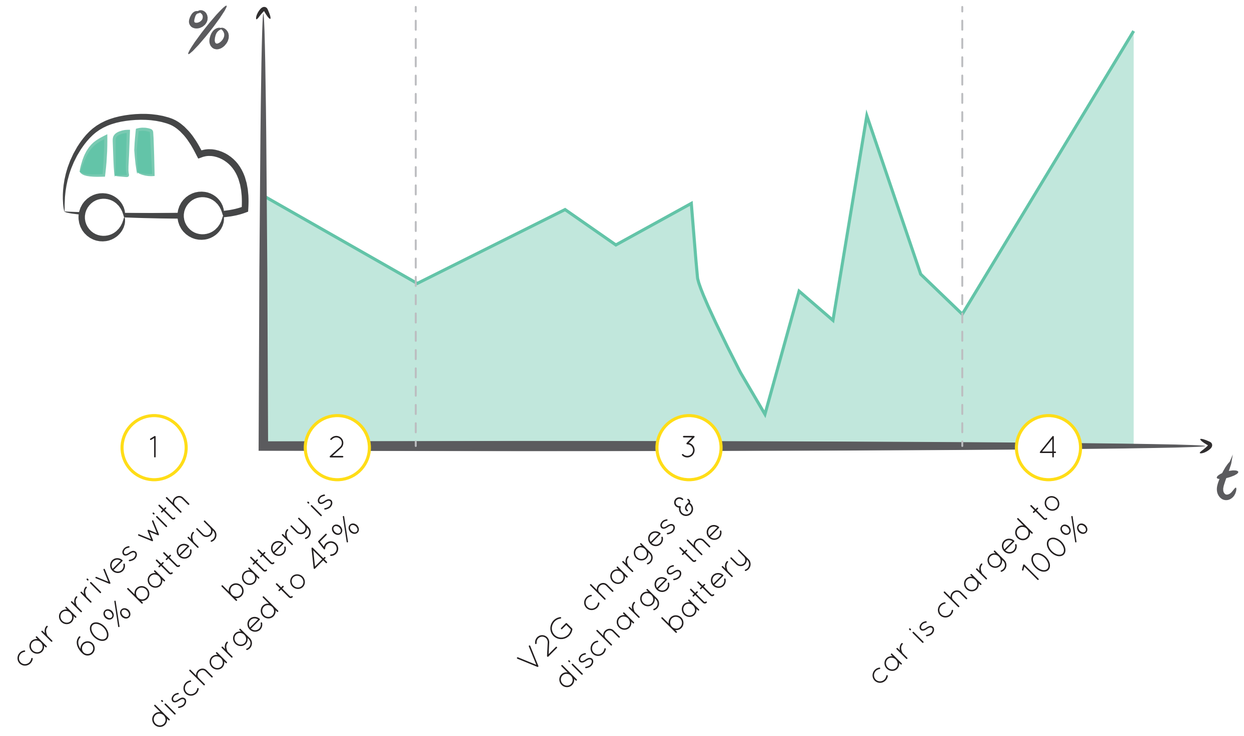


figure 10 | charging and discharging the car within V2G

unit that has been installed either at his home or in a public area. Via an app on his phone, he can indicate his departure time.

2) The system starts by either charging or discharging the car until the battery is 45% charged.

3) Then, based on the energy balance in the grid, the car will either be charged to store a surplus of energy, or discharged to compensate for a shortage of energy in the area. A network of algorithms designed by Enervalis is optimalised to achieve both a minimal congestion of the grid and a maximum profit by selling energy when the price is high and buying it when it is low. Optimisation is achieved via reinforcement learning, see box 3.

4) An hour before the deadline, the V2G unit will stop using the car as a battery and only charge it, so the user will find his car fully charged. In case the user suddenly decides to leave earlier, he can press an emergency button, and the car can be charged

box 3 | Reinforcement learning

Reinforcement learning, like most machine learning, uses artificial neural networks (ANN). In other words, they create intelligent functions by building up networks that are similar to those in a human brain. This enables the algorithms to respond to a completely new situation. Usually, a machine learning algorithm will try to find 'the correct' action that will give him his reward. But a reinforcement learning algorithm can take several actions one after the other to maximise a certain reward, and it can adapt to changing circumstances. (Karanasiou & Pinotsis, 2017)
An example of what (deep) reinforcement learning can do, is the triumph of Google DeepMind's AlphaGo. In March 2016 this computer programme defeated
Lee Sedol, the second highest-ranked

continues on next page >>

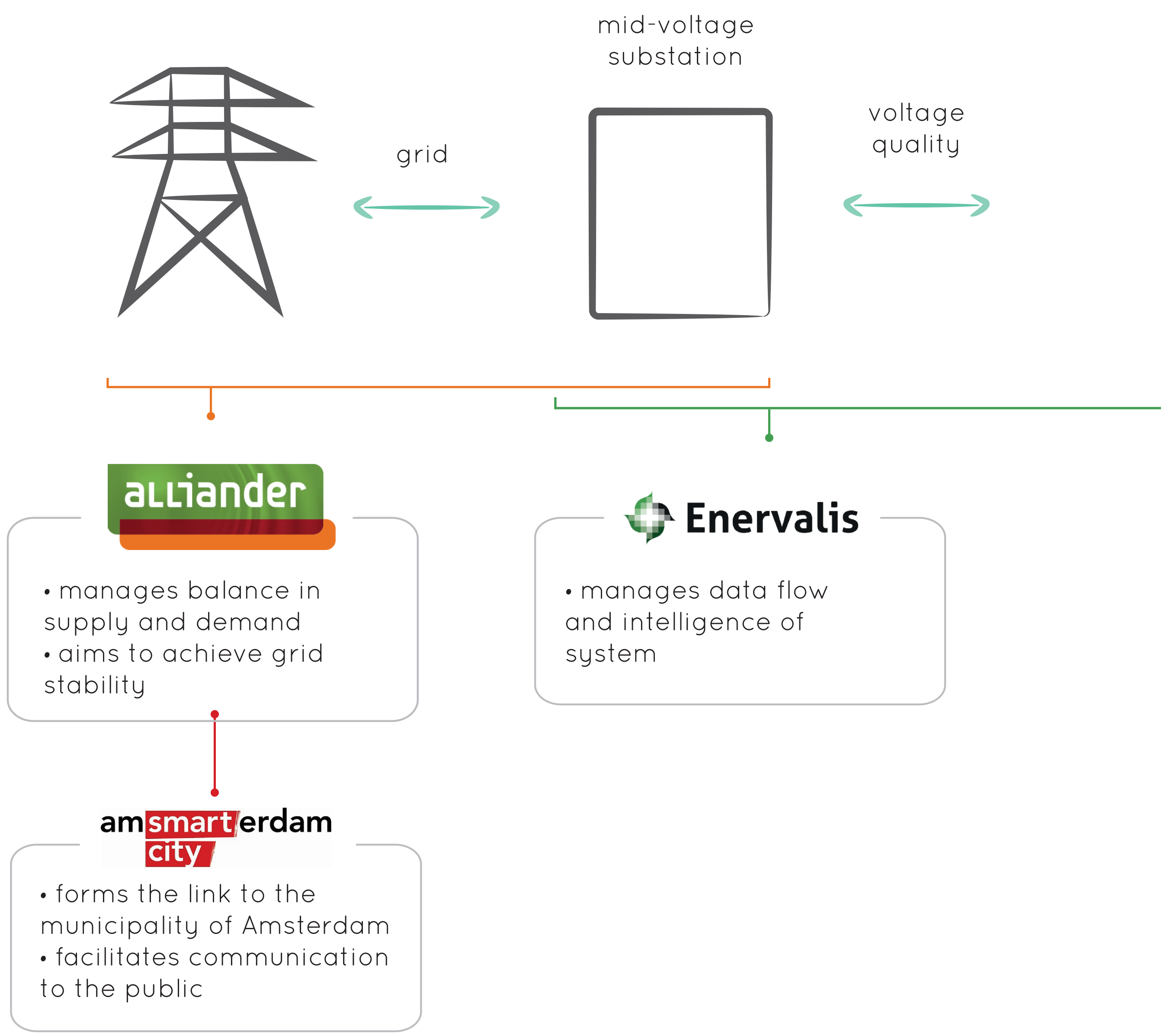
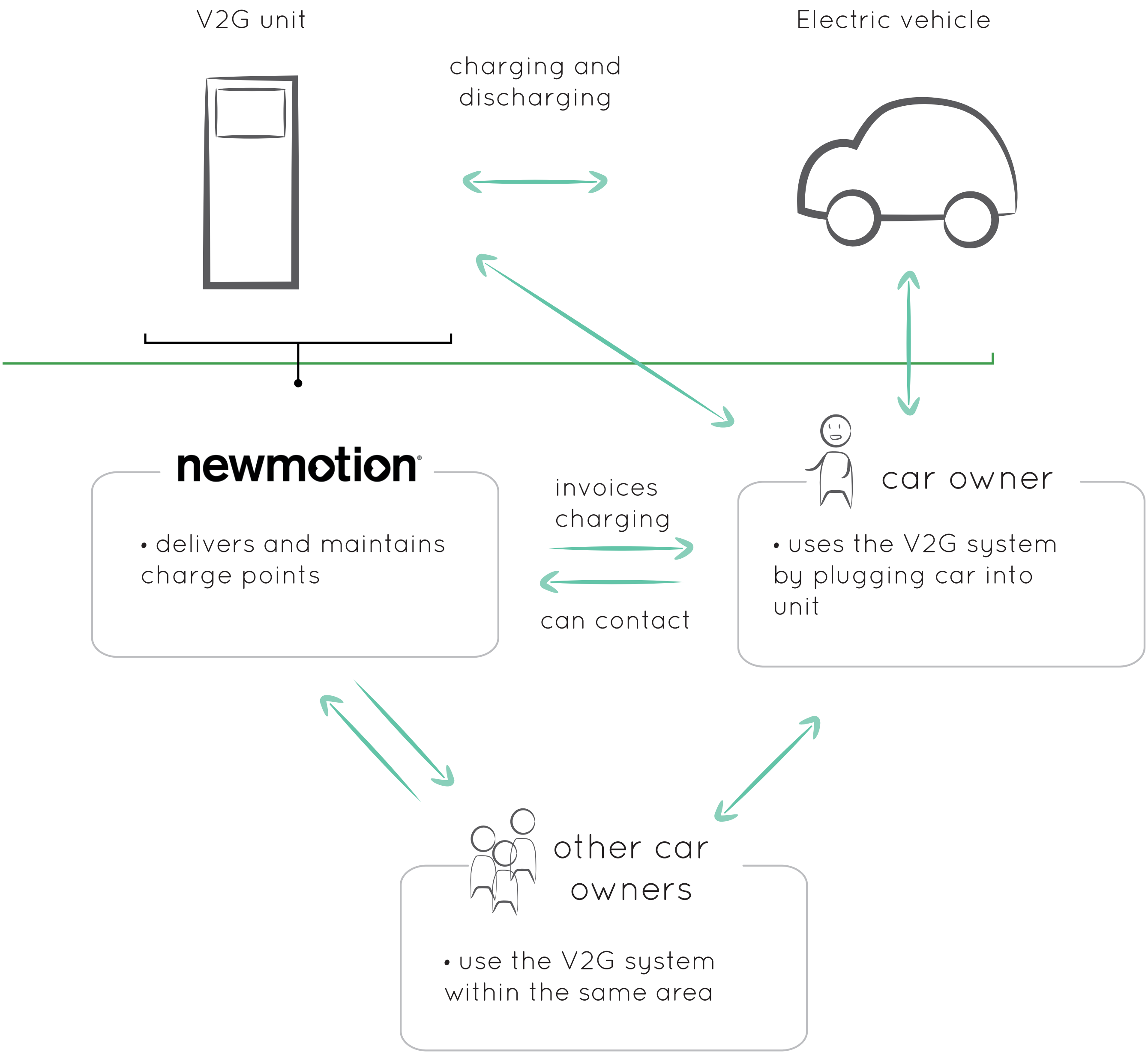


figure 11 | stakeholders and their tasks and interactions in the system of the V2G pilot

professional Go player, in a best of five Go competition. Compared to chess, Go is regarded as more intuitive, and it has a higher complexity because of the many moves that can be made at one time. The computer programme was trained by playing against another computer and being rewarded for winning the game. Based on this, it developed ANN that would lead to this reward (Burger, 2016).

within half an hour.

Once the system starts recognising consumption patterns, it can start playing around with the final charge amount. For instance, a user might only travel short distances, for which a battery level of 40% would suffice. Or an EV driver might always arrive half an hour late, meaning the system could keep the car connected to the grid for 30 more minutes. One of the pilot's research objectives is to find out how users will respond to this, and how comfortable they will be with the system making these adaptations.



6.2. Method

In order to investigate the different aspects of algorithmic accountability within the V2G case, the stakeholders involved in the project were interviewed. Another aim of these interviews was to get insight into the concerns of the stakeholders.

In semi-structured interviews, the possibilities and limitations for accountability were discussed with the different stakeholders, by asking questions about each of the five aspects of accountability that were presented in the framework in chapter 4. Also, the interviewees were asked for their company's motive for developing the V2G system. The semi-structured quality of the interviews allowed some flexibility in the order in which the questions are asked and the amount of attention that is given to each topic, while also ensuring that all relevant topics were discussed (Bryman, 2012).

The interviews were held with a representative of each of the five companies involved in the V2G project. The interview with the project manager (the representative from Alliander) was held in person, the rest of the interviews were conducted by phone. The interview protocol can be found in Appendix C.

The interviews were transcribed and then analysed in NVivo ("NVivo qualitative data analysis Software," 2017), by coding the interviews according to the five aspects of accountability: explainability, debatability, auditability, responsibility and sanctionability. Quotes related to these topics were then searched for opportunities and barriers. Statements related to the stakeholders' motives and concern in joining the V2G project were also coded. The code tree can be found in Appendix C as well.

6.3. Results

This resulted in a better understanding of the stakeholders' motives and in a list of opportunities and barriers for each of the five aspects. I'll first discuss the former for each company, before presenting the latter in a table.

6.3.1. Stakeholder motives

For Alliander, the V2G system is a possible way of managing the grid in the most cost efficient way. The electric grid that is currently installed is very likely to become overburdened as the energy use continues to grow. Especially electric transport, solar panels and heat pumps all have a big impact. Smart solutions like V2G can minimalise the investments that are needed to facilitate this increasing load.

At the moment, is uncertain whether they would be able to implement V2G on a larger scale; As a public organisation, they have to account for their investments to the government.

Newmotion is a commercial company, that provides a charging service for EVs. They act as a link between the EV drivers and the energy companies that supply the electricity. Providing the service in the V2G system therefore fits in their portfolio and is part of the strategy to stay ahead of the market. In the words of the interviewee, Newmotion tries to find 'a balans between sustainability ideals and a good business plan'.

Enervalis is a commercial software company, that strives to facilitate 100% renewable energy. The mismatch in supply and demand of renewable energy sources makes an energy grid that relies on only green energy unaffordable. Their core business is to come up with software solutions that predict the energy market, collect data on the working of heat pumps and electrical vehicles and facilitate a balance in the supply and demand of energy. Enervalis' interest in V2G comes from the possibility of combining the fluctuation in the energy market with the prevention of congestion. This could lead to a more solid business model.

Amsterdam Smart City plays a role in communicating the research results of the V2G pilot, but isn't directly involved in building the system.

6.3.2. Barriers and opportunities

The barriers and opportunities have been summarised in table 1.

6.4. Discussion

This is not an exhaustive list of all the barriers and opportunities to achieving algorithmic accountability, but it does give an overview of the main possibilities and limitations that should be considered. The stakeholder motives for setting up the V2G system are relevant to understand how the algorithmic system will be set up and optimalised.

barriers	opportunities
Explainability	
• the algorithmic system is complex • financial cost of explanatory sofware is high • Enervalis wants to protect intellectual property • user doesn't want information overload	• adding explanatory software *is* possible • providing information via the app interface • sending push notifications as a warning when car isn't charged
Debatability	
• terms of agreement are often unflexible • no direct connection between user and Enervalis • Enervalis wants to protect intellectual property • privacy of other users	• customer service of Newmotion • information sessions for users • Amsterdam Smart City platform as a place for debate • social connections between users
Responsibility	
• user is highly dependent on system • inscrutable terms of agreement leads to a user not knowing what he signs up for • lack of control over Enervalis	• certification of ethical principles • clear division in roles of parties • offering alternative services (competition) • single contact point for users • one party with ultimate responsibility
Auditability	
• the costs of auditing could be high • protection of intellectual property • privacy of other users • Enervalis has control over transparency of the system	• external authority for inspection • contractor can force transparency • providing an information portal for installers can enable auditing
Sanctionability	
• undefined who should sanction	• contractor can impose a sanctioning over the contracted company

table 1 | barriers and opportunities to achieving algorithmic accountability within V2G

Alliander, Newmotion and Enervalis have together built the V2G system, which has the combined goal of preventing congestion in the grid and playing the energy market. No conclusions were drawn from the interviews yet, but the barriers and opportunities that were found were used in the construction of a causal loop diagram, which will be presented in the next chapter.

TO SUMMARISE...

7. Causal loop diagram

To investigate how the barriers and opportunities found in the interviews relate to each other and to the five aspects of algorithmic accountability, a causal loop diagram was constructed based on the findings (7.1). This chapter presents the diagram itself (7.2) and the critical node that was chosen to focus on within the case study (7.3, 7.4).

7.1. Method

The causal loop diagram is a tool used in system dynamics methodology. It is a representation of a feedback structure, communicating the principal causal mechanisms in a system. The causal relationship of two variables can be either positive or negative:

- an increase of one variable leads to the increase of another (↑ causes ↑). This is a positive relationship indicated with an arrow with a '+' sign.
- an increase of one variable leads to a decrease of the other (↑ causes ↓). This is a negative relationship indicated with an arrow with a '-' sign.

When two variables have a causal effect on each other in both directions, this can result in a reinforcing loop or a balancing loop.

- when the relations are all positive or have an even amount of negative relations, the loop reinforces itself. This reinforcing loop is indicated with an 'R'
- when the relations have an odd amount of negative relations, the loop balances itself. This balancing loop is indicated with a 'B'.

(Bala, Arshad, & Noh, 2017)

7.2. Results

The resulting visualisation is shown in figure 12 on page 54. Different markings have been used to indicate whether the insight came from an interview, from literature, both or was based on own interpretation.

7.3. Discussion

In the causal loop diagram, three 'critical nodes' were discovered: a cluster of elements in the system that captures the complexity of the system but also provides a focus on a more specific problem (van der Sanden & de Vries, 2016).

The first critical node is the 'inscrutability of the terms of agreement' and is related to the user dependency on the system, to taking responsibility for harm and to the responsiveness to comments. This is the contract between the user and the responsible parties and therefore could provide an opportunity to explain the working of the system. But similar to how privacy agreements are usually not read because they are too dense and extensive, it is expected that users of the V2G system might accept terms without reviewing them, because they are too complex to understand for a layman. Even if they review them and disagree, they will most likely have no choice but to accept because of their dependence on the service.

By improving this and making the terms more intelligible and open for discussion, the whole system might become more accountable.

The second critical node is 'the interest of the user', which impacts the urgency of implementing explanations in the system. As it is now, the perceived necessity of information about the algorithm's working is very low. But by shutting the user out like that, the algorithmic decision-making will remain a mystery for him and it will be harder to hold the responsible party accountable for its actions. A more active engagement of the user might lead to more debate with the creators of the V2G system.

The third critical node is the 'certification of ethical principles'. It is related to the parties taking responsibility for the working of the algorithm and it might be related to the aspects of explaining, debating, auditing and sanctioning as well depending on the content of the principles. By having a clearer set of principles to which all parties consent, it would be easier to hold them accountable for a specific action that opposes this principle.

The causal loop diagram is of course not purely objective nor is it complete. To increase the reliability of the diagram, the sources of the elements have been indicated. Linking the elements was a more intuitive process and so another person might come to a slightly different conclusion. For now, however it provides a reasoned starting point for further exploration in a design process.

7.4. Conclusion

The second critical node of 'the interest of the user' was chosen to focus on for the next steps in the process. The literature review of algorithmic accountability (chapter 4) showed that many solutions are currently being sought in the design of the algorithm itself or in the rules concerning this design. The third node, focused on clear ethical principles, would be similar to this and would lead to a solution that works on the boundaries of the system instead of playing out inside of it. The aim of this research is to design algorithmic accountability *into* the working of the system. Of the three critical nodes, the second node offers the best opportunity for this, by providing a user-centred perspective to algorithmic accountability.

Of the three critical nodes that were found in the causal loop diagram, the one revolving around the interest or engagement of the user in the system was chosen to pursue. This node offered the most opportunity of designing algorithmic accountability into the system.

TO SUMMARISE...

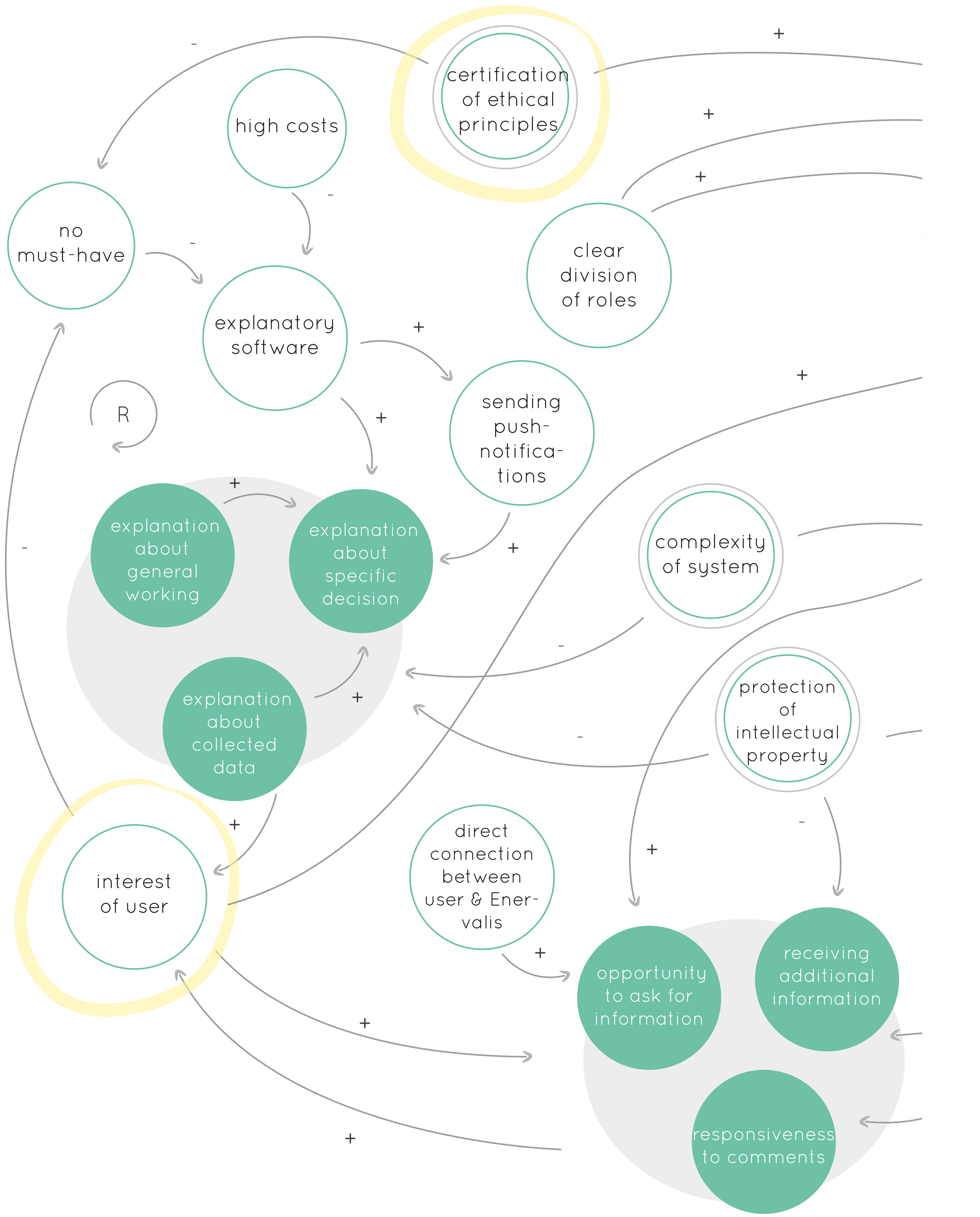


figure 12 | causal loop diagram showing the relations between opportunities and barriers to achieving algorithmic accountability in the V2G system

user dependency on system
availability of alternatives
taking responsibility for harm
clear point of contact
R
effective sanctioning method available
control over software company
inscrutability of terms of agreement
opportunity to sanction
ultimate responsibility of non-soft-ware company
B
power of contractor over software company
opportunity to audit
accessibility to 3rd party
high costs of audit
portal for installers
privacy of users
interview insight
literature insight
literature & interview

8. Design fiction

One of the findings of the design excursions described in chapter 3, was the importance of imagining how algorithmic decision-making might lead to undesired effects. Creating scenarios can facilitate this exploration. In this case, a design fiction (8.1) was developed to illustrate a possible future. This chapter describes how the fiction was developed (8.2), and what insights were gained in doing so (8.3, 8.4)

8.1. Narrating future technology

Design fiction, a term first coined by Bruce Sterling, has been defined by him as “the deliberate use of diegetic prototypes to suspend disbelief about change” (Bosch, 2012). ‘Diegetic’, a derivative from the Greek word of diēgeisthai ‘to describe, narrate’, is a crucial term in this description. It means the prototype should tell people a consistent story that explores the working of a future technology or service. A design fiction can serve as an inspiration and motivation for design, by exploring possible design requirements within a fictional scenario, and serve as a tool for communicating innovations to researchers or the general public. (Tanenbaum, 2014)

In this case, a slightly dystopian design fiction video was used to tell the story of a V2G unit that starts making increasingly complex decisions on more and more data, causing the user to lose control over it and be unable to understand the outcomes. Working out this story helped in examining, understanding and communicating the problem that might develop. In the workshop with EV drivers presented in chapter 9, the video was shown to start a discussion about this future.

8.2. Method

The video was developed in two iterations (figure 13). The first iteration was more experimental, with two actors improvising and acting out various scenes. To grasp the feeling and the problem, a metaphor was used in which the car driver became a customer of a bakery, and the charging station became a baker that refused to sell bread to the customer. The different scenes were edited together.

In the second iteration, the scenarios were first developed on paper and divided in three phases: a scene that shows the benefits of the smart V2G charging station, a scene that shows a more uncomfortable side of this unpredictable and all-knowing charge point, and a scene where the car driver has lost all control.

figure 13 | iterations of the design fiction video

8.3. Results

Since the first iteration was based on improvisation, this also led to some surprising ideas about how the V2G unit might behave if it was making increasingly automated decisions, based on an increasing amount of data. This led to the V2G unit anticipating traffic, warning the driver to drive safely and locking the car in place when the driver kept 'bothering' the charging point.

The final result was a short video, which can be viewed by following this link: https://tinyurl.com/V2Gdesignfic

8.4. Discussion & conclusion

The video 'brought to life' the possible future in which the V2G system is unaccountable and is automated to the point where its decisions are inexplicable. Developing the scenario meant having to imagine exactly what might happen and how, which led to a better understanding of the system and new ideas of how the V2G unit might act. It was helpful that the actor playing the V2G unit had a background in Computer Science and could play out scenarios based on this knowledge.

Only one scenario was worked out here, and it might have been even more valuable to work out several options. Nonetheless, the design fiction showed how the automation might slowly increase until suddenly there is no control anymore. It seems likely that only at the point where this has negative consequences, a user will become interested and react. This raises the question whether it would be possible to engage the user earlier on .

Improvising the future scenarios for the working of the V2G system led to new ideas about what the algorithmic decision-making might lead to. These ideas were captured in a design fiction, which was later shown to participants of a contextmapping workshop (chapter 10).

TO SUMMARISE...

9. Critical areas

cycle II conclusion

Throughout the cycle presented in this part, I researched the Vehicle to Grid case study, interviewed stakeholders and imagined how algorithmic decision-making could develop. The barriers and opportunities to establishing algorithmic accountability were plotted in a causal loop diagram to come to a focus on the lack of interest that a car driver is expected to have in the exact working of the V2G system.

In this reflection on the second cycle, conclusions will be drawn based on the insights from the research and design activities. With the findings from this cycle, SQ 2 and SQ3 can be answered.

SQ2. What is the socio-technical context of the Vehicle to Grid case study?
The V2G project is a collaboration between three parties: grid manager Alliander, aggregator Newmotion and software company Enervalis. Newmotion is the connection to the car driver; they deliver the service of V2G to the user and invoice the car driver for charging his car battery.

The algorithms that direct the V2G charging are based on reinforcement learning, and are constructed to optimise to maximise profit by buying and selling energy, and to minimise congestion in the grid. This optimisation is to be achieved by predicting energy usage in the neighbourhood and departure times of EV drivers.

SQ3. What are barriers and opportunities to achieving algorithmic accountability in the V2G algorithmic system?
By interviewing the stakeholders, possible barriers and opportunities to achieving algorithmic accountability were uncovered. Analysing these barriers and opportunities in a causal loop diagram led to the identification of three critical nodes. The interest of the user was chosen as the focus for the next phase.

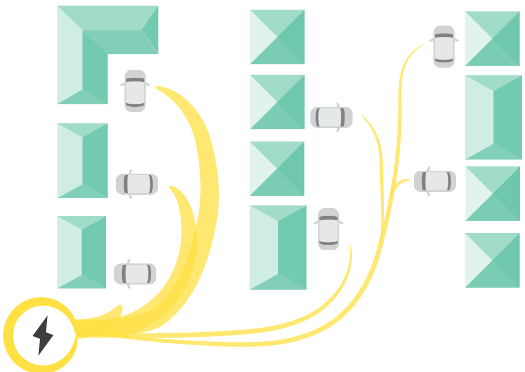
As was discovered in the design excursions of the previous cycle, it is important to understand what the undesired effects of the algorithmic decision-making might be. This was explored on an individual scale in the design fiction that was created, but apart from that two other problem-scenarios are also imaginable. All three problems are listed in box 4.

While all three problems are relevant, the first and second are more related to the challenges of algorithmic decision-making and therefore more emphasis will be placed on them to scope the challenge.

box 4 | critical problems in the V2G system

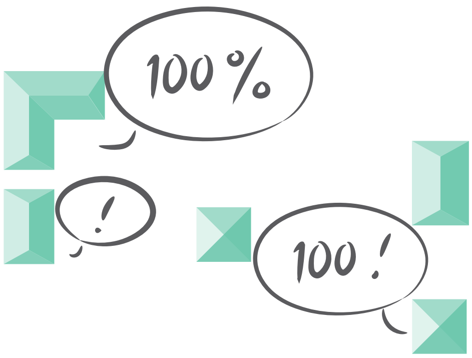
individual

The first problem was visualised in the design fiction. The EV driver might arrive at his car to find it insufficiently charged. He will have no idea of how this happened and what to do next if the V2G unit doesn't offer any explanation.



scaled

The second problem is less predictable and also less visible to an individual. Over time, as the algorithmic system develops some drivers might find their car uncharged more often than others. This could be because there is more energy 'reserved' for streets in which the energy consumption is generally high. Or they might be less predictable drivers. This problem was derived from the book by O'Neil (2016).



collective

Third, a social problem could develop if users can keep on indicating the battery level they desire. If there is no trust in the system, all drivers might keep on asking for a fully charged battery an hour before they actually leave, just so they have a safety margin.

The second and third subquestion have now been adressed. Moroever, through choosing a critical node and identifying the challenges that algorithmic accountability should adress within the V2G system, the next cycle can be entered with a clearer focus within the complex system of V2G.

TO SUMMARISE...

10. Contextmapping

11. Dystopia/utopia

12. Design challenge

III
Framing the design challenge

With the choice to focus on the interest and engagement of the user, this cycle set out to better understand the perspective of the EV driver. A workshop was organised with three EV drivers (chapter 10), and a desired and undesired scenario were sketched out (chapter 11). As a conclusion to this cycle, a design challenge was set that should build towards the positive future vision (chapter 12).

10. Contextmapping

With the user-centred approach revolving around the interest of the user chosen in the previous part, an investigation into the perspective of the EV driver was desired. To this end, a contextmapping workshop was organised (10.1). The approach to this workshop is presented in 10.2. The session gave various insights into the context of EV drivers and based on the results (10.3), relevant user needs and values were identified (10.4, 10.5)

10.1. Mapping the user context

The generative method of contextmapping was used to discover more about the context of EV drivers and to find needs and dreams that a design should 'hook on to'. In this technique, participants are guided towards expressing deeper levels of knowledge about their interactions and experiences as they make designerly artefacts. These might be drawings, collages or models. As the participants explain what they have made, they share rich stories about their current experiences, their fears, their dreams (Visser, Stappers, van der Lugt, & Sanders, 2005).

Within the experience domain (figure 14), the path of expression is followed: participants reflect on their present and past experiences to then enable them to imagine future experiences. In this case, participants first discussed their current electric car and their experiences in driving it, then had to remember aspects of their first car ever, and finally map out their ideal future electric car.

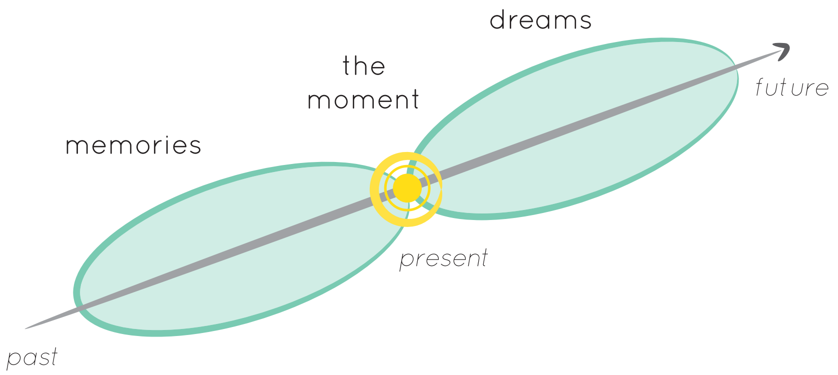


figure 14 | The experience domain, adapted from Visser et al. (2005)

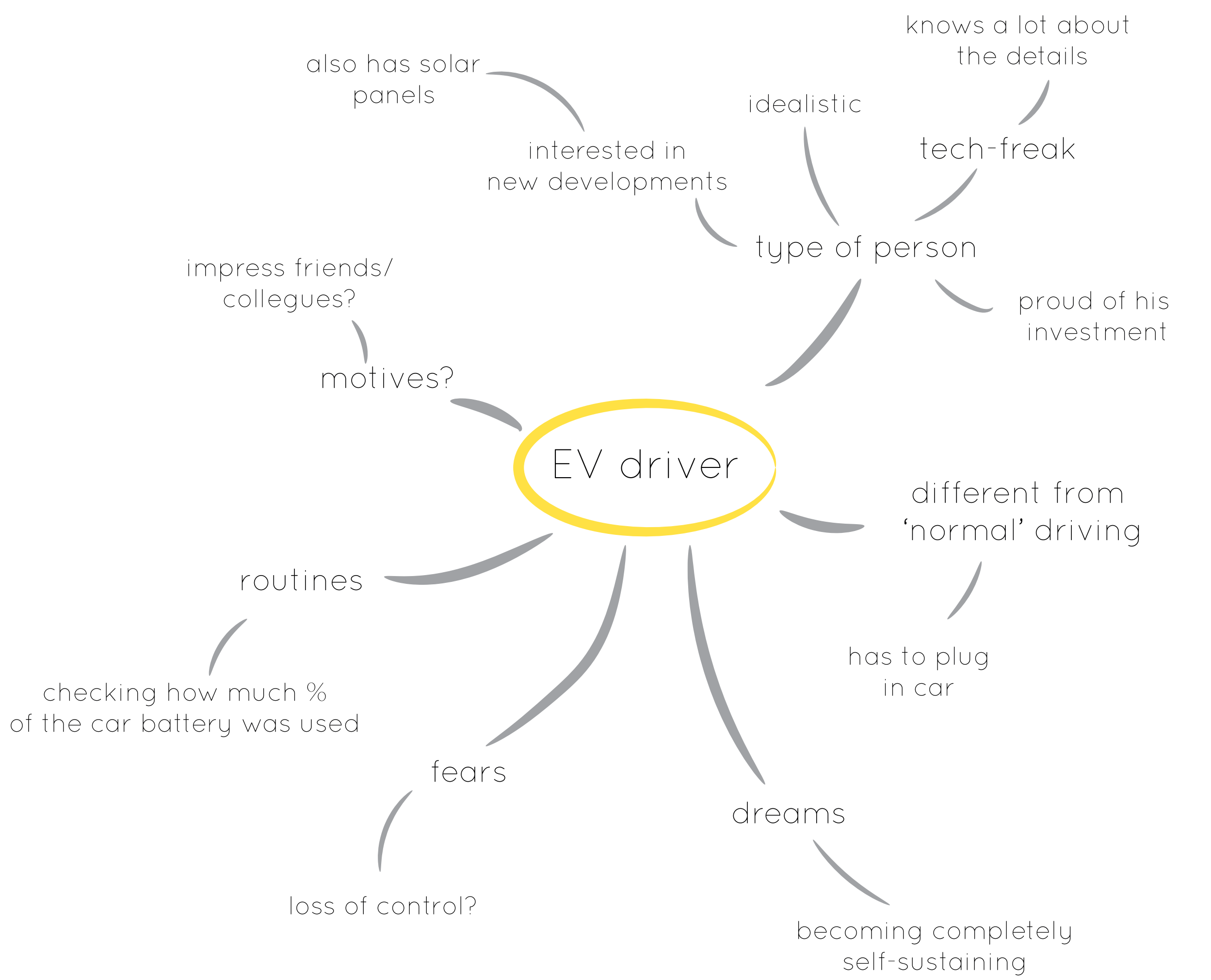


figure 15 | mindmap of assumptions of EV driver

10.2. Method

10.2.1. Preparation

Participant recruitment was done by placing flyers on charging points and electric and plug-in cars around Delft. The same flyer was shared on LinkedIn (Appendix D). This resulted in a total of 7 responses. Two people were not available on the exact date but were willing to help out. 5 respondents agreed to attending the workshop, of which 2 cancelled on the day itself. Therefore, the workshop was held with three people, all men, aged between 35 and 60 years old.

To purge my current knowledge and assumptions concerning electric driving, I created a mindmap of what I expected to find during the workshop (figure 15).

10.2.2. Workshop build-up

The complete workshop plan has been included in Appendix E. First, the participants were asked to sign an informed consent form, which explained what data would be collected and how it would be processed (Appendix F). Afterwards, the video camera and recording were turned on.

The workshop was then kicked off with each participant introducing himself and briefly reflecting on why he had decided to buy an electric car and what he liked about it best. Next, I introduced myself and presented the first exercise, in which the participant would write down recollections of the first car they owned. After about ten minutes, each participant presented their memories while the other participants were encouraged to ask questions. This cycle of individually completing an exercise, presenting it and discussing each other's ideas was repeated in the second exercise. The participants were instructed to create a collage of their ideal electric car with the help of contextmapping materials: stickers with icons and pictures. The icons were related to the topic of driving, energy, and the home, while the pictures were ambiguous, allowing multiple interpretations.

After the participants had created their ideal electric cars in the contextmapping part of the workshop, some more information was given about the content of my research and the design fiction was introduced as a less ideal version of the future. The video was then played for the participants. Afterwards, they were asked what they had seen, what stood out to them and how they felt about that.

10.2.3. Data collection & analysis

Data was collected through the collages that the participants had made (figure 16 and Appendix G), the presentations of their collages, and the discussions about electric driving in general; about their first car; about their ideal electric car; and about the design fiction. The responses were transcribed and then written into statement cards for data analysis. A statement card presents a quote from a participant and a paraphrase formulated by the researcher, combining raw data with an interpretation step (Stappers, 2012). A total of 112 statement cards were clustered with the help of two other IDE master students.

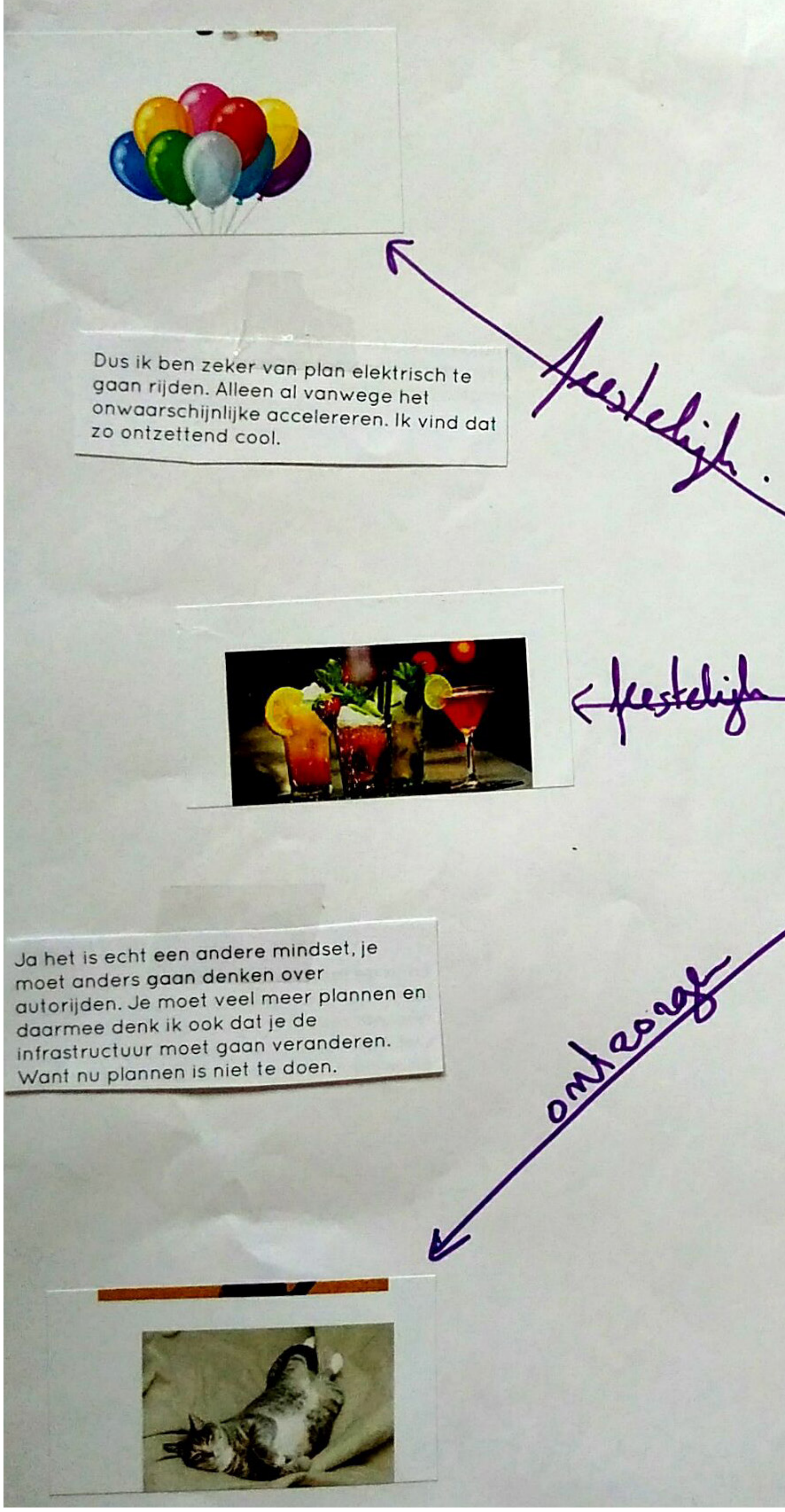
10.3. Results

The clustering resulted in 12 themes. Each theme will be discussed briefly.

1. Challenge accepted

"What I try to achieve, is that I charge at my work [for free], then don't charge at home, and then manage to arrive at my work again… It's a close call. But that's the thrill of course."

The participants exchanged tactics to increase the electric range of the car. Turning off the air-conditioning, trying not to brake, not turning on the music. They saw a challenge in maximalising the electric kilometres, even if it meant handing in comfort.



2. Keep it simple

"You can just turn [the car] on- and off. I really like that, haha. Just 'click'."

Being able to turn the car on and off with just one click was a requirement for the future car of one of the participants. The desire for simplicity also became clear from the appeal of a car that charges directly from the sun, requires no hassle with charge points and thinks for itself about when it should be charged.

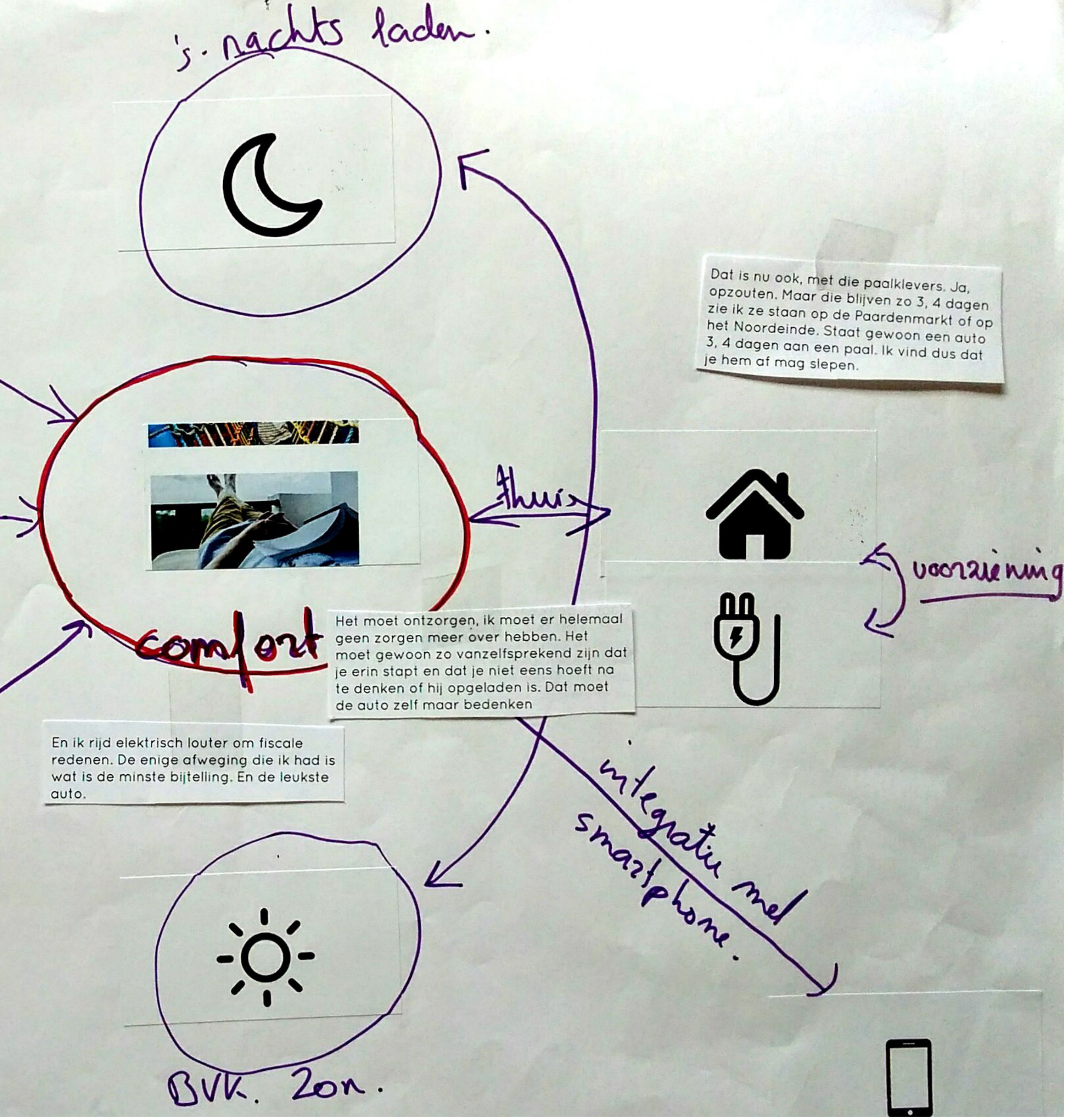


figure 16 | a collage made by one of the participants with descriptive quotes added

3. Car means freedom

"It was my first car, and it was especially the independence that it gave me. Free from anyone else, I could decide to go wherever I wanted, whenever I wanted"

The car, in the past and in the future, was seen as an important facilitator to travel and go wherever they wanted. As part of this role, one participant stated that his car would have to be able to store a bike or a walker, and preferable drag a caravan.

4. The electric driving experience

"…it's especially the peace and quiet. It's unfortunate, because when the motor turns on after 40 to 45 kilometres, well, there's the racket again!"

Various positive aspects of driving an electric car where mentioned: the quiet, listening to music, the absence of a gas smell, and the fast acceleration.

5. My car is your car?

"I think… cars stand still a lot of the time, and you see how the streets get clogged because of it. Delft is at its prettiest when there is a royal funeral and all the cars are gone from the Oude Delft."

Two of the participants wanted to keep on having a car of their own, the third wasn't quite as adamant. He believed a decrease in car ownership would make the city a more beautiful place. His experience with Uber had made him believe that there were promising alternatives.

6. Nostalgia

"What I cherish [about my first car]… we had some problems with the starting motor. But you could push that thing on your own, because it was so light. So you would walk next to it, push it, and then jump in and then shift the gear and it would turn on."

The sound of the motor, having to push the car, the first drive on your own through the rain and the dark… these were all 'first car' memories that were now cherished.

7. Green and financial sustainability

"I like charging to the sun. The local and the sustainable, I like that, even though it's an investment to buy such a car."

Two of the participants were driven by the sustainability of an electric car, while one had specifically bought it because of the low tax addition.

8. Charging troubles

"I skip the charging of my car more and more often. It's the hassle. I have to walk quite a bit, well, half a kilometre. And finding a spot is a problem. The city centre doesn't have an abundance of charging points."

One participant could not charge at his own home and this resulted in a lot of annoyances with charging at public charging points and dealing with people that would take a charging parking space for days. Another participant said the wait at fastcharging stations could be boring.

"It's simple. The computer should always be serving me and therefore I should always be able to overrule it."

workshop participant

9. (Not) my problem

"If it is because "the grid" can't manage [to charge my car]… Well! That's not my problem, they should take care of that. I'm not going to contribute to that."

"But if you have button to overrule system… I can imagine people will think 'I don't want this hassle' and they'll just push the button every time."

The opinions were divided about the ownership of the problem of a grid that can't handle the energy supply and demand. One participant especially felt everyone should just become self-supporting on energy supply. He would only accept giving up a fully charged car if it would help someone he knew, a neighbour who would return the favour. Another participant prospected that people would feel this way, and think from their individual concerns.

10. Techno-optimism

"In the last phase he didn't have control. And the guy [in the video] became nervous. When… maybe it was correct. Maybe it was charged to 45% and the system is so smart that it has considered 'you have to go there, and 45% will be sufficient'. While he was thinking that he needed 55%"

The design fiction also resulted in some positive comments, about the convenience of a system that would adapt to your routines and the traffic of that day. Moreover, during the contextmapping, the participants had shared that their future car would contain all the new smart technologies and a full integration with their smartphone.

11. Staying in control
"It's simple. The computer should always be serving me and therefore I should always be able to overrule it."

While there were also some positive responses to the idea of a smart V2G charging system, it was important for all participants to have some degree of control, and to have some kind of brake when something really went wrong.

12. Privacy matters
"What I'm thinking… why do you want to know all that, man? What are you going to do with that information? It's going to be Big, Big Data, what's going to happen with all that. That's what I'm suspicious about."

One participant was very outspoken in his distrust of big data and the collection of personal data.

10.4. Discussion

There were some limitations to this research that should be discussed. There were only three participants, and they had actively responded to the call for participants, so they were a bit of a unique group that is very interested in development in electric driving. The electric car driving community is in itself quite a homogenous group, with mainly male and high educated drivers. It's not a group that can represent society. It's therefore important to realise that others might not be as interested in electric driving or solar panels and the integration of various technological systems as these pioneers are.

10.5. Conclusion

While all clusters help in painting a picture of the context of the EV driver, there are three that stood out and that I will discuss some more.

First of all, to the participants, the car is a means to go wherever you want, whenever you want. Freedom to travel is thus relevant individual value in this context. If the V2G unit would restrict this freedom, it leads to a conflict with the system. Therefore, a design would have to balance this well, for instance by having the user keep control of how much freedom he offers.

The second cluster that I wanted to emphasise, is the 'challenge accepted' cluster. This might offer a way to motivate the user give up more of his freedom. As the results show, the limited radius that the driver has also leads to a form of gamification in which the user feels challenged to drive as far as possible with the limited energy he has.

Lastly, clusters 10 and 11 show another balance that should be found: on the one hand, the participants are optimistic about what technology can do for them in terms of comfort and convenience and they welcome smarter, automated and integrated systems. On the other hand, after seeing the design fiction they did express that they would want to have the final word or have some sort of brake when things spiral out of control. So a sense of control is an important need in this situation.

The contextmapping workshop with EV drivers provided insight into the user context driving and charging an electric vehicle. The value of the freedom in travelling that a car offers was identified, and also the need to have some control over technology. These findings provided input for the step described in the next chapter.

TO SUMMARISE...

11. Dystopia/utopia

To clearly define the possible, unwanted future and the desired future, a dystopian (11.1) and a utopian vision (11.2) were created. The unwanted future was based on the values and needs that were uncovered through the creation of the design fiction and the contextmapping workshop, and on the possible 'unaccountable' scenarios that were thought out in creating the design fiction.

figure 17 | dystopian vision

figure 18 | utopian vision

11.1. Dystopia

Ffigure 17 shows the dystopian vision, in which the V2G charging unit looms over the user, with a mysterious data creature hides in the shadows. In this scenario, the user has no idea of the responsible party behind the machine and can't ask for further explanation than 'the computer says no.' This not only means that he might have an uncharged car without understanding why, he has no means of taking back control. It also means the algorithm might be disadvantaging certain specific groups, obstructing their freedom to travel. Other negative effects might prove difficult to predict. They develop gradually as the algorithms develop in their socio-technical context.

11.2. Utopia

figure 18 shows the utopian vision. Here, the V2G unit operates in a more inclusive way, as a doorway through which the car driver can collaborate with the responsible party behind it to improve the working of the V2G system and achieve their collective goals. The charging allows the driver to go to work when necessary and when he doesn't use his car he can make some extra money and contribute to the neighbourhood. He has control over the amount of freedom that he needs. Both sides recognise each other as partners in making this work.

The insights from the design fiction, the critical areas in the V2G and the contextmapping workshop were brought together in two possible futures. This provided the stepping stone to the last chapter of this cycle, in which the design challenge is formulated.

TO SUMMARISE...

12. Design challenge

cycle III conclusion

In the previous chapter, a dystopian and utopian vision were sketched out. By considering what should be changed to step towards the utopian vision, a design challenge can be formulated.

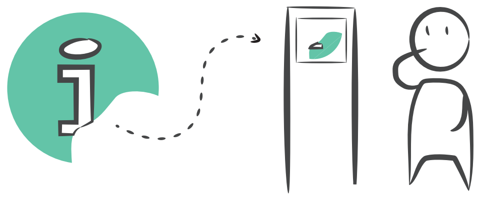
The important difference between the two visions is whether or not the user and responsible parties can 'see' each other. What I'd want to achieve, is that the users and Newmotion, Enervalis and Alliander all feel and work like a team in improving a correct and fair working of the V2G system. For this to happen, the user must be offered 'a dosed transparency' that allows him to understand what is happening without being overwhelmed. In the meantime, the user should also provide feedback about the working of the V2G system, so it can be improved. The interaction that facilitates this exchange of information should be designed. This is summarised in figure 19.

The design challenge that was formulated was to 'design an interaction between the user and the V2G unit that optimises the system while providing relevant transparency to the user'. This raises the question of what an optimal system would be. As the contextmapping conclusions showed, a car means going where you need to go. The importance of a safety option also came up in the creative session: The idea arose that a user would be able to set a minimum based on their emergency locations. This led me to be inspired by the capability approach.

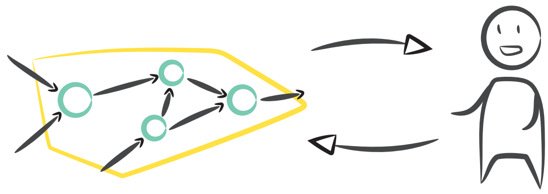
Oosterlaken (2012) describes this philosophy by Nussbaum, according to which the focus for topics such as justice and equality should be on human capabilities, and not on well-being or resources. Capabilities are the positive freedoms that people have to enjoy valuable 'beings and doings' ; so-called 'functionings'. This might be working, being healthy or... being able to travel (Oosterlaken, 2012). To achieve this, a person might need certain resources, but is also dependent on personal, social and environmental circumstances or 'conversion factors'.

This provides a framework to look at the V2G system once again. The capability would be for car drivers to have the freedom to travel. The car is an important resource to do this, and the driver is also dependent on the V2G infrastructure and the amount of energy that other drivers use.

I want users & the companies to feel and work like a team
in improving a fair working of the V2G system.



By providing them a 'dosed transparency' based
on their needs and interests.



Therefore, my goal is to design an interaction between
the user & the V2G unit, which optimises the system
while providing relevant transparency to the user.

figure 19 | the design challenge

This cycle ended with a design challenge that tackles the problem of the lack of the interest of the user, which was chosen as a focus within the V2G system at the end of the previous cycle.

TO SUMMARISE...

13. Ideation

14. Final design

15. Design evaluation

IV
Solution finding

The final cycle started off with the ideation of solutions for the design challenge that concluded the previous cycle. This process will be presented in chapter 13, after which the final design is presented in chapter 14. The evaluation of the design with the V2G project manager and the theoretical frame of the first cycle are discussed in chapter 15.

13. Ideation

Various techniques were applied to generate ideas that answered to the design challenge that was formulated in chapter 12. In a creative session, the meaning of this design challenge was explored (13.1). Based on these ideas, a general user scenario was constructed that would accomplish the challenge (13.2). This provided a starting point to generate more concrete ideas. Finally, an interaction vision (13.3) and a set of criteria (13.4) shaped and framed these seperate ideas.

13.1. Creative session with design students

To get started on the idea generation and gather fresh views on the subject, a creative session was organised with seven students. The session revolved around the problem statement presented in the previous paragraph. The session plan has been included in Appendix H..

13.1.1. Method

The first part of the creative session was dedicated to exploring the design challenge. An important question raised was what 'relevant information' would be. As this would be different for various users, possible roles were created and played out during a role play (figure 20).

After this explorative phase, four subquestions were formulated for which solutions would be generated in a brainwriting session. These questions were:

- *How to tailor relevant information to the user?*
- *How to motivate users to contribute to the system / How can users contribute to the system?*
- *How to make sure the user is & feels safe?*
- *What are indirect effects of the system and how can you prevent them?*

After a brainwriting session in couples, all ideas were put up on a wall and the participants could indicate useful and novel ideas with a sticker. These ideas were discussed collectively.

13.1.2. Results

All ideas were clustered afterwards, which led to six clusters of ideas in the same direction.

Finetuning the message

The timing, tone, medium and content of a message are all aspects that might be adapted to the user.

Building a user profile

Using personal data about the user to get a profile that both predicts driving behaviour and helps in

figure 20 | role playing the V2G unit and the EV driver

providing the right, relevant information for this specific user.

Offering additional information
Also providing other information that is related to driving, such as traffic jams on your route. In this way, the V2G unit connects to your other needs and isn't just the bad guy that sometimes leaves your car uncharged.

Kinds of feedback by user
These were all ideas of what kind of feedback a user might provide, such as warning the system that his routine had changed, setting favourite destinations and rating the driving experience.

Setting certain minimum and average routes
This was seen as a way to make the user feel safe and also to allow for some spontaneity: A quick drive somewhere would always be possible.

Thinking along with the user/providing alternative solutions
Several ideas were proposed for alternative solutions that could be offered if a car turns out to be insufficiently charged, such as allowing your car to be lent out for emergencies, having emergency bicycles, or an emergency battery in the V2G unit.

13.1.3. Conclusion

After reviewing the results of the creative session, three ideas arose as most interesting to pursue. These will be briefly discussed here.

First of all, the information that should be presented as explanation might be divided in three levels: crucial information, which every car driver should see and be able to review regularly, custom information, which is more extensive and should be personalised based on the interest of the car driver; and finally reference information, which will only be consulted in case the car driver has a specific question. Examples are given in figure 21.

Another idea is to have the car and charging point communicate with each other, with the car 'representing' the car driver, while the charging point represents the community (figure 22).

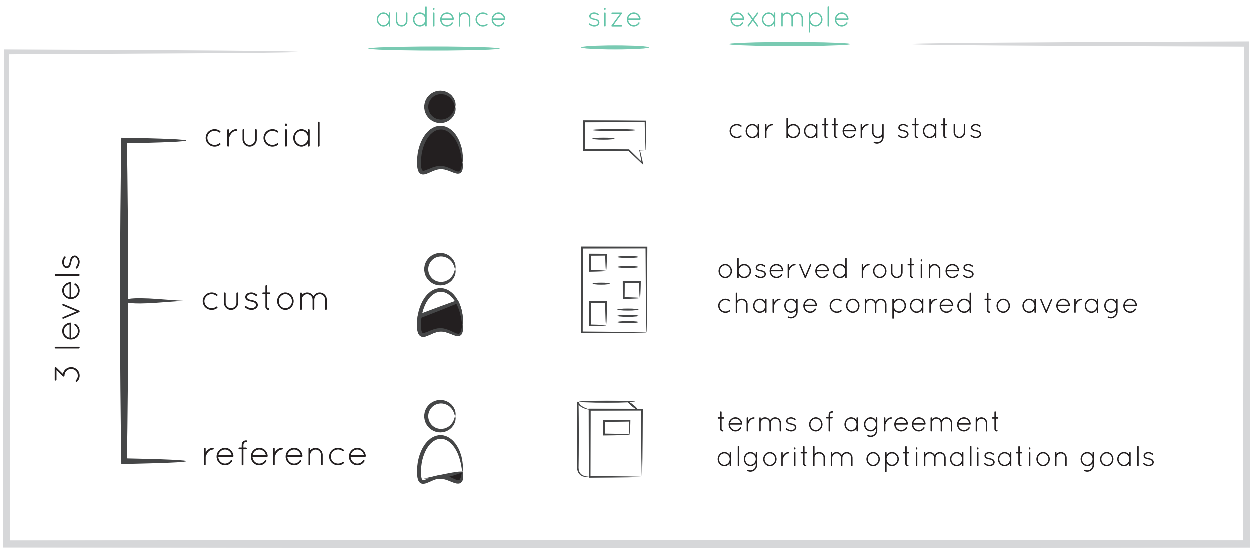


figure 21 | levels of information

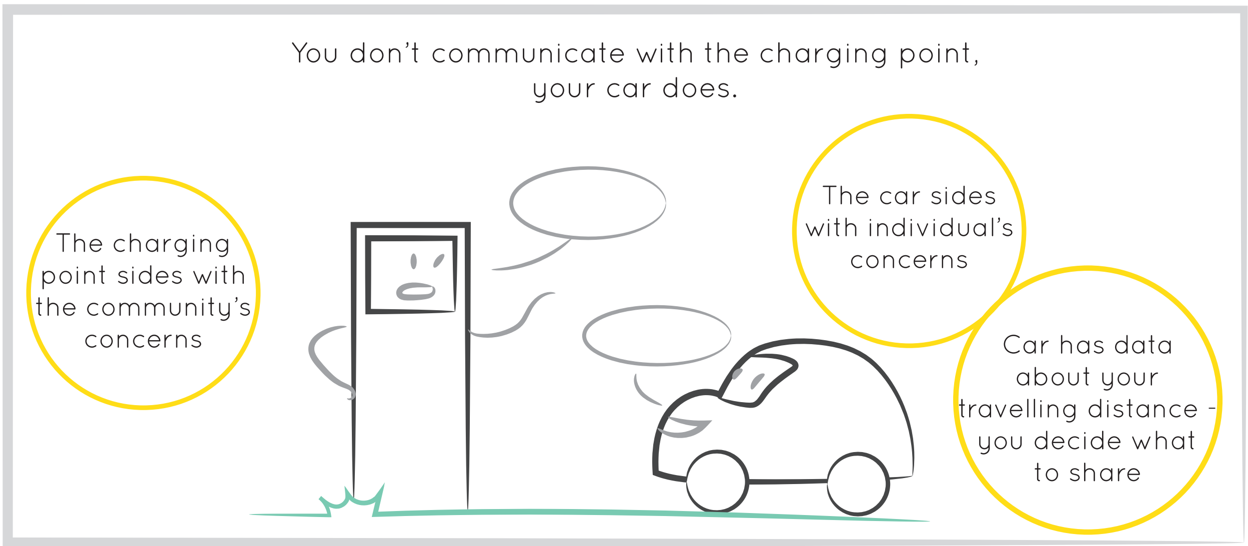


figure 22 | the car and the V2G unit representing different needs

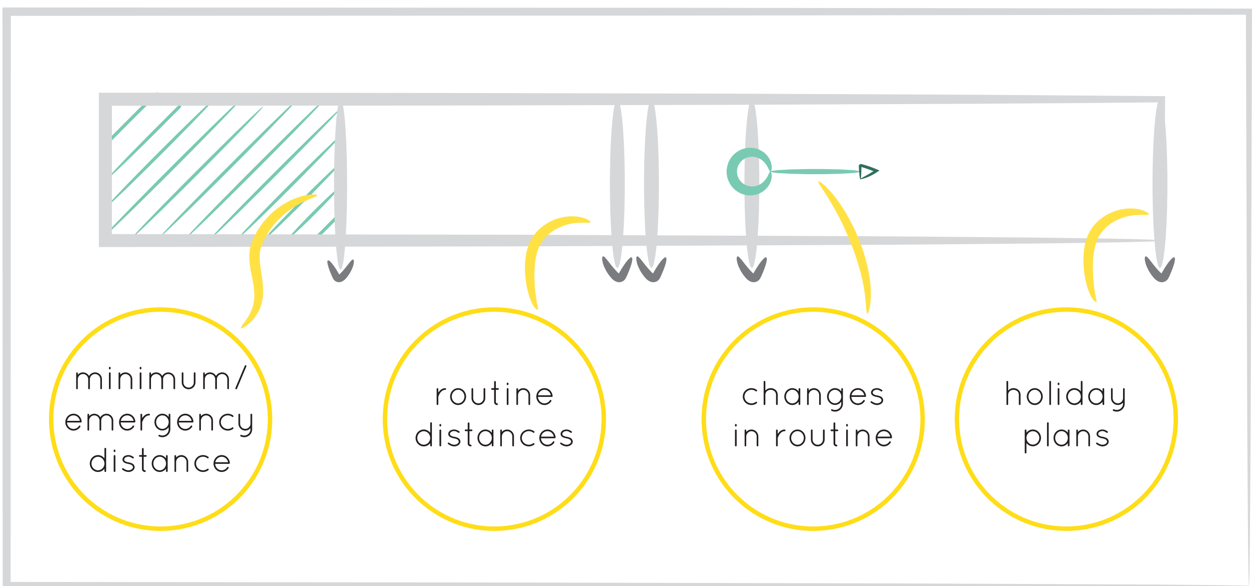


figure 23 | indicating required battery levels

Lastly, an idea was for the user to indicate his routine energy needs and the minimum distance he would need to travel. A change of routine could also be indicated, instead of having to wait for the system to catch up with this (figure 23).

13.2. User scenario

To imagine how the system would work in order to achieve the design goal, a user scenario was created that described three situations: 1) the acquaintance of the user with the system, 2) the regular working of the system and 3) the 'doom scenario' in which the car is insufficiently charged.

This resulted in the formulation of five solution-directions, which can be found in the user scenario.

1. Providing a profile with all data that was collected about the user, and all decisions he/she has made
2. Presenting the day's conditions that might influence the division of energy
3. Enabling the user to indicate that charging is

figure 24 | user scenario showing the five solution directions insufficient and presenting alternatives

4. Providing a way to rate the experience/ comfort level before & after drive

5. Offering insight into the general working and optimisation goals of the system

To further develop the solution directions in concrete ideas, the SCAMPER method was used. In this idea generation method, questions are asked to stimulate the designer to come up with new solutions based on the current ideas. The name of the method is an acronym for the ways in which this is done: Substitute, Combine, Adapt, Modify, Put to another use, Eliminate and Reverse (van Boeijen, Daalhuizen, Zijlstra, & van der Schoor, 2013). The ideas can be found in Appendix I.

13.3. Interaction vision

To get a feel for how the design challenge should be accomplished, an interaction vision was developed that can describe the interaction in an analogy. By analysing this analogy, interaction qualities for the design can be derived.

The interaction should be like being passed a ball in a basketball match. You see the ball coming towards you, and you know you're up next, you have to move the game forwards to the basket. You feel valued because of the trust that your team mate puts in you, and you feel like you can contribute to the game. You feel like a team, you know you will have to pass on the ball as well to score a point.

Translating this back to the V2G system, there are three components that should be present in the design. The car driver should be passed a clear opportunity to direct they working of the system, he should feel this made an impact and he should trust the system sufficiently to hand the control back over.

13.4. Criteria

Now, based on the findings until now, a set of criteria for the final design were set up. These are related to the five aspects of algorithmic accountability and to the new aspects that were found to be of importance.

1. Responsibility
1.1. The role of Newmotion should be clear to the user
1.2. The division in roles should be clear
1.3. The algorithmic system should deliver Newmotion the data necessary to watch over the fair and accurate working of the system.
1.4. There should be a single contact point for the user

2. Explainability
2.1. The optimalisation goals of the system should be clear to the user
2.2. The user should be able to vaguely understand the reasoning of the decision-making
2.3. The user should be able to look up what data is being used about him/her

3. Debatability
3.1. The user should be able to ask for further explanation about a decision

3.2. The user should be able to gather evidence of a negative charging session.

4. Efficiency
4.1. The user should be able to correct inaccurate assumptions in the algorithms decisions about his/her energy needs
4.2. The user should be able to indicate when a charging session has resulted in an insufficiently charged car

5. Fairness
5.1. The system should strive to provide all users with sufficient energy to allow them to live a valuable life (capability approach)

6. Capability
6.1. The user should always be able to leave in an emergency
6.2. The system should measure the battery left once the user has arrived at his destination.

The design challenge as defined at the end of the previous cycle was still quite broad. The process of ideation and creating a user scenario led to a better understanding of the challenge: the user should experience interacting with the V2G system as getting the ball passed to him in a basketball match. Boundaries were set for the final design by defining criteria for responsibility, explainability, debatablity, auditability, sanctionability, efficiency and fairness.

TO SUMMARISE...

14. Final design

The final design will be elaborated in this chapter. First, the assumptions that scoped the problem are presented (14.1). Next, the working principle (14.2) and intended use scenarios of the V2G system (14.3) will become clear. Finally, the accountability mechanisms that were implemented are explained more explicitly (14.4).

14.1. Assumptions

In order to scope the solution space and to lower the amount of uncertainty in the design context, a few assumptions were made. These are listed here.

- Alliander is ultimately responsible for the system. They have to account for their actions and investments to the government, as they are a public organisation. They contract Newmotion & Enervalis. Newmotion asks the municipality for permission to install the poles. This assumption was based on Alliander's current responsibility in the project.
- All car drivers have a charging station at their own house, there is no sharing.
- Newmotion is the point of contact for customers, just like in the pilot.
- The user is the only driver of the car.

14.2. Working principle

The final design is a design for the system that connects the V2G unit, the car and the user, and integrates some of the ideas presented earlier. It is built on the principle of the car that stands between the V2G unit and the user, with the car representing the needs of the user and the V2G unit the needs of the community. The division of 'tasks' is summarised in figure 25.

14.3. Use phases

As the design challenge and interaction vision described, the design should make the different parties of the system into a team, in which the teammates trust each other and feel they make a contribution. A team goes through a few stages before achieving this teamwork. And so do the car driver and the V2G system. These phases have been illustrated in a Product-Service System blueprint (Geum & Park, 2011). This is a visualisation tool that shows the actions of the user, the responses in the interface and the support processes that run in

DATA	travel distances travel destinations charge sessions remaining battery after charge	overall demand & supply of energy
FUNCTIONS	navigation locating other charging units	charge and discharge batteries
REVIEWING	# times individual's car was insufficiently charged	overall division of # times individual's car was insufficiently charged

figure 25 | division of tasks between the car and the V2G unit

the background. I've adapted the scheme to show the support processes in the car and in the V2G unit separately. The PSS blueprint can be found in Appendix J.

See the next pages and figure 26 for a description of the use scenarios. An interactive version of the interface that was created as part of the design can be found at: https://invis.io/W3N9B13D42X

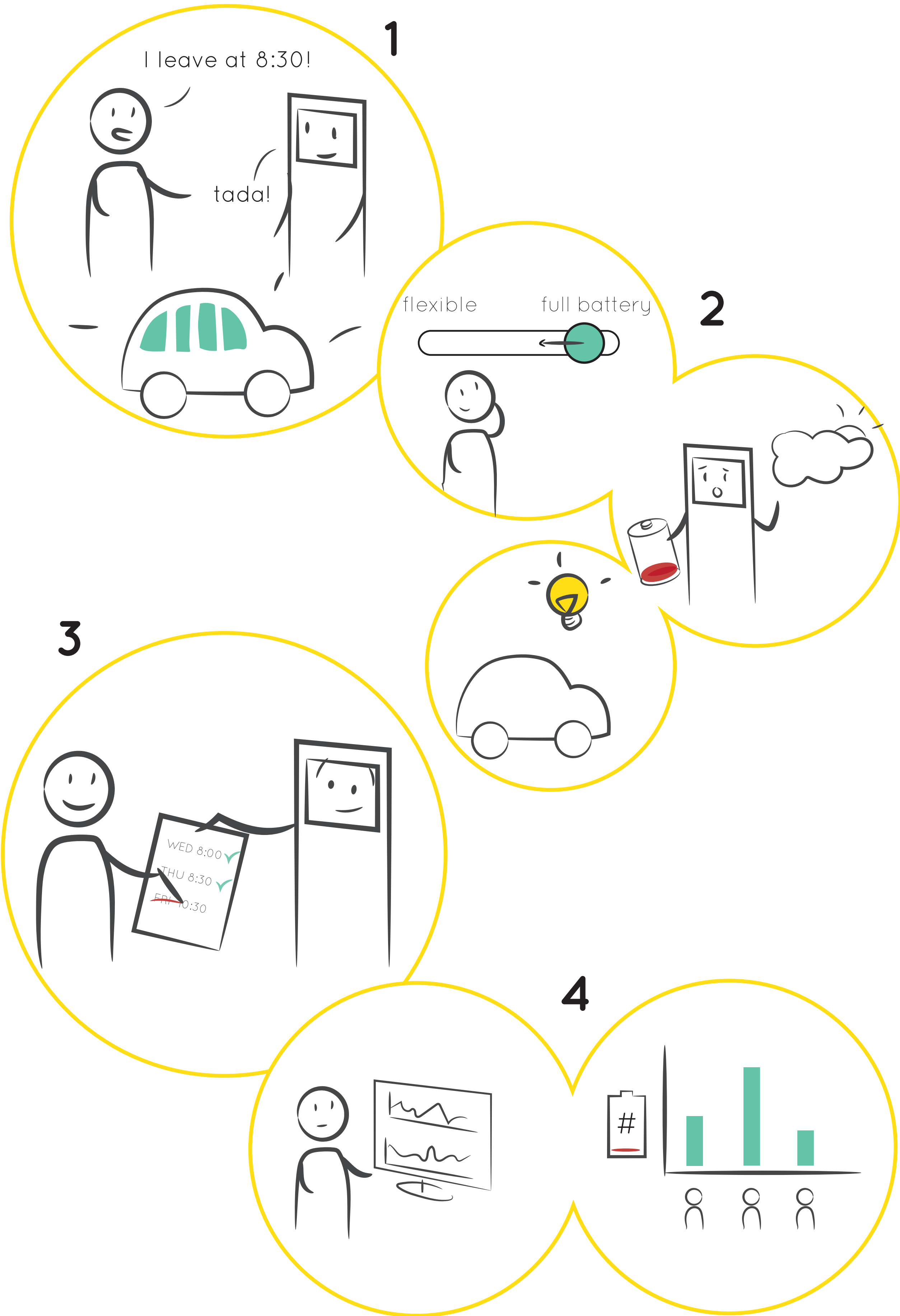


figure 26 | use scenarios of the final design

1

The user is new to the system. He is introduced to its working: how it optimalises to prevent congestion and buy and sell energy at the right time.

At first, the user indicates all his rides and his battery will always be charged up to 100%.

2

Once the user feels comfortable in allowing more flexibility, he can indicate that the system can start predicting his rides and charging his battery only to the amounts that he is expected to need.

Whenever this goes wrong, and the car is charged too minimally, the user can choose to report this and is offered some alternative options: travelling via a fastcharger, fastcharging now (if possible) or finding a shared electric car in the neighbourhood. The user can also review what the current circumstances are regarding the available sustainable energy and the demand from other cars to charge.

3

The user gets used to the system, knows to look ahead and to set predicted rides right if necessary. The system, meanwhile, gets better in predicting the user's routines.

4

The user can keep track of his charging sessions, and label them with his satisfaction level. If he detects a recurring mistake, he can select the related charging sessions and present this to Newmotion.

Newmotion, meanwhile, reviews the battery level that remains after users have arrived at their destination, and the amount of times they had to stop by a different charger. By mapping this out, the company can check whether there are certain people that have to deal with an empty battery more often than others.

14.4. Accountability mechanisms

Within the design, various accountability mechanisms have been implemented. Some of these are also explained in the use scenarios, but in this paragraph they will be explained expllicitly.

14.4.1. Explainability

When the system is introduced to the user, the optimisation goals - money, a balance in the grid, and the use of renewable energy - are presented (figure 29).

Then, in the main screen of the car interface, a graph shows general info of the supply and demand of energy. This allows the user to interpret an explanation for himself (figure 28).

Moreover, the various levels of automation that the user can set, help the user understand what the various stages of automation are and what data will be used. By changing these levels and seeing how this effects his battery level, profit, and the information that he has to provide concerning his departure times, the user can 'sandbox' with the algorithm to see how it works (figure 27).

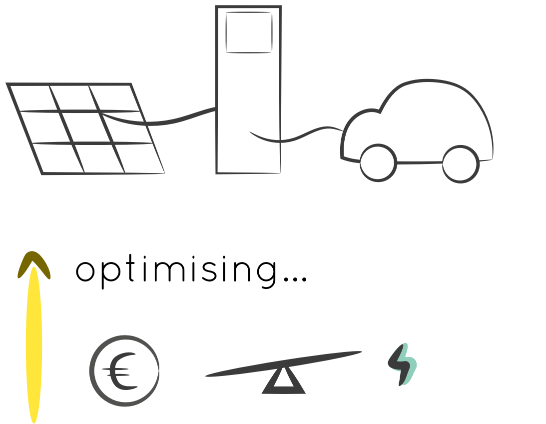


figure 29 | introducing the optimisation goals

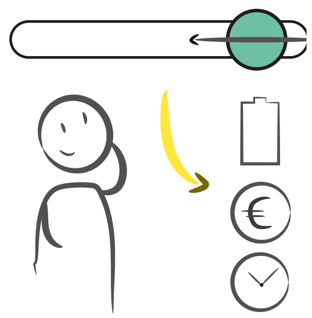
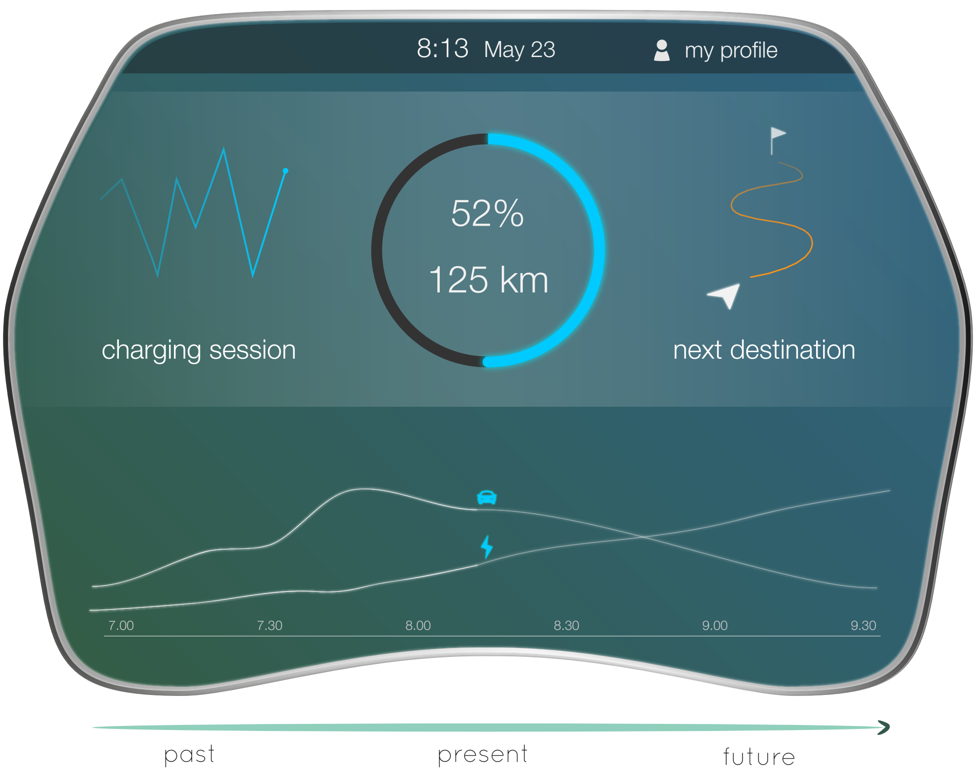


figure 27 | sandboxing to understand the system

figure 28 | main screen with the explanatory graph



14.4.2. Debatability

First of all, if the battery level is too low, the user can report this within the interface (figure 30)

Second, the user can track his charging sessions and combine multiple records and send a more extensive report, commenting and asking for more information. For instance, because it was the same day each week when his battery was charged insufficiently (figure 31)

The third option is a collective act of debating. If a large group of EV drivers lowers their automation level at the same time, Newmotion will feel compelled to present an explanation of what happened, to gain back the trust of the users in the system's working (figure 32).

figure 31 | reporting a set of charging sessions

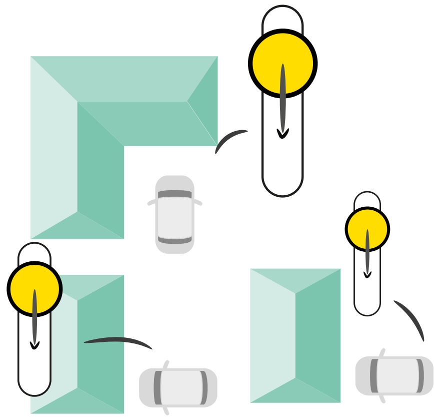


figure 32 | lowering the automation level collectively

figure 30 | reporting an uncharged battery

14.4.3. Responsibility

When introducing the working of the system to the user, a clear contact person explains the working of the system. This could be in person, but also via a video. The collaboration between Alliander, Newmotion and Enervalis is explained here and can also be reviewed later figure 33.

14.4.4. Auditability

As an objective measurement, the battery percentage that remains once a user arrives at its destination can be used, along with whether he has had to resupply his battery along the way. A subjective measurement of rating with stars has been added to this, because what might be comfortable for one, might be stressful for someone else (figure 34).

These measurements are reviewed by Alliander and/or a third party.

14.4.5. Sanctionability

It's in Alliander's and Newmotion's advantage for users to move their automation level upwards and allow more automatic decisions concerning the timing and battery level in charging. If people don't trust the system and stay on the lowest level, their car batteries can only be used minimally. There will be less freedom to play around with energy. Therefore, the way of debating presented in figure 32 on page 85, is also a way of sanctioning.

In case the auditing shows that some users are disadvantaged, Alliander as the ultimately responsible party can enforce Newmotion to change this, and in turn Newmotion can discuss with Enervalis how this should be changed (figure 35).

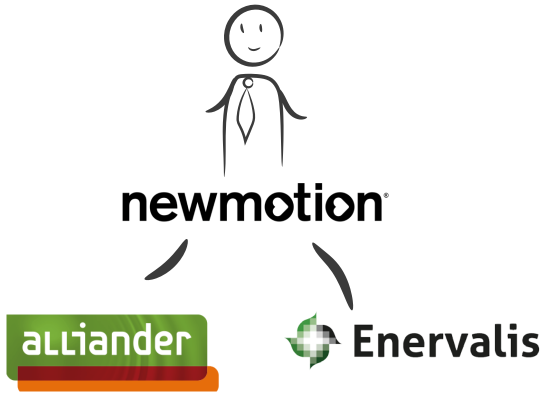


figure 33 | presenting the responsible parties

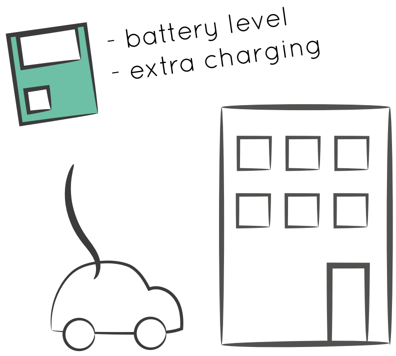


figure 34 | saving and sending data for auditing

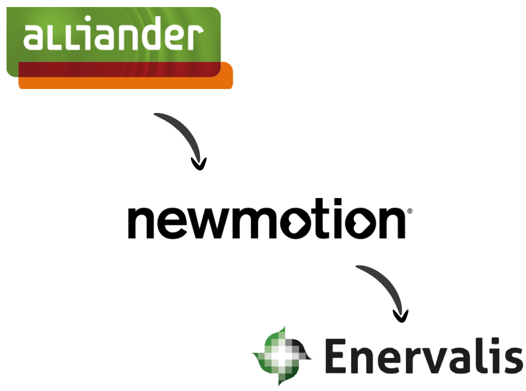


figure 35 | the chain of sanctioning

This chapter presented the final design of the V2G system. This included accountability mechanisms and was illustrated by working out the interface through which the user can understand and control the system.

TO SUMMARISE...

15. Design evaluation

cycle IV conclusion

The design presented in the previous chapter, was evaluated in two ways. The feasibility of implementing such a design in the V2G system was evaluated in an interview with the project's manager (15.1). The extent to which the mechanisms presented in 14.4 would make the system more accountable was evaluated theoretically (15.2). The results of both evaluations are presented in this chapter.

15.1. Feasibility evaluation

To evaluate the design and discuss its possible working in practice, another meeting was planned with the projectmanager of the V2G pilot. During this meeting, the findings of the research and the design were presented and discussed. The various functions were discussed and ideas about further development came to light.

15.1.1. Method

A short presentation was given of how the research had taken course since the last meeting, after which the use scenario as presented in the previous chapter was elaborated. Central to this evaluation were the questions:

- *How would (or wouldn't) the design work in the V2G system?*
- *What are ideas for improving the concept for implementation in the V2G system?*

The discussion was recorded with permission. This recording was used to complement the notes that were also taken.

15.1.2. Results

This paragraph summarises the points that were made in the discussion of the design.

- The design assumes that the user will follow this process from forming to performing. But what if he doesn't? Can/should he be tempted or convinced to take these steps in some way? There should be some sort of incentive for handing over your certainty. Perhaps it could be communicated how much more money you can make.

- For Alliander and Newmotion, it could be important to see on which automation level users are, as a feedback mechanism. Especially if an update would lead to a massive lowering of levels, this could be a clear sign that something is wrong or that perhaps the update was too big of a change. 'This could be valuable information that we did something wrong.'

- Removing planned rides easily is important for improving the system. Otherwise the system might think a user will need a charged battery when this is not the case.

- The 5-star rating seemed too vague. On what would this rating be based then? Is it arriving to your destination? Or is it having a full battery? It might be more interesting to ask more concrete questions, 'did you arrive at your destination without concerns?' Perhaps it might be a different question each time

- To interpret this rating correctly, it would be important to link it to the comfort level that the user has selected. If a driver would have settled for a low comfort level, he will probably also be less dissatisfied with an insufficiently charged car, because he has accepted that risk.

- The analysis for whether certain drivers are disadvantaged would also have to separate different 'automation levels' from each other.

- The way the user's settings are arranged now, the comfort level that a user can move to would be dependent on his settings for data sharing. This should perhaps be indicated more clearly, or a higher level might be blocked, unless you provide the data that allows correct predictions to be made.

- Privacy might be a challenge: will this work without personal data? Could the system provide more information about the circumstances than just the supply and demand of energy, i.e. when there has been an emergency, or a special priority person.

- It would be interesting to research how people like doctors would use this system. Would they just stay on the lowest level because they know they can't take any risks? And how large will this group be that can't afford the risk of an uncharged battery? Perhaps it's very small.

Especially if an update would lead to a massive lowering of levels, this could be a clear sign that something is wrong or that perhaps the update was too big of a change. 'This could be valuable information that we did something wrong.'

- The current way of presenting the supply and demand energy would probably work well, and shouldn't be extended to much or it might become too complex.

- What if the user could review how many people are in the same level of automation as he? And how many people are higher? Perhaps this could be a motivating factor to move up. "Forty percent of the users are already on level four" "You are one of the 12 percent fools that are still down below…"

- Perhaps the rise in level could be linked to the amount of information the user can view. As you go up, what happens will be more unpredictable, so perhaps you would need more information to understand what's happening. It's hard to expect how much information people need, but maybe this need will change. And having access to more information will also increase the feeling that you're part of a complex energy system. Also because as he moves up, he'll have been in the system longer and is also more prepared for new information.

- It's interesting how the user can have influence in this way. It could still happen that a battery is empty, but he will have a way of responding to this. But it means that if there is a problem with energy supply or demand the V2G system can still do its work.

It's interesting how the user can have influence in this way. It could still happen that a battery is empty, but he will have a way of responding to this.

15.1.3. Discussion & conclusion

It would have been even more interesting to have done this evaluation with more different stakeholders, i.e. also having Newmotion and users at the table. Nonetheless, the evaluation provided some valuable insights from a more practical point of view.

It seems that especially the shifting of the comfort level was regarded a promising mechanism for giving the user some control over the system and providing feedback to the service provider. Further developments of the design could work out this mechanism further, for example by investigating how the user can be motivated to move up the ladder. This could perhaps happen through challenges, as the contextmapping research showed how this triggers EV drivers now. There could also be a social motivator, like showing that most people nearby are on a higher level. Finally, of course, there can be a financial stimulus.

15.2. Theoretical evaluation

After evaluating the design from a practical point of view, this paragraph will evaluate the extent to which the accountability mechanisms fulfil the requirements for accountability that the theoretical framework set. This will be done by discussing each of the five aspects: explainability, debatability, responsibility, auditability and sanctionability.

15.2.1. Explainability

Explainability was defined as the ability to explain the working of the algorithmic system towards the people affected by algorithmic decision-making. This encompasses both the general working of the system and the explanation of specific decisions. In the design, a precise explanation was replaced for general info of supply and demand of energy. As discussed in the evaluation with Alliander, this could be extended with information about priority users. This allows the user to interpret an explanation for himself. Moreover, the various comfort levels can also help the user understand what the various stages of automation are and what data will be used.

Further testing would need to explore whether such an explanation would suffice for users. Research suggests that it might be better not to disclose all details to the user. Kizilcec (2016) investigate trust in algorithmic interface design in the context of peer assessment in a MOOC (Massive Open Online Course). His finding was that the transparency should be neither too little nor too much, in order for the user to hold a positive attitude to its outcome. Especially if the expectations of a participant regarding their score was violated, more explanation about this result was required. In an interview with MIT Technology Review Jeff Clune, an assistant professor at the University of Wyoming, points out that the impossibility of a detailed explanation of AI decisions is actually similar to how many aspects of human behaviour can't be completely clarified. ""It might just be part of the nature of intelligence that only part of it is exposed to rational explanation. Some of it is just instinctual, or subconscious, or inscrutable." (Knight, 2017)

15.2.2. Debatability

Debatability is the opportunity to ask for further information about decisions from the responsible party.

In the design, this was implemented with the possibility to report charging sessions and collect different charging sessions from history. But as it turned out during the evaluation with Alliander, the setting of the comfort level might also become a mechanism for debating the working of the system. When many users move down a level in response to a new update, Newmotion or Alliander might feel obliged to send more information about their recent change to the system. In this way, an indirect mechanism is provided through which more information can be requested. A challenge that remains, however is how exactly Newmotion should respond to these requests. It will most likely not always be possible to provide specific information about a decision. It would require some more thought to come up with ways to work around this, and also more information from Enervalis about what can be communicated.

15.2.3. Responsibility

The aspect of responsibility involves being responsible for the effects of an algorithmic system, and providing clarity over who has the authority to step in when the algorithm has unwanted individual or societal effects.

In the design, the working of the V2G system is introduced by the party that is responsible, that represents the collaboration of stakeholders. So, the communication about responsibility has been thought out in the design, but not the 'behind the screen' organisation. Also, the steps after choosing to report a charging session were not developed yet in the design. but this could of course be an opening for establishing a link between the user and Newmotion as the service provider; for instance, by receiving a personal message in response to the report.

15.2.4. Auditability

Auditability is the possibility for a third party to review the working of the algorithm.

In creating the design, I considered what data could be used to evaluate a correct working of the system. As an objective measurement, the battery percentage that remains once a user arrives at its destination can be used, along with whether he has had to resupply his battery along the way. A subjective measurement of rating with stars has been added to this, because what might be comfortable for one, might be stressful for someone else. These measurements should be seen in the context of the comfort level that an EV driver has chosen. This information could be reviewed by Newmotion, but also by a third party, some kind of oversight authority. Further developments of the design might come up with ways of visualising this data to facilitate the auditing.

15.2.5. Sanctionability

Sanctionability is the possibility of consequences being attached to giving an insufficient account for algorithmic decisions.
In this case as well, the comfort level setting turned out to be a helpful mechanism. It's in Allianders advantage for users to move their level upwards and allow more automatic decisions concerning the timing and battery level in charging. If people don't trust the system and stay on the lowest level, their car batteries can only be used minimally. There will be less freedom to play around with energy. This is a loss for Alliander, because they want to lower congestion. Therefore, instead of the consequences being legal, or related to loss of reputation, this offers a way of sanctioning that is part of the socio-technical system itself. It has become a matter of trust that needs to be established between the system and the user. Trust is a recurring topic in the interaction of humans with algorithmic systems. Eslami (2017) describes how reviewers that notice the biased algorithms in the grading system of Booking.com lose their trust in its working and start warning other users through their review. By creating an 'algorithm-aware design' Eslami hopes to shape interactions between users and socio-technical algorithmic systems that are more engaged and trustworthy.

15.2.6. Conclusion

In conclusion, the design seems to touch upon all aspects in some way, but there also open issues that came to light in this reflection. The steps that follow after the user decides to report a charging session would be a part of making the system more debatable, and also could make the responsible party clearer to the user. Moreover, the exact way in which auditing can be done has not been worked out yet. In general, the design still focused very much on the interface and less on the organisation around it. Some ideas have been proposed to improve this. To further develop most of these, more input from the stakeholders would be required. Setting this aside, the design does provide solutions for each of the five aspects and therefore provides starting points for implementing accountability.

By evaluating the design from a practical and theoretical perspective, the promising features could be identified as well as those that could be further developed. In general, the design offers opportunities for implementing accountability, but could still benefit from another design iteration in which the arrangements between the stakeholders that should complete the accountability are captured.

TO SUMMARISE...

16. Reflection on design process

17. Conclusion

18. Discussion

The final act

The final part of this thesis revolves around answering the main question and placing the findings and implications.of this research in a broader context. Chapter 16 first reflects on the design process of the case study. This led to the proposal of a set of guidelines, based on the theoretical foundation and enriched with the insights from the design process (chapter 17). In the discussion, the overall limitations of the research and the conclusion will be discussed and recommendations for further research will be provided (chapter 18).

16. Reflection on design process

By reflecting on the design process that I followed in designing for accountability in the V2G system, insights might be gained on how a similar process could be executed more effectively and in a more structured way. The reflection discusses each of the three cycles that were made within the V2G case study context.

16.1. Cycle 2

Cycle 2, which was the first cycle focused on designing for the V2G case study, was aimed at understanding the context of the case; the connections between the stakeholders, the working of the V2G system and the purpose for which it was built. I did this by separately interviewing a representative from each company, and I think this worked well, because it gave each party a chance to talk freely. Some different ideas about how the system works came to light here, as well as the uncertainty about who is responsible.

The theoretical framework guided the questions of these interviews, making sure that all aspects of accountability were discussed. Analysing the context in light of the aspects of accountability also showed the complexity of the situation. A bit too much, perhaps. This had to be scoped down in some way, and that is when I chose to focus on the user and more on explainability and debatability.

In this same cycle, I developed a design fiction. In the process of creating this story, I tried to imagine the possible future interactions with the V2G system. Because of the unpredictable nature of algorithmic decision-making, this requires some creativity. I believe now that I could have approached this more 'openly'; I was tempted to create doom scenarios but thinking in more different directions about what might happen could have led to more inspiring and surprising insights. In any case, it's important to understand general flaws that an algorithm might have, such as working most effectively for the most standard and predictable people, it's difficulty in embodying complex concepts such as 'fairness' and the limitations to explaining its reasoning.

16.2. Cycle 3

This cycle started with a contextmapping research. I explored the user perspective more and gained a better understanding of the context here. In retrospective, this research might have been more in place in the previous cycle, as a part of understanding the context.

In framing a design challenge, the utopian vision that I constructed for the ideal situation gave some direction, but at this point I also had some difficulty in keeping apart design purposes related to the ideal working of the algorithm and those related to designing accountability. Looking back now, I would say it's better to keep these goals more separated. In other words, to first get a clear picture of the context and the desired working of the algorithmic decision-making (preventing an empty car battery), and then to start thinking of accountability mechanisms that can protect this vision (providing an explanation when it is empty). By keeping these goals separated, the design process would have become more focused and the design challenge more specific, which facilitates the generation of fitting solutions.

16.3. Cycle 4

As I said, in framing the design challenge, a clear distinction was lost between the goals of a fair and inclusive V2G system and the goals of an accountability mechanism. I wanted to develop a design that would work effectively for users, that would quickly adapt to their routines so they would experience as little trouble in charging their car as possible. At the same time, I wanted to design accountability mechanisms for if the system would obstruct the user's travelling freedom. The problem here lies in the fact that I was unaware of this two-sided design challenge that I was pursuing. This led to a fairly chaotic ideation. Here, the interaction vision led to more clarity, because it captured the essence of what I wanted to achieve with my design. What I would suggest based on these experiences, is that cycle 2 should have ended with a formulation of a goal concerning a fair and accurate working of the system, perhaps also with an interaction vision that illustrated the desired interaction. Cycle 3 then, might have ended with the formulation of the design challenge concerning accountability. This would have demarcated the solution space for cycle 4 more clearly.

In creating the final design, some more guidance would also have been helpful. Intuitively, I generated ideas for the different aspects of accountability (ways of explaining the level of the battery, opportunities for debating the charging sessions, etc), but along with this came a feeling of uncertainty whether I was covering all problems. While some uncertainty in the design process isn't necessarily bad – and in any way, it's inevitable – I believe the ideation and the development of the final design would have benefited from considering each of the accountability aspects more structurally. The barriers and opportunities found in the interviews could have provided more input.

As the evaluation showed, the final design offers opportunities for implementing accountability, but could still benefit from another design iteration in which the arrangements between the stakeholders that should complete the accountability are captured. I believe a closer collaboration with the responsible parties for the V2G system would have led to a more elaborate design, because this would have provided more input about how arrangements can be made concerning who audits the data and how Newmotion would respond to a report from a user that wants more information about a specific decision.

Overall, because this was a first shot at designing for algorithmic accountability, there was still a lot to discover and there was sometimes a lack of structure. Through this reflection, I have sought ways to improve this. The main reflections for each cycle were:

- Cycle 2 could have focused more on just understanding the context from all sides, the parties responsible for creating the system, the algorithm itself, and the users or decision subject.
- The design fiction then might've fit better in cycle 3, and it could have been a more extensive process of imagining possible futures. The method proved effective for imagining possible developments of the algorithm and this can help in designing for effective accountability measures.
- For the ideation cycle, a clear differentiation in the various goals would have helped in coming up with more fitting ideas. In finalising the implementation of algorithmic accountability, a close collaboration with stakeholders would likely be necessary to make the design complete and decide on specific arrangements for achieving accountability.

Based on the insights that were gathered by reflecting on my design process, I will present guidelines that answer the research question in the next chapter, the conclusion.

TO SUMMARISE...

17. Conclusion

The aim of this research has been to work towards a framework that structures the challenge of implementing algorithmic accountability, providing guidance for designing solutions for implementation thereof. In light of this, I formulated the following research question: How can algorithmic accountability be implemented 'by design'? In this chapter, an answer to this question will be provided.

In a research process of four cycles, a literature review of algorithmic accountability has been executed, concluding with the proposal of a framework that defines accountability. Based on this framework, the socio-technical context of the Vehicle to Grid (V2G) case was investigated and a focus was chosen within this complex system. A contextmapping session of the experiences of the EV driver followed, leading to a better understanding of the individual values that are to be protected. Lastly, a final design containing various mechanisms that support accountability was presented.

Based on the findings from this design-based research and in answer to the research question, I have developed three main guidelines. I will argue each of these in the following paragraph. The guidelines will be presented from the perspective of an outsider looking in, but it could also be done by a party that is part of the system. To illustrate the guidelines, I will refer to the design excursion and V2G case study. The guidelines are presented in the form of stages in the design process. Like any design process, it isn't linear, and it might be necessary to go back and forth.

Step 1. Understand the context in which the algorithm is created and employed. This can be achieved by interviewing stakeholders and using design research methods such as contextmapping to investigate the context in which the algorithm will be used.

This is important for two reasons. First of all, the values and needs of users that have to be protected should be known. This is related to designing the system in a way that is fair and accurate, for which responsibility should be taken. This should be seen separately from the goal of designing algorithmic accountability or else it might cause confusion when coming up with solutions for accountability. Second, algorithms develop based on the way they are set to optimise, so the purposes of the creators of the system need to be clear. Also, once the algorithm is implemented it will continue to develop based on how people interact with the algorithms. It is important to understand this, in

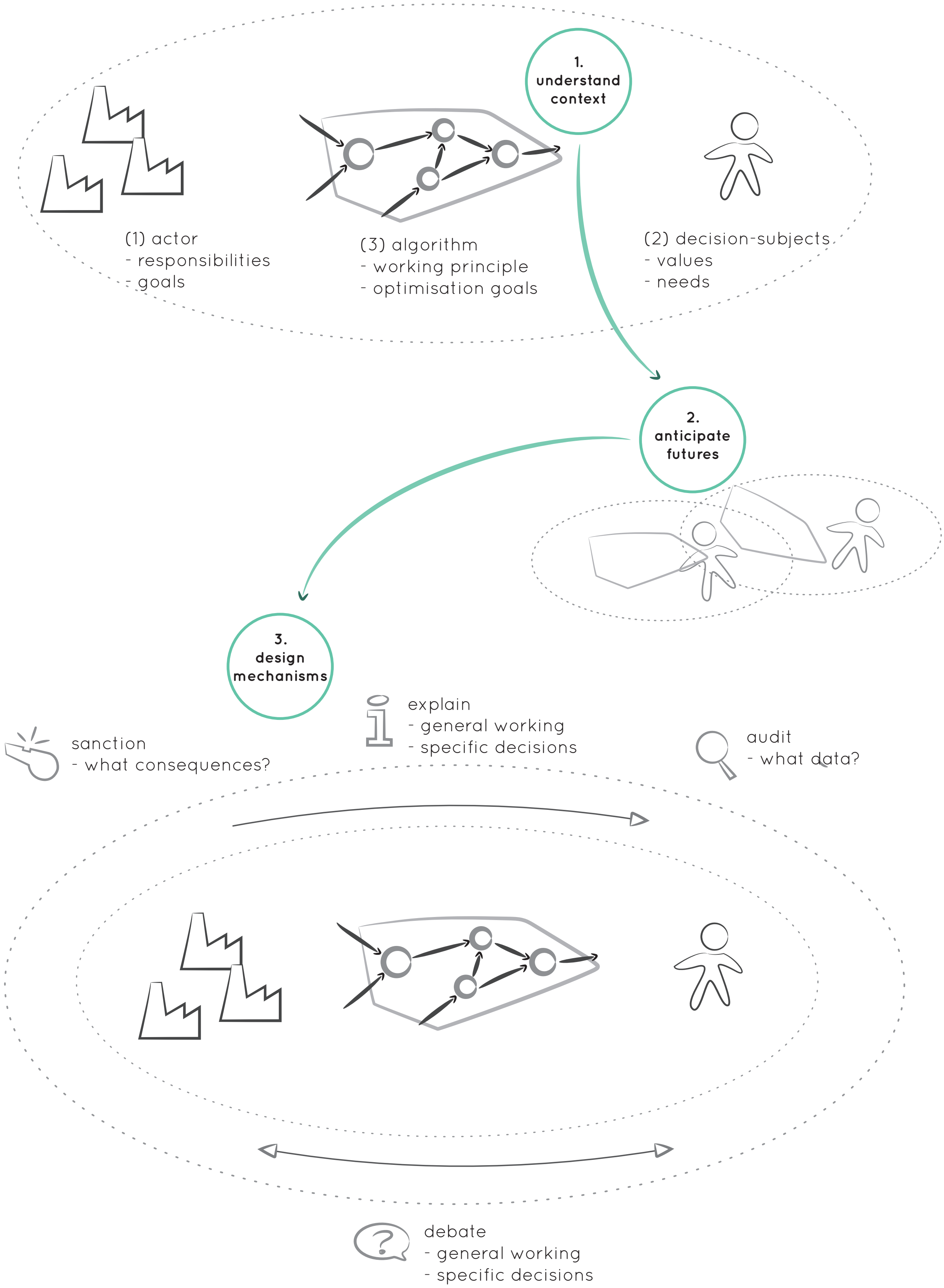


figure 36 | guidelines for implementing algorithmic accountability by design order to follow the second guideline.

In the case of the V2G project, the purposes of the stakeholders were to prevent congestion and make profit from playing the energy market. These motives determine how the algorithm will be programmed to optimise. An EV driver would be able to benefit financially from this, but it might be more important to him to be able to leave spontaneously or always be able to travel a certain distance to visit his mother.

Once the context is sufficiently understood, step 2 can be executed.

2. Anticipate unwanted developments or effects of the system. This research has investigated accountability as an 'after-the-fact' way of governing algorithmic decision-making. So, holding the algorithm and its creators accountable is seen as a response. But the options for accountability mechanisms such as explanations and auditing are limited, because of the technical complexity of algorithms, the limited opacity, and the limited resources or technical knowledge of those reviewing the working of the algorithm. Extensive explanations, opportunities to debate and measurements for auditing won't be effective or even be feasible. By anticipating what undesired effects of the algorithmic decision-making might be based on your knowledge of the context, specific accountability mechanisms can be placed like a safety net.

In the case of V2G, these unwanted scenarios were thought of in the creation of the design fiction, but. Furthermore, they were based on the insights from the contextmapping workshop and based on a general understanding of the flaws of algorithmic decision-making. These scenarios were captured in a design fiction and later in a dystopian vision and included situations in which the EV driver couldn't go to his work or had no way to control the working of the V2G unit.

3. Design context-specific accountability mechanisms.

This is where the four aspects of accountability come in: explainability, debatability, auditability and sanctionability. Responsibility has been left out, because this is more related to the first stage of understanding the context and defining specific responsibilities concerning the fairness and accuracy of the system. Based on this and on the prediction of possible unwanted effects, decisions can be made for how the four aspects can be implemented.

Each aspect had been defined in the theoretical framework of algorithmic accountability, but the description has been extended based on the findings in the case study.

- Explain. Explainability entails the information that the user needs to understand the general working of the system and the specific outcomes of the algorithmic decision-making. It should be determined what the relevant information is for the user in order to reach his goals in using the system.

- Debate. For debatability, the user should have the opportunity to respond to the working of the system and ask for more information if desired. It should be considered what the user might want more explanation about, in what situation this might be and how he might receive more information.

Explainability and debatability are more closely linked and play out through interaction between the algorithm and the user or the creator and the user. In the final design for the V2G system, the interface presented a graph of the supply and demand of energy throughout the day as an explanation for the battery level of the car. If the explanations for the decisions are valid, the user will be less likely to debate the working. This shows how the two aspects depend on each other.

- Audit. Auditability is the possibility for a third party to review the working of the system. To enable effective auditing, it should be determined what data should be audited. These should be relevant measurements that show whether the desired situation is still achieved. Besides this, possible auditors should be identified.

- Sanction. The fourth aspect of accountability is sanctionability. This is the possibility to impose sanctions when the account of the system doesn't suffice. To consider this, one should ask what possible consequences might be and who might execute these.

Auditability and sanctionability are more likely to depend on arrangements on a 'higher', organisational level. External auditing requires a third party that has the required technical knowledge and is given access to the data that is deemed relevant for auditing. In the V2G case study, sanctioning was achieved by giving the users control over the extent to which V2G unit automates its decisions. But if we consider the case of the smart lanterns once more, this would likely have to rely on an external party sanctioning an improper working of the system. Because of the focus chosen based on the causal loop diagram, these aspects have been worked out less thoroughly.

It might be of course, that the accountability relies on one aspect more heavily than another. The set up of the V2G system offered a clear touch point between the system and the user, through which the user could debate decisions. Because the EV driver would use the system regularly, there would be time to slowly learn how the system works. But in the case of the smart lanterns in the design excursion, this would prove more difficult. Here, it might be better to rely more on auditing by an external party.

Based on the insights from the research and design activities in this research, I have sought to answer the research question. I conclude that the following guidelines will structure the process of designing accountability mechanisms: (1) understand the context in which the algorithm is created and employed, (2) Anticipate unwanted developments or effects of the system, (3). Design context-specific accountability mechanisms based on these findings. There are four aspects to consider: how to explain the decisions of the algorithm to the user, how to offer a course of action to debate the algorithmic outcomes, through what data and by whom auditing should take place and how to sanction the responsible parties when the account of the system's working doesn't suffice.

TO SUMMARISE...

18. Discussion

This thesis has investigated how algorithmic accountability can be implemented by design in the context of the Vehicle to Grid system, with the ultimate goal of governing algorithmic decision-making and making sure its outcomes are fair and efficient. This chapter will review the limitations of the methods applied, discuss the implications of the findings of this research, and present recommendations for further research.

18.1. Research approach

18.1.1. Combining research and design

As stated in chapter 2. "Approach", the research field related to governing algorithms requires more practical guidelines and frameworks for designing algorithms. That is why a design-based research approach was chosen in which a design process was followed in a case study. Throughout this design process, I've encountered some struggles, which have been described in the reflection. These practical obstacles show the power of researching by means of a design process, because only by encountering them we realise their existence. In this way, practice and theory can be linked together.

I believe that starting out my research with a design excursion, looking at the practical side before defining the theory also led to a more practice-minded approach. In this case, this worked out well, because the aim of the research was to move away from the philosophical discussion of how algorithmic accountability should be achieved, and to focus on how it is actually designed.

However, it should also be mentioned how this design excursion steered my process. The focus on accountability towards the user started in this excursion and continued for the rest of the research. It should be acknowledged that this doesn't mean that this is the only or most effective of achieving algorithmic accountability. In paragraph 18.2 I will further discuss the implications of this.

18.1.2. Validity and generalisability

In judging the validity of this research, it should be noted that a design process is always dependent on the choices by the designer and so it is very unlikely, that someone following the same approach will come to the same design or even the same design process. Nevertheless, the evaluations of the final design did show that the process had provided solutions that implement algorithmic accountability, and so the guidelines provide a valid answer to the research question.

Concerning generalisability, the guidelines are presented in a very general way, not specifically for the energy infrastructure. In this, I have followed the example of other authors concerned with the challenges of governing algorithmic decision-making. It seems that the field has yet to develop more, before more specific frameworks might arise. I expect the guidelines will provide support in many different contexts, but do imagine three situations to which they are best transferable. First of all, situations where there is a direct relation between the responsible party and the user. Second, situations in which the creators of the algorithmic system are interested in building it in an accountable way. Third, situations in which there is a clear touchpoint between the user and the system.

18.1.3. Limitations

There are some limitations to the research that should be mentioned.

First of all, the validity of the literature review. The search included the terms 'algorithm' and 'algorithmic decision-making', but some authors might use the terms 'big data', 'machine learning' or 'AI' while discussing a similar topic. It could be that some relevant studies were missed as a result. Moreover, the research on algorithmic accountability is still limited and so I combined literature from a range of different contexts, placing different types of algorithms on the same pile.

Second, if the design had been tested with users, more definitive conclusions might have been drawn about the extent to which it established accountability. Especially relating to the final design, some hypotheses have been generated about how the interface will lead to a better understanding of the working of the system, more engagement with its working, and ultimately to a more accountable system. A longitudinal study would have to show whether this is indeed the case or if it isn't, what the other effects are on the working of the V2G system.

18.1.4. Integration of SC and DfI methods

Concerning the approach, I would also like to share my thoughts on the value of integrating Science Communication and Design for Interaction methods. I believe these complemented each other well. The combination of a systems perspective with an investigation into the individual user's context helped in understanding the complete picture. When I was in 'SC mode', I would be more focused on grounding my findings in theory and looking for a strong rationale for my steps, whereas in 'DfI mode' I would be more likely to let intuition and creativity lead the way for instance in creating the design fiction and utopian vision.

At times, however, it was hard to 'play' both roles. Therefore, I can imagine that the benefits of such an interdisciplinary research would be best reaped in an interdisciplinary team, where each approach can be defended and encouraged by a different person.

18.2. Implications of the research

I will discuss the implications of my research by posing three statements and discussing these.

I. To truly achieve algorithmic accountability, we need both top-down regulations and bottom-up solutions

As I stated in the previous paragraph, my approach to algorithmic accountability was focused on establishing accountability towards the user. In choosing to include Bovens' aspects of accountability, I also decided to design this accountability as a mechanism. I don't see this process as the only or best way of achieving algorithmic accountability. As Veale & Binns (2017) say, "we cannot expect to find simple or universal panaceas. We are stuck with layered, messy techniques to define, resolve and manage these complex challenges."

Bottom-up solutions

I regard the guidelines that my conclusion offered and the way in which accountability was designed into the V2G system as a bottom-up solution. If algorithmic accountability is to be established, then it needs to be clear how to do this. How do I imagine these guidelines being used? Ideally, by the creators of this systems themselves, perhaps in a co-creation setting with their future users. For this, the guidelines will need to be developed into a tool that can easily be used. Also, in the theoretical evaluation of the design, it was determined that the solutions didn't yet fully address accountability. By developing this further, the guidelines might be extended to form a more complete picture.

Also, right now it might cost a lot of effort to think of these accountability mechanisms. But once there are more examples, this might also lead to an easier adaptation. Sarah Gold, a designer invested in privacy and ethics in IoT, phrases this as follows in the video 'ethics for design': "Designers are good at copying, but when it comes to ethics I don't think we have any blueprints to copy" (Roussilhe, n.d.). I believe that in designing for accountability, the situation is similar.

So what might be other bottom-up solutions? Abdul, Vermeulen, Wang, Lim, & Kankanhalli (2018) present an overview of the developments concerning explainable, accountable and intelligible algorithms, based on an extensive literature analysis. In discussing the research on fair, transparent and accountable algorithms they point out that while researchers in machine learning communities are working to make their algorithms explainable, there tends to be a 'lack of usability, practical interpretability and efficacy on real users'. According to Abdul et al. (2018), this is where the HCI community might step in to develop technology that actually empowers the people that use it.

Veale, Binns & Van Kleek (2018) also explore areas in which HCI can contribute to achieving what they call 'GDPR-compliant machine learning' in relation to algorithmic fairness and accountability. By designing new ways of transparency, of receiving meaningful human input, and of communicating the consequences that automated decision-making might have, data subjects can be supported in exercising the rights that the GDPR presents them.

Top-down regulations

Nonetheless, for these accountability mechanisms to actually be implemented, there is likely to be a need for external forces or top-down governance mechanisms to stimulate this. Saurwein, Just & Latzer (2015) discuss various options. A company could voluntary make changes to an algorithmic system to protect reputation, or a new company might offer technical self-help solutions that protect a consumer from bias and violation of privacy. Self-organisation could include company principles and standards, while collective self-regulation would refer to standards across an industry, such as the technical standardisation of a 'do-not-track' option in web browser providers. State intervention would then be the provision of public services and the instalment of command-and-control regulations for manipulation or violations of privacy and copyright or freedom of expression.

In conclusion, top-down mechanisms such as regulations might be required to force companies to take responsibility in designing transparent and accountable algorithms. But at the same time, bottom-up solutions need to be facilitate the actual implementation of these accountable algorithms.

II. Designers should understand both the potential and the risks of designing for solutions that are based on algorithmic decision-making

When I reflect now on my own experiences in designing 'smart solutions' previous to doing this research project, I realise now how the focus was mostly on the possibilities it provided. Of course, my project team and I would also make considerations of how the technology wouldn't be experienced as 'creepy' or too intrusive. However, we didn't think of whether it might obstruct certain values and needs of the user.

Making these considerations should become more integrated into the approach of designing for algorithmic systems, or Internet of Things applications. I think that anticipating the possible undesired of algorithmic decision-making is something that designers can be quite good at, considering they are also good at imagining possible positive futures. As anticipation is also one of the four conceptual dimensions of Responsible Research and Innovation (Burget, Bardone, &

> *'To reap the advantages and avoid the risks, we must talk about the negative sides of algorithms'*
> Michal Kosinski

Pedaste, 2017), current design methods might benefit from the tools that have been developed in this field.

Right now, predicting the development of algorithms might still be difficult but it will get easier as there are more examples. Designers should familiarise themselves with these examples.

III. Public organisations should set the example of how algorithms and accountability should be implemented.

Would these guidelines that I developed actually be used by practitioners? I believe that there is an increasing pressure from the public for organisations to implement Big Data analyses and algorithmic decision-making in a fair, inclusive and transparent way. In the public debate, people are getting aware of the need for governance over the algorithmic systems that are implemented. The topic of algorithms in our society is increasingly being addressed, through various media.

- Tegenlicht, a Dutch documentary series produced an episode titled 'Verslaafd aan het algoritme' ('Addicted to the Algorithm'), in which the threats and risks of algorithmic decisions are shown in various contexts. As Michal Kosinski, psychologist and data scientist at Stanford University states: 'to reap the advantages and avoid the risks, we must talk about the negative sides of algorithms' (Kieft, 2018).

- Another example is Media lab SETUP. They made a pledge to create "Civil Weapons of Math Retaliation" in response to Cathy O'Neill's book on algorithmic models titled 'Weapons of Math Destruction." Through artworks and design, they intend to start a public debate concerning the question of how much and what kind of decision-making we are prepared to accept from algorithms (van der Ster, 2018).

- A similar initiative is TADA city in Amsterdam, a manifest by citizens and companies, that aims to contribute to a responsible digitalisation of the city, based on pillars such as inclusion, legitimacy and a human dimension (Tada, n.d.).

- Lastly, the Rathenau institute published a report about securing public values in a digital society, through which the government, industry and civil society are encouraged to improve the governance of digital technologies.

Based on these developments, I would say that non-transparent and unaccountable algorithms are becoming less accepted by the public. As a result, the urgency to public organisations like Alliander will be higher to instead develop algorithmic decision-making responsibly. Here is a chance for these organisations to set an example. This might be challenging because public organisations might not have the expertise of high-tech companies like Facebook and Google. This highlights the importance of researches into how accountability can be achieved.

By setting these examples, people will develop higher standards in regards to privacy, transparency and accountability, and commercial companies will have to follow their example. Right now, high-tech companies driven by revenues are setting the standard and people don't know better than to accept this. We don't think twice about the fact that we can't know how our Facebook News Feed works, but I hope that in 10 years we will have developed a different, more critical attitude.

18.3. Recommendations for further research

I will summarise and extend the recommendations for further research in this paragraph.

As stated previously, this research is a first exploration into design for algorithmic accountability and it provides a lot of openings to investigate further.

In this research, I've come across many crossroads, and that means there are still many other possible paths to pursue. One example of such a crossroad is the causal loop diagram. I identified three critical nodes – and there are probably more in there. Choosing a different critical node will lead to a whole different process and other results. For example, focusing on an ethical certificate could open up the discussion between stakeholders, about what standard would be feasible and how it would be communicated. Focusing on the critical node of the legibility of the terms of agreement could result in solutions that translate the inscrutable terms.

Furthermore, it would be interesting to explore other ideas that came up in the process. I chose to develop a new design for the interface for V2G cars, but other possible ideas were a separate platform, or a physical object. By developing these designs and comparing them with each other and with the current design, new insights could be generated about what design is more effective in accomplishing its purpose.

Lastly, I consider the current approach a bottom-up way of achieving algorithmic accountability as opposed to the top-down approach of the GDPR. It could be interesting to identify the approaches in between and explore how they can complement each other. Right now, the guidelines are like the result of a first design iteration. They require being applied in different contexts, preferably in collaboration with the stakeholders of a project like V2G, not just from the outside. In this way, the guidelines could be further shaped and polished to provide a more extensive framework for designing algorithmic accountability.

18.4. Outlook

So what's next? First of all, the guidelines that I developed should be applied in different contexts in order to gather more examples of accountability mechanisms and improve and extend these guidelines at the same time. As accountable algorithms become the rule instead of the exception, this should raise the standard for companies that implement algorithmic decision-making. For further encouragement, regulations such as the GDPR should be defined to set more specific boundaries.

Why go through all this effort? So we can enjoy the benefits that algorithms provide us. Throughout this project, for instance, I have relied on algorithms to find me relevant news articles when Googling, to provide me with accurate translations from Dutch to English, and to supply me with playlists of 'maximum concentration' piano music on Spotify. Algorithms can extend the possibilities we have, as long as they don't forget to pass the ball.

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, 1–18. http://doi.org/10.1145/3173574.3174156

Amsterdam SmartCity. (2018). Vehicle to Grid onder de loep: gebruiken we de elektrische auto straks ook als batterij voor groene stroom? Retrieved September 2, 2018, from https://amsterdamsmartcity.com/p/vehicle-to-grid

Ananny, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 146144481667664. http://doi.org/10.1177/1461444816676645

Bala, B. K., Arshad, F. M., & Noh, K. M. (2017). Causal Loop Diagrams. In System Dynamics (pp. 37–51). Singapore: Springer. http://doi.org/https://doi.org/10.1007/978-981-10-2045-2_3

Binns, R. (2017). Algorithmic Accountability and Public Reason. Philosophy & Technology. http://doi.org/10.1007/s13347-017-0263-5

Bosch, T. (2012). Sci-Fi Writer Bruce Sterling Explains the Intriguing New Concept of Design Fiction. Retrieved May 7, 2018, from http://www.slate.com/blogs/future_tense/2012/03/02/bruce_sterling_on_design_fictions_.html

Bovens, M. (2006). Analysing and Asessing Public Accountability: A conceptual framework. European Governance Papers (EUROGOV), (C-06-01), 1–37. Retrieved from http://www.connex-network.org/eurogov/pdf/egp-connex-C-06-01.pdf.

Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a Mechanism. West European Politics, 33(5), 946–967. http://doi.org/10.1080/01402382.2010.486119

Broeders, D., Schrijvers, E., van der Sloot, B., van Brakel, R., de Hoog, J., & Hirsch Ballin, E. (2017). Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. Computer Law and Security Review, 33(3), 309–323. http://doi.org/10.1016/j.clsr.2017.03.002

Bryman, A. (2012). Social Research Methods (4th ed.). New York: Oxford University Press.

Burger, C. (2016). Google DeepMind's AlphaGo: How it works. Retrieved August 9, 2018, from https://www.tastehit.com/blog/google-deepmind-alphago-how-it-works/

Burget, M., Bardone, E., & Pedaste, M. (2017). Definitions and Conceptual Dimensions of Responsible Research and Innovation: A Literature Review. Science and Engineering Ethics, 23(1), 1–19. http://doi.org/10.1007/s11948-016-9782-1

Caplan, R., & Reed, L. (2016). Who controls the public sphere in an era of algorithms: Case studies. Data & Society, 10.

City-zen. (n.d.). City-zen. New urban energy. Retrieved April 24, 2018, from http://www.cityzen-smartcity.eu/nl/home-nl/

Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. Philosophy and Technology, 29(3), 245–268. http://doi.org/10.1007/s13347-015-0211-1

Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62. http://doi.org/10.1145/2844110

Diakopoulos, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., … Zevenberge, B. (2017). Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. Retrieved February 6, 2018, from http://www.fatml.org/resources/principles-for-accountable-algorithms

Edelson, D. C. (2002). Design Research: What We Learn When We Engage in Design. The Journal of the Learning Sciences, 11(1), 105–121. http://doi.org/10.1207/ S15327809JLS1101_4

Eslami, M. (2017). Understanding and Designing around Users' Interaction with Hidden Algorithms in Sociotechnical Systems. Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion, 57–60. http://doi.org/10.1145/3022198.3024947

Eslami, M., Vaccaro, K., Karahalios, K., & Hamilton, K. (2017). "Be careful; things can be worse than they appear": Understanding Biased Algorithms and Users' Behavior around Them in Rating Platforms. Icwsm 2017, (Icwsm), 62–71. Retrieved from https://pdfs.semanticscholar.org/4ad3/4d39396d8bf8752fd-6fd96eb53c033e67331.pdf?_ga=2.161737276.1247488861.1504751314-783989974.1504751314

Geum, Y., & Park, Y. (2011). Designing the sustainable product-service integration: A product-service blueprint approach. Journal of Cleaner Production, 19(14), 1601–1614. http://doi.org/10.1016/j.jclepro.2011.05.017

Hollander, G. de, Vonk, M., Snellen, D., & Huitzing, H. (2017). Mobiliteit en elektriciteit in het digitale tijdperk. Publieke waarden onder spanning. Den Haag: Mobiliteit en elektriciteit in het digitale tijdperk. Publieke waarden onder spanning.

Karanasiou, A. P., & Pinotsis, D. A. (2017). A study into the layers of automated decision-making: emergent normative and legal aspects of deep learning. International Review of Law, Computers and Technology, 31(2), 170–187. http://doi.org/10.1080/13600869.2017.1298499

Kieft, M. (2018). Verslaafd aan het algoritme. The Netherlands: VPRO Tegenlicht. Retrieved from https://www.vpro.nl/programmas/tegenlicht/kijk/afleveringen/2017-2018/verslaafd-aan-het-algoritme.html

Kitchin, R. (2017). Thinking critically about and researching algorithms. Information Communication and Society, 20(1), 14–29. http://doi.org/10.1080/1369118X.2016.1154087

Kizilcec, R. F. (2016). How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, 2390–2395. http://doi.org/10.1145/2858036.2858402

Knight, W. (2017). The dark secret at the heart of AI. Retrieved August 12, 2018, from https://www.technology-review.com/s/604087/the-dark-secret-at-the-heart-of-ai/

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2017). Fair, Transparent, and Accountable Algorithmic Decision-making Processes. Philosophy & Technology. http://doi.org/10.1007/s13347-017-0279-x

Lindberg, S. I. (2013). Mapping accountability: Core concept and subtypes. International Review of Administrative Sciences, 79(2), 202–226. http://doi.org/10.1177/0020852313477761

Mulgan, R. (2000). "Accountability": An ever expanding Concept? Public Administraton, 78(no 3), 555–573. http://doi.org/10.1111/1467-9299.00218

Neyland, D. (2016). Bearing Account-able Witness to the Ethical Algorithmic System. Science Technology and Human Values, 41(1), 50–76. http://doi.org/10.1177/0162243915598056

NVivo qualitative data analysis Software. (2017). QSR Internation Pty Ltd.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. [MOBI], London: Penguin Books Ltd.

Oosterlaken, I. (2012). The Capability Approach, Technology and Design: Taking Stock and Looking Ahead BT - The Capability Approach, Technology and Design. In I. Oosterlaken & J. van den Hoven (Eds.) (pp. 3–26). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-3879-9_1

Roussilhe, G. (n.d.). Ethics for design. Retrieved from http://www.ethicsfordesign.com/

Saurwein, F., Just, N., & Latzer, M. (2015). Governance of algorithms: options and limitations. Info, 17(6), 35–49. http://doi.org/10.1108/info-05-2015-0025

Smart Data City. (n.d.). Stratumseind Living Lab. Retrieved January 23, 2018, from http://www.smartdatacity.org/stratumseind-living-lab/

Snowden, D. (2002). Complex acts of knowing: Paradox and descriptive self-awareness. Journal of Knowledge Management, 6(2), 100–111. http://doi.org/10.1108/13673270210424639

Spielkamp, M. (2017). Transparent Data Mining for Big and Small Data, 32, 207–215. http://doi.org/10.1007/978-3-319-54024-5

Stappers, P. J. (2012). Teaching Principles of Qualitative Analysis to Industrial Design Engineers. DS 74: Proceedings of the 14th International Conference on Engineering & Product Design Education (E&PDE12) Design Education for Future Wellbeing, Antwerp, B(September), 109–114.

Sullivan, D. (2015). How Machine Learning Works, As Explained By Google. Retrieved August 12, 2018, from https://martechtoday.com/how-machine-learning-works-150366

Tada. (n.d.). Tada - duidelijk over data. Retrieved August 13, 2018, from https://tada.city/

Tanenbaum, J. (2014). Design fictional interactions. Interactions, 21(5), 22–23. http://doi.org/10.1145/2648414

van Boeijen, A. G. C., Daalhuizen, J. J., Zijlstra, J. J. M., & van der Schoor, R. S. A. (2013). Delft Design Guide. Amsterdam: BIS Publishers.

van der Sanden, M. C., & de Vries, M. J. (2016). Innovation in science and technology education and communication through design thinking. In Science and Technology Education and Communication (pp. 129–145). SensePublishers, Rotterdam. http://doi.org/https://doi.org/10.1007/978-94-6300-738-2_8

van der Ster, J. (2018). Civil weapons of math retaliation. Retrieved August 13, 2018, from https://www.setup.nl/magazine/2018/01/civil-weapons-math-retaliation

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. Big Data & Society, 4(2), 205395171774353. http://doi.org/10.1177/2053951717743530

Veale, M., Binns, R., & Van Kleek, M. (2018). Some HCI Priorities for GDPR-Compliant Machine Learning, (April 2018). http://doi.org/10.1145/2872427.2883028

Vedder, A., & Naudts, L. (2017). Accountability for the use of algorithms in a big data environment. International Review of Law, Computers and Technology, 31(2), 206–224. http://doi.org/10.1080/13600869.2017.1298547

Visser, F. S., Stappers, P. J., van der Lugt, R., & Sanders, E. B.-N. (2005). Contextmapping: experiences from practice. CoDesign, 1(2), 119–149. http://doi.org/10.1080/15710880500135987

Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. Science Technology and Human Values, 41(1), 118–132. http://doi.org/10.1177/0162243915605575

Žliobaitė, I. (2017). Measuring discrimination in algorithmic decision making. Data Mining and Knowledge Discovery, 31(4), 1060–1089. http://doi.org/10.1007/s10618-017-0506-1

Zouave, E. T., & Marquenie, T. (2017). An Inconvenient Truth: Algorithmic Transparency & Accountability in Criminal Intelligence Profiling. 2017 European Intelligence and Security Informatics Conference (EISIC), 17–23. http://doi.org/10.1109/EISIC.2017.12