Curbing Algorithmic Kafka

Embedding design processes of public algorithmic systems in a democracy and Rule of Law context

Nouws, S.J.J.

DOI
10.4233/uuid:9477b6ce-3d7e-4e4e-82b4-43fd98b2a5fb

Publication date
2026

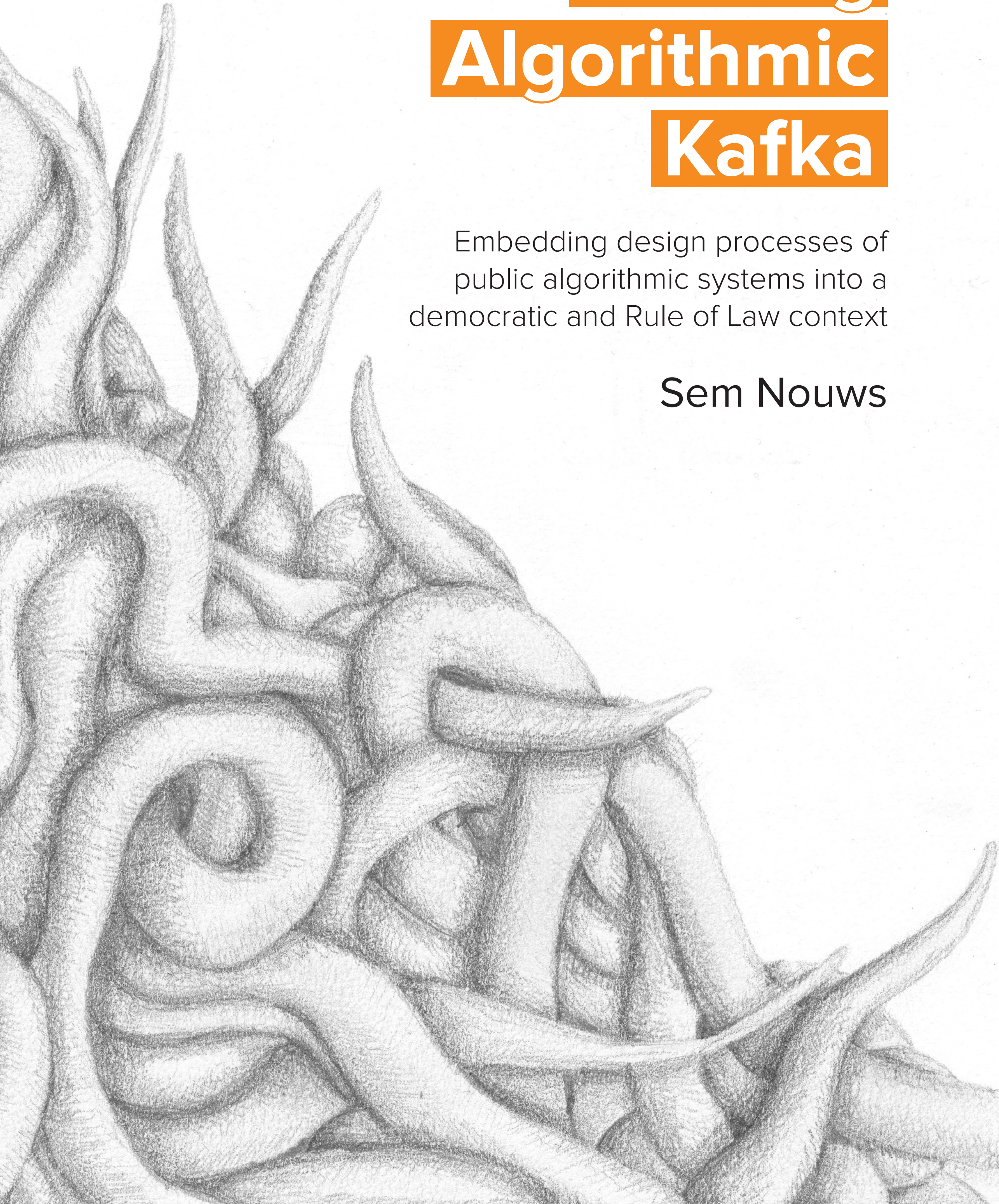
Document Version
Final published version

Citation (APA)
Nouws, S. J. J. (2026). *Curbing Algorithmic Kafka: Embedding design processes of public algorithmic systems in a democracy and Rule of Law context*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:9477b6ce-3d7e-4e4e-82b4-43fd98b2a5fb

Curbing Algorithmic Kafka

Embedding design processes of public algorithmic systems into a democratic and Rule of Law context

Sem Nouws



Curbing Algorithmic Kafka

Embedding design processes of public algorithmic systems
into a democratic and Rule of Law context

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Wednesday 14 January 2026 at 15:00 o'clock

by

Sem Johannes Joseph NOUWS
Master of Science in Complex Systems Engineering, Delft University of Technology,
The Netherlands
Born in Breda, The Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof.dr.ir. M.F.W.H.A. Janssen	Delft University of Technology, promotor
Dr.ir. R.I.J. Dobbe	Delft University of Technology, copromotor

Independent members:

Prof.dr.mr.ir. N. Doorn	Delft University of Technology
Prof.dr. A.J. Meijer	Utrecht University
Prof.mr.dr. H.S. Taekema	Erasmus University Rotterdam
Prof.mr.dr. M. Veale	University College London
Prof.dr.ir. N. Bharosa	Delft University of Technology

Cover art: Sem Nouws
Layout: Sem Nouws & Don van Baar
Printed by: Gildeprint

ISBN 978-94-6384-890-9 (print)
ISBN 978-94-6518-217-9 (digital)
An electronic version of this dissertation is available at http://repository.tudelft.nl

Voor Oma Nouws

Propositions accompanying the dissertation

Curbing Algorithmic Kafka

by **Sem Nouws**

1. As long as public servants do not consider themselves socio-technical designers, they underestimate their influence on the design of public algorithmic systems (this thesis, part I Diagnosis)
2. Design processes of public algorithmic systems that lack a democratic and Rule of Law basis will produce algorithmic systems that are undemocratic and that mediate arbitrary use of power (this thesis, part II Appraisal)
3. Despite being intangible and intractable, institutional interventions are the only means for public organisations to transform their design practices (this thesis, part III Create and Assess)
4. A call for AI ethics opens up opportunities for politicians to stifle political debates on AI systems
5. The occupation with 'new' and 'disruptive' technologies distracts scholars from doing more relevant, but less 'sexy' research on the structuration of human behaviour by everyday technology
6. Designing a design science approach to formulate a design theory on design processes will fail when the researcher lacks reflectivity and responsiveness
7. Using AI for coding data is equivalent to involving an opportunistic and negligent interpretive researcher into your research project
8. The practice of providing recommendations at the very end of research papers invites authors to make superficial, trivial, and trendy claims outside their expertise
9. The use of English terms in Dutch-speaking workplaces points out an intellectual void
10. Square-shaped vehicles for communication are barriers to meaningful deliberation

These propositions are regarded as opposable and defendable, and have been approved as such by the promotors Prof.dr.ir. M.F.W.H.A. Janssen and Dr.ir. R.I.J. Dobbe.

I’ve looked at clouds from both sides now
From up and down, and still somehow
It’s clouds illusions I recall
I really don’t know clouds at all

– Joni Mitchell; Both Sides Now

That the people have the power
To redeem the work of fools
Upon the meek the graces shower
It’s decreed: the people rule

– Patti Smith; People Have the Power

Table of contents

Preface

All language is but a poor translation
– Franz Kafka

Riding my bike from my hometown Rotterdam to work in Delft, I often listened to the podcast *Betrouwbare Bronnen*. This podcast on politics and political history was a great inspiration for my PhD. In their preview on the year 2024, the presenters pointed out the upcoming centenary of Franz Kafka's death and referred to their episode dedicated to the writer's life and work. Relistening their excellent episode on Franz Kafka convinced me of using the writer's work as metaphor for problems in current public algorithmic systems.

The episode was a reaction to the use of 'Kafkaesque' to characterise the *Toeslagenaffaire* (childcare allowances scandal in which parents were falsely accused of fraud by a public algorithmic system) by Dutch members of parliament. The podcast's presenters pointed out that the metaphor associated to Kafka's name is often narrowly focused on harms inflicted on citizens by a cold-blooded government. But his vast body of work paints a much richer picture of individual citizens or persons faced with an anonymous apparatus. Kafka's work starts from the perspective of individual persons and, therefore, is an appropriate, adequate and meaningful way to show the impact of systems or technologies on individuals. Moreover, while Kafka is often associated with critique on bureaucratic organisations with incomprehensible and inconsistent rules, the writer also indicated the role of technology in alienating individuals. As such, the qualification 'Kafkaesque' relates to several aspects resulting in the detriments of the *Toeslagenaffaire* and comparable cases in other national contexts.

A government has a duty to protect its citizens from Kafkaesque situations caused by its own actions. This is a fundamental pillar in the *democratische rechtsstaat* – the organisation of a state based on principles of democracy and the Rule of Law. The concept of *democratische rechtsstaat* brings me to another author that influenced my research on the design process of public algorithmic systems considerably. Before starting my PhD, I read *Groter denken, Kleiner doen* by Herman Tjeenk Willink, a Dutch minister of state. The book eloquently delineates the deterioration of the *democratische rechtsstaat* because of negligence by the actors that should preserve the fundamental institutions associated with democracy and the Rule of Law.

So, we are confronted with harmful, Kafkaesque systems, but the democratic and Rule of Law mechanisms at our disposal to address Kafkaesque situations are failing. Therefore, this research focuses on strengthening the *democratische rechtsstaat*. I purposefully chose to scope the research on an aspect that is often overlooked in research on democracy, the Rule of Law and algorithmic systems: the design process of algorithmic systems. As such, I hope to contribute situational knowledge to the often-theoretical ideas concerning the *democratische rechtsstaat* and provide public organisations with useful and meaningful interventions in their own design processes. All this to prevent, mitigate and correct Kafkaesque situations in public algorithmic systems.

Considering that Franz Kafka already wrote about citizens cornered by bureaucratic and technological apparatuses more than a hundred years ago, I do not want to claim that this thesis will solve all harmful public algorithmic systems. At best, it can support in reducing the emergence of Kafkaesque situations. As implied by the thesis' title, this research produces interventions that support public organisations in curbing algorithmic Kafka.

4 April 2025

Research statement

The data management plans, informed consent forms, and human research ethics application related to the studies involving human participants presented in this thesis were approved by the assigned bodies of Delft University of Technology. Interview transcripts and fieldnotes are stored at secure servers of Delft University of Technology. Following from the consent forms signed by interviewees, this data cannot be shared on a data repository. The research data can be requested until November 2027. The request should state a clear goal for gaining access to the data and can be send to Sem Nouws (s.j.j.nouws@tudelft.nl), Roel Dobbe (r.i.j.dobbe@tudelft.nl) or Marijn Janssen (m.f.w.h.a.janssen@tudelft.nl). All transcripts and fieldnotes have been pseudonymised by replacing personal and organisation-specific information by tokens. In addition, all quotes presented in this thesis have been translated from Dutch into English. Interview recordings have been deleted at the end of this PhD project.

This thesis was written without the assistance of large language models. Moreover, we did not use these models in conducting the research presented in this thesis.

Chapter 1
Introduction

1

> Er is een chronisch gebrek aan ruimte voor reflectie. Veel politici, bestuurders en ambtenaren zouden dat anders willen. Maar ze zitten gevangen in hun eigen rationaliteit, taal en tijd. Daaruit breken lukt alleen door tegenwicht, tegenspraak en tegenmacht; in het publieke debat, vanuit de burgersamenleving, door de professionals op de werkvloer en door de rechterlijke macht.[1]
>
> *– Herman Tjeenk Willink; Groter denken, kleiner doen, p. 73*

1.1 Public organisations designing public algorithmic systems

The role of public organisations in designing socio-technical systems requires more attention because of digitalisation and automation efforts within governments. Digital technologies such as algorithmic applications are core elements in the regulative and executive practices of public organisations. At the same time, algorithmic applications deployed in the execution of critical public services put citizens in Kafkaesque situations and inflict harm on these citizens. Such algorithmic Kafka is partly the result of the organisation of design processes of public algorithmic systems. Accordingly, public organisations have to redefine their role in designing these systems and develop their design competencies. This section elaborates on the link between shortcomings in design processes of algorithmic systems and the emergence of algorithmic Kafka.

1.1.1 Public organisations and algorithmic systems

Reinforcing the trend of automating and augmenting processes in public administration, public organisations are integrating data analytics, statistics, information systems, and algorithmic functionalities into the execution of their policies and the provision of their services (Veale & Brass, 2019). This trend is currently strengthened by the emergence of big data and artificial intelligence (AI) technologies, such as machine learning (ML) and large language models (LLM). Broadly defined, these technologies are *algorithmic applications* in which automated

1 There is a chronic lack of room for reflection. Many politicians, administrators and public servants would like to see this changed. But they are caught in their own rationality, language and time. Breaking out of that situation is only possible through counterbalance, dissent and countervailing power; in public debate, from civil society, by professionals on the ground, and by the judiciary.

rule-based and/or data-driven algorithms encoded in software-based technologies predict or generate output. The promise of facilitating fraud prediction, risk assessment, and allocation of public resources generates interest in these algorithmic applications among public organisations (Dencik et al., 2019; Beer, 2017). Algorithmic applications come with a promise of increased efficiency, which makes them instrumental, through financial cuts in public services, in achieving austerity goals of public organisations (Dencik et al., 2019).

Consequently, public organisations increasingly establish *algorithmic practices* for the execution of public policies, laws and regulations. Algorithmic practices refer to organisational and human practices in which the enactment of algorithmic applications has become an inextricable part. Considering algorithmic applications from a practice perspective expands the narrow focus on the technology to accentuate the role algorithmic applications play in politics, power relations and decision-making (Beer, 2017). In public administration, algorithmic practices in public organisations are reinforcing, transforming and challenging 'established institutions and administrative practices' (Veale & Brass, 2019, p. 142).

A characteristic reshaping of practices is related to the efficiency gains associated with using algorithmic applications. Algorithmic applications formalise societal phenomena (Newman et al., 2022) and, therefore, pertain to the standardised and impersonal decision-making characteristic of coercive formalism in bureaucratic organisations (Adler & Borys, 1996; Peeters & Schuilenburg, 2018). With a focus on efficiency and rationality inherent to the bureaucratic ideal type, public organisations shift towards *technocratic governance*. Janssen and Kuk (2016, pp. 371–372) observe that this shift is based on the assumption 'that complex societal problems can be deconstructed into neatly defined, structured and well-scoped problems that can be solved algorithmically and in which political realities play no role.' The inflexibility of code (which underlies the algorithms) takes over the deliberative practice of a legal system based on speech and text that enables front-line workers to apply rules and regulations to specific and contextual individual circumstances (Hildebrandt, 2019). The rigidity of algorithmic applications restructures work practices and compromises the discretion of front-line workers (Alkhatib & Bernstein, 2019; Giest & Klievink, 2024). Their discretion is shifted to the application's designers, who have gained influence over what decisions are made in particular circumstances (Bovens & Zouridis, 2002; Zouridis et al., 2020).

Instead of increasing efficiency, rigid algorithmic practices emerge that 'capture' the complexity of societal problems and human behaviour in algorithms and computer code. Accordingly, algorithmic applications create their own reality that does not cover the actual situation that citizens are in (Agre, 1994; Alkhatib, 2021; Yeung, 2018). For example, studies show that algorithmic decision-making is often discriminatory (e.g., Eubanks, 2017), can come to incorrect decisions (e.g., Peeters & Widlak, 2018), can reduce agency of both citizens and users (e.g., Peeters, 2020; Wagner, 2019), and can be inscrutable (e.g., Burrell, 2016; Kroll, 2018). Consequently, algorithmic practices challenge the protection of citizens from misconduct by governments. The position of citizens in relation to governments has deteriorated because automation of government has become more intrusive, more rigid, and more obscure (Alkhatib, 2021; Dencik et al., 2019). We refer to this deteriorated position of citizens as algorithmic Kafka.

Algorithmic Kafka creates situations in which citizens are cornered in a digital bureaucratic system without knowing how to solve problems caused by those systems (Bayamlıoğlu & Leenes, 2018; De Laat, 2019; Ossewaarde, 2023; Susser, 2016). Such algorithmic Kafka can have a detrimental impact on citizens. Fraud prediction models are illustrative of algorithmic Kafka when they produce false accusations and erroneous decisions that result in citizen harm. In the Dutch childcare allowances scandal, false accusations of fraud have pushed a large group of citizens into debt, resulting in the tragic destruction of people's lives (Bouwmeester, 2023; Fenger & Simonse, 2024; Konaté & Pali, 2023; Peeters & Widlak, 2023). In Australia, the use of Robodebt – a data-matching algorithm that automated debt detection among welfare recipients – was discontinued after a Federal Court ruling that followed years of criticism on the algorithm by, amongst others, victims and national inquiries (Braithwaite, 2020; Galloway, 2017; Rinta-Kahila et al., 2022; Whiteford, 2021). In both cases, the affected citizens were cornered and had very little possibilities to find recourse.

Algorithmic Kafka is contrary to the principles of democracy and the Rule of Law. Democracy guarantees the freedom of citizens through self-government and ensures equal participation in decision-making (Christiano, 1996; Cunningham, 2002). The Rule of Law protects citizens from arbitrary use of power by the state through making law the mode to rule people (instead of being ruled by, an elite of, man) (Krygier, 2009; Raz, 1979; Waldron, 2002). However, the protection of citizens through democracy and the Rule of Law is violated by algorithmic Kafka. Similarly, authors have argued that algorithmic practices can be a threat to democratic and Rule of Law values (cf. Bayamlıoğlu & Leenes, 2018; Greenstein, 2022; Grimmelikhuijsen & Meijer, 2022; Hildebrandt, 2018; König & Wenzelburger, 2020).

To curb algorithmic Kafka, public organisations should ensure that algorithmic systems are aligned with principles of democracy and the Rule of Law. This requires public organisations to ensure legitimate and safe integration of algorithmic applications into their organisational practices. Most research focuses on the algorithmic practices that have already been implemented and deployed. More specifically, this research focuses on how these practices change the role of public organisations (e.g., Dencik et al., 2019; Redden et al., 2020; Veale & Brass, 2019), how public organisations use algorithmic applications (e.g., Kolkman, 2020; Lorenz et al., 2022; Giest & Klievink, 2024), and what the possible merits and challenges of algorithmic applications are for public organisations (e.g., Pencheva et al., 2020; Straub et al., 2023). While it is important to study the use of algorithmic applications in public organisations, this undervalues the role of design activities in shaping algorithmic systems. The research presented in this thesis examines the role of designing algorithmic systems in creating algorithmic Kafka in order to identify design practices in which harms inflicted on citizens by algorithmic systems are prevented, mitigated, or corrected.

1.1.2 Designing public algorithmic systems

This thesis focuses on *designing public algorithmic systems* and provides an overview of definitions that are elaborated in Chapter 2. We use the term designing to signify design-as-a-verb, since we concentrate on the interactions between designers of public algorithmic systems and how these interactions result in design choices. A public algorithmic system is a socio-technical system with the purpose of automating, supporting or augmenting public administration practices. Being a socio-technical system, algorithmic systems consist of three types of components: technology, institutions and human agents (Orlikowski, 1992). The distinguishing artefact of the socio-technical system in this thesis is the algorithmic application.

Designing a socio-technical system comprises the creation of both institutional as well as technological artefacts, and establishing interactions between these artefacts (Koppenjan

& Groenewegen, 2005). This research focuses on a specific part of the design process of public algorithmic systems: the formulation of its *socio-technical specification*. The socio-technical specification describes the technological artefacts, institutional artefacts, and human agents – including the interactions between those components – that constitute the public algorithmic system. In other words, it specifies the form and function of each system component, who the system needs to serve, and the system's purpose (Dobbe et al., 2021). The specification is formulated through *design practices* – i.e., patterns and routines in the interactions between designers. Everyone who performs *design activities* is a designer (Simon, 1969/1996). Design activities are all creative actions that affect or result in design choices about the socio-technical specification (Bots, 2007; Van Aken, 2005b).

The formulation of a socio-technical specification is the moment in a system's life cycle in which dedicated designers have considerable influence on the form and function of the algorithmic systems. They can formulate alternative specifications, deliberate on the consequences of design choices, and create a comprehensive overview of the system that will be created. As soon as the system is implemented and deployed in the organisation, users such as operators will also affect the form and function of the system, for example, by adapting the meaning designers have given to an artefact (cf. Kroes et al., 2006). Users in their role of designer are outside the scope of this thesis. Accordingly, we consider actors involved in or responsible for making design choices, that influence at least one socio-technical component of public algorithmic systems, without being directly affected by the system as *designers*.

The process of designing socio-technical systems can be improved by establishing a process design (Koppenjan & Groenewegen, 2005). More importantly, the structure and enactment of the design practices within these processes influence the form and function of the output of these practices (Simon, 1969/1996). Thus, the approach taken in the design process will eventually be observable in the outcome of that process. In other words, harms inflicted on citizens by Kafkaesque algorithmic systems can partly be related to how the design process is organised.

Consequently, the design process of public algorithmic systems should be organised following principles that counter algorithmic Kafka. This thesis focuses on principles that guarantee the freedom of citizens and concerns the protection of citizens against misconduct by government: democracy and the Rule of Law. The literature suggests that current design processes are contrary to these principles (as are the algorithmic practices that follow from them, see Section 1.1.1). The next section provides a preliminary overview of that literature. Chapter 4 will take a closer look at existing design practices and elaborate on the problems in those practices.

1.1.3 Public organisations as naive designers

Hence, the structure of design practices affects the algorithmic systems that result from design processes, and, therefore, eventually shapes algorithmic practices. However, the scant literature available on design processes of algorithmic systems provides hunches that these processes are rudimentary. Public organisations do not seem to be able to anticipate or react to algorithmic systems that are potentially harmful to citizens. This section discusses how a lack of democratic legitimacy in the design process, inadequate instruments, and changes in the organisation of design practices are related to the institutional void that design processes are in.

First, a lack of democratic legitimacy of algorithmic practices is observed in the processes used to create and establish algorithmic practices (Grimmelikhuijsen & Meijer, 2022). Van Zoonen (2020) discusses the emergence of undesirable practices in data-driven social policy in the Netherlands, such as the lack of a public debate and a lack of direct involvement of citizens in the design processes of such policies. In the United States, public organisations tend to focus on technical fixes, trump certain singular values over other values, lack proper tools to perform the act of designing, and shift towards 'processes that make technological choices appear inevitable and apolitical' in the design processes of public algorithmic systems (Mulligan & Bamberger, 2018). In response to this democratic deficit, authors recommend to institutionalise citizen participation or other democratic mechanisms in the design process (Mulligan & Bamberger, 2019; Van Zoonen, 2020).

Second, the instruments that public organisations develop and implement to alter their design processes – e.g., algorithm registers, watchdogs, or impact assessments – are inadequate in protecting citizens from harmful algorithmic systems. Public organisations explore such instruments in response to the call for more control over erroneous automated and algorithmic decision-making by authorities and NGOs. Although these instruments potentially can strengthen the detection, mitigation, and correction of harmful algorithmic systems, their translation into practice turns out to be hard to accomplish. Green (2022) examines 41, mostly national, policy documents on human oversight of algorithmic applications. He observes a lack of empirical evidence that supports these policies on human oversight, increasing the risk of legitimising 'flawed and unaccountable algorithms in government' (Green, 2022, p. 9). Others have studied the use of ethical and legal frameworks (a subset of policy instruments) by public organisations in practice (Fest et al., 2022; Siffels et al., 2022). Fest et al. (2022) identify the challenges public organisations face in translating legal frameworks into their daily practice. These frameworks are clear and meaningful on paper and have the potential to support developers in considering legal aspects. Nevertheless, the aims behind the frameworks are not achieved in practice since actors working with the frameworks mostly lack the time or capacity to use the frameworks as intended. Similarly, Siffels et al. (2022) show the difficulties of bringing ethical frameworks to practice. They observe a lack of expertise, impaired data literacy, and low ethical awareness in Dutch municipalities as important barriers to putting data ethics into practice. The instrumental focus of public organisations also indicates a lack of a normative idea behind protecting citizens from governmental misconduct. The Rule of Law provides a normative basis for organising such protection.

Third, the position of designers of algorithmic systems is changing in public algorithmic systems. As mentioned before, the position of (technical) designers has been strengthened by the shift of discretion from street-level bureaucrats to system-level bureaucrats (Bovens & Zouridis, 2002; Zouridis et al., 2020). However, at that system-level, technological- and policy-related responsibilities are separated in public organisations, which impedes a socio-technical approach towards public algorithmic systems (Van Der Voort et al., 2019; Lorenz, 2023; Selten & Klievink, 2024). At the same time, the design of algorithmic applications is often handed over to private or external parties through procurement (Mulligan & Bamberger, 2019). This dependency on external parties pushes the design process out of the public domain and, as a result, reinforces the lack of democratic legitimacy of the process. Accordingly, Mulligan & Bamberger (2019) argue that public organisations should critically consider what design choices cannot be outsourced but should be made through public deliberation.

The lack of democratic legitimacy, inadequate instruments, and shifts in the position of designers of algorithmic systems can be associated to an *institutional void* that design processes of public algorithmic systems are in (Van Zoonen, 2020; Oldenhof et al., 2024). In an institutional void, policymaking is performed while 'there are no clear rules and norms according to which politics is to be conducted and policy measures are to be agreed upon' (Hajer, 2003, p. 175). An institutional void does not mean that institutions – i.e., social rules that structure human behaviour (Hodgson, 2006) – are absent. Rather, it signifies that political processes of making trade-offs have not yet been institutionalised, formalised or translated to a specific practice. In such situations, rules and norms emerge along the way, often without explicit deliberation or reasoning behind those institutions. Van Zoonen (2020, p. 9) argues that 'the institutional void around using data for social policy goals, is … slowly filled by a set of practices that is undesirable from the perspective of the legal frameworks of the GDPR [EU regulation on data protection] and the social domain, the quality principles of responsible data use and general ethical standards.' As discussed in this section, the design practices that emerged in this institutional void are also undesirable or unfit for designing public algorithmic systems (Oldenhof et al., 2024; Van Zoonen, 2020). By underestimating the importance of institutionalising design processes of public algorithmic systems, public organisations can be qualified as ignorant designers.

1.1.4 Research aim

There is an urgent need to examine the institutional void in current design practices and address the undesirable and inadequate institutions that have emerged in this void. This research focuses on studying the institutions that currently structure design practices of public algorithmic systems. In addition to understanding the current situation, it prescribes *institutional interventions* that shift the focus of current design practices towards protecting citizens from harmful and Kafkaesque public algorithmic systems. Institutional interventions are deliberate changes to the institutional structure of design practices. The starting point for these interventions is to embed design processes of public algorithmic systems in established democratic and Rule of Law practices. Accordingly, the main aim of this research is to *create design practices that reduce the emergence of Kafkaesque situations in public algorithmic systems by stimulating designers to reflect on and respond to the consequences of their design choices.*

1.2 Research questions

This section presents three research questions that need to be answered in order to achieve the research aim.

Research question 1

Scientific literature lacks empirical insight into current design practices of public algorithmic systems and the initiatives that public organisations undertake to adapt their design practices (cf. Zuiderwijk et al., 2021). As discussed above, the empirical research available mostly focuses on the use or enactment of algorithmic applications, how deployment of such applications changes practices in public organisations, or how ideologies are ingrained in or re-enacted by using public algorithmic systems. Although using and designing algorithmic systems (and their effects) cannot be fully distinguished in practice, most researchers concentrate on developed systems that are operated. Moreover, empirical research on public algorithmic systems is based on small samples of public organisations or algorithmic applications, and many studies are based in the United States, limiting the generalisation of insights to design practices in other jurisdictions with their specific institutional context (e.g., Mulligan & Bamberger, 2018).

While there are theoretical frameworks concerning designing socio-technical systems (e.g., Baxter & Sommerville, 2011; Clegg, 2000; De Bruijn & Herder, 2009), a comprehensive understanding of the design process of currently deployed public algorithmic systems is missing. Moreover, the literature on socio-technical designing has not considered the relevant specificities of designing within the context of public organisations. In sum, the specific context of designing algorithmic systems in public organisations is missing in empirical research.

In response to this gap in scientific literature, this research maps current design practices of algorithmic systems within public organisations. To better understand the design practices that have emerged, this thesis also examines and reconstructs the presuppositions that underlie these practices. Presuppositions are ideas, ideologies, or institutional logics (implicitly) shared by actors within an organisation or community. These presuppositions shape organisational practices by structuring the behaviour of actors. We arrive at design practices and presuppositions by answering research question 1:

What presuppositions underlie the design practices for public algorithmic systems that have emerged in public organisations?

Research question 2

As discussed before, public organisations have shifted from street-level bureaucracies to system-level or infrastructure-level bureaucracies (Bovens & Zouridis, 2002; Widlak & Peeters, 2025). This shift has consequences for design practices of public algorithmic systems. System engineers have obtained a more central role in shaping government practices (Bovens & Zouridis, 2002). However, democratic control over system engineers is lacking (Zouridis et al., 2020). Similarly, Widlak & Peeters (2025) argue that infrastructural data exchange, inherent to public algorithmic systems, obscures and impedes democratic control over the procedural lawfulness in automated decision-making. These concerns are related to more general concerns about the lack of democratic legitimacy of algorithmic systems (see Grimmelikhuijsen & Meijer, 2022; König, 2020; König & Wenzelburger, 2021).

Accordingly, there is a need for studying how to guarantee democratic control and legitimacy in design processes of public algorithmic systems. The intersection of technology, design and democracy has been an area of academic interest for a longer time (e.g., Feenberg, 1999; Hajer, 1995; Sclove, 1995). The research on democracy and technology provides ideal types of democratically designing technology (Hajer, 1995; Sclove, 1995), or discusses philosophical reflections on the nature of a democratic society in which technology plays a prominent and significant role (Feenberg, 1999). In addition, scientific research has developed new design approaches, including participatory design, adversarial design, and speculative design (Ozkaramanli et al., 2022).

The combination of algorithms, AI and democracy has also been attracting considerable interest in the last few years. Scholars regularly call for democratising AI and its design process to address the intrusive and disruptive consequences of algorithmic systems (cf. algorithmic

Kafka). These calls range from democratising AI in a narrow sense (offering as many people as possible access to AI), ensuring participation and public involvement in the development and use of algorithmic systems, to the correction of power imbalances in algorithmic systems (Aizenberg & Van Den Hoven, 2020; Liu, 2018; Noorman & Swierstra, 2023). At the same time, literature on democratising AI is critiqued for being opportunistic, superficial, and having little engagement with political philosophy literature (Himmelreich, 2023; Noorman & Swierstra, 2023; Sætra et al., 2022; Sloane et al., 2022).

Like in the case of democracy, scholars express their concerns of the Rule of Law being under threat by public algorithmic systems (Bayamlıoğlu & Leenes, 2018). The Kafkaesque implications of algorithmic systems show that public organisations are not capable of protecting citizens from arbitrary conduct in these systems. This is, in essence, the very aim of the Rule of Law: organising a state or administration in such a way that citizens are protected from arbitrary use of power (Krygier, 2014). The Rule of Law can, therefore, also be a source of inspiration for institutionalizing mechanisms to protect citizens from harmful algorithmic systems. This requires a thorough understanding of how Rule of Law mechanisms are (re)structured by algorithmic applications or, more general, by technology. However, legal philosophy has disregarded the role of technology in engendering or mediating arbitrary conduct (Nouws & Dobbe, 2024). Consequently, it is unclear how the Rule of Law can play a role in organising design processes of public algorithmic systems.

In sum, researchers have identified democracy and the Rule of Law as important presuppositions to base algorithmic systems and their design processes on; especially when public organisations want to curb algorithmic Kafka. However, the synthesis of these three dissimilar presuppositions (established in different scientific disciplines with their own vocabulary and methods) has not yet been pursued by scholars. Considering this research' main goal – embedding design process of public algorithmic systems into a democracy and Rule of Law contexts –, we aim to elaborate on how socio-technical designing, democracy, and the Rule of Law can reinforce each other. In addition, we also explore the difficulties of synthesising these three presuppositions.

Where the first research question provides insight into the current situation of design practices, the second research question focuses on the desired situation. Accordingly, we pose research question 2:

What design practices that curb algorithmic Kafka are prescribed by the synthesis of the presuppositions of socio-technical designing, democracy, and the Rule of Law?

Research question 3

Where research question 2 explores the theoretical basis for the institutional interventions this research prescribes, research question 3 focuses on the translation of that theoretical basis to practice. We examine the institutional interventions needed to arrive at design practices that curb algorithmic Kafka and evaluate how these interventions play out in practice.

By focusing on institutional interventions that adapt the organisational structure of design practices, we adopt a different focal point compared to prior scientific inquiries in the field of governance of algorithms and AI. As discussed in Section 1.1.3, both scientific literature and practice have proposed a variety of initiatives to alter algorithmic practices. These initiatives comprise both technical instruments (e.g., debiasing tools and XAI) as well as policy instruments (e.g., impact assessments and governance frameworks). Nevertheless, public organisations still struggle with steering both their algorithmic practices as well as the associated design processes towards democratic and Rule of Law practices. Most of the time, the instruments do not urge public organisations to effectively challenge well-established but obsolete and unproductive design practices (Nouws et al., 2022). Instead of concentrating on concrete and specific instruments, public organisations should consider implementing a complementary set of institutional interventions that restructure design practices (cf. Grimmelikhuijsen & Meijer, 2022).

The struggle of public organisations suggests that the translation of theoretical conceptualisations of design processes embedded in democracy and Rule of Law contexts to practice is challenging. We aim to understand how institutional interventions can reshape interactions between designers of public algorithmic systems and, thereby, transform design practices. Accordingly, we will also evaluate to what extent the interventions succeed in stimulating the expected interactions between designers. This evaluation enables us to reiterate the formulated institutional interventions. Research question 3 is as follows:

What institutional interventions engender interactions between designers of public algorithmic systems that align with democratic and Rule of Law principles?

1.3 Research approach

This section elaborates on the design science approach we used to arrive at a design theory for institutional interventions (Section 1.3.1). The institutional interventions are supposed to reshape design practices and, consequently, provide a socio-technical specification in which the possibilities for algorithmic Kafka are reduced (Section 1.3.2). Section 1.3.3 introduces the research methods used in this thesis.

1.3.1 Design science approach

The three research questions formulated in Section 1.2 follow the structure of design science as a research approach. Through design science, scientific and practical knowledge is generated by creating, implementing and evaluating artefacts (Hevner et al., 2004; Peffers et al., 2007). This type of research is based on Simon's (1969/1996) characterisation of design: actions that aim to change an existing situation into a preferred one. As a research approach, design science aims to learn from the changes needed to arrive at the preferred situation. Our research questions also reflect this research approach. Research questions 1 and 2 study the current and desired design practices of public algorithmic systems in public organisations. Subsequently, research question 3 examines interventions in the institutional context of design processes that bridge the gap between current and desired design practices.

Through design science, we can provide public organisations with the means to transform their design practices and embed them in a democracy and Rule of Law context. Apart from the societal relevance of findings resulting from design science, it also generates knowledge with scientific relevance (Meijer, 2025). Scientific relevance in design science is ensured by generalising learnings that follow from studying the implications and effects of interventions in practice.

With its focus on creating artefacts for specific problems, design science can also be considered a solutionist perspective on research. In this thesis, the designed artefact – i.e., a design theory – is not necessarily a definitive solution or quick fix for the undesirable and inadequate design practices currently found in public organisations. It is a way of generating learnings that public organisations can use in reshaping their own practices. This follows from the pragmatic and abductive approach to design science that is in line with the critical realist ontology (see Chapter 2) and constructionist epistemology (Chapter 3) underlying this research. Accordingly, we regard design science as a way to problematise the enactment of technology in practice through a deliberative effort. Arguments are developed by studying problems identified earlier in other domains or other technology (see Bannister, 2023; Meijer & Löfgren, 2015). Subsequently, this knowledge base is used to create new situations in which researchers can explore the contextual additions or modifications of known problems. The arguments in the deliberation are further developed by studying the implications of artefacts created within the research.

This thesis focuses on institutional artefacts, more specifically, institutional interventions. However, the final form and function of these institutional interventions heavily depend on the contextual configuration of particular public organisations. Consequently, the deliverable presented in this thesis is not a detailed articulation of specific institutional interventions. Instead, we generalise the findings from developing specific instantiations of institutional interventions to a design theory. A design theory is a prescriptive theory that translates other forms of theory (theories for analysis, explanation, prediction, and combinations of explanation and prediction) to practice (Gregor, 2006; Gregor & Jones, 2007). We consider a design theory to be a set of design principles that prescribe institutional interventions to transform the design practices of public algorithmic systems. These principles provide universal mechanisms behind the interventions and, therefore, enable public organisations to design interventions that fit their own context. In addition, the design theory comprises a meta-theory (Love, 2000). This meta-theory synthesises the assumptions behind the design principles.

In order to formulate design principles, we followed the three-cycle model of design science (Hevner et al., 2004; Hevner, 2007). A design science project iteratively moves through a relevance, a rigor, and a design cycle in this model. The relevance cycle studies the problems addressed in a design science project in practice. This cycle ensures the societal relevance of the research. The rigor cycle focuses on scientific relevance. It studies the scientific knowledge base relevant to the design problem and, therefore, grounds design science in scientific research. The artefacts are created and evaluated in the design cycle. Through this evaluation (and the iterative nature of design science) generalisable insights on the implications of interventions are generated. In this research, evaluating the design theory in a simulation of a design process provides such scientific insights. The three cycles can be completed simultaneously or in varying sequences.

Nonetheless, the choice for using design science to address the problems in the institutions that structure design practices of public algorithmic systems is also related to the background and training of this thesis' author. I was trained as a complex systems engineer with socio-technical theory as the point of departure. This means that my focus is on comprehensive engineering – i.e., jointly designing institutional, technical, and process artefacts in social systems. Besides my training as an engineer, I studied law. My main interest was in the philosophical and sociological perspectives on law, rather than the application of written law. In this thesis, I bring both academic backgrounds together. I consider institutions as the concept that connects both backgrounds and, therefore, as a natural focus for my research.

1.3.2 Unit of analysis and research context

As mentioned before, the interventions this design science study prescribe concern the institutions that structure interactions between designers of public algorithmic systems. This subsection discusses how this point of intervention is related to the unit of analysis of this research. Where the institutions in design practices are the main point of intervention, the design practice as a whole is the unit of analysis, see Figure 1.1.

The focus on interactions and institutions in design practices is based on the following considerations. The organisation of design processes influences the form and function of the design output (Simon, 1969/1996). The design choices made by designers depend on, for example, the information available to them or the stakeholders they interact with. These elements are partly structured by the institutions that structure the design process. As a result, institutions in design practices play a role in determining what and how design choices are made. This perspective on design practices corresponds with the IAD framework by Ostrom (2005). Chapter 2 elaborates on how the IAD framework is used as an analytic lens for this research's unit of analysis.

Figure 1.1 also shows that public algorithmic systems fall out of the scope of the unit of analysis. That follows from the fact that the actual public algorithmic system is not only created by designers. Designers only have limited control over its shape. Therefore, this research focuses on an artefact that designers do have control over; the socio-technical specification that underlies the public algorithmic system.

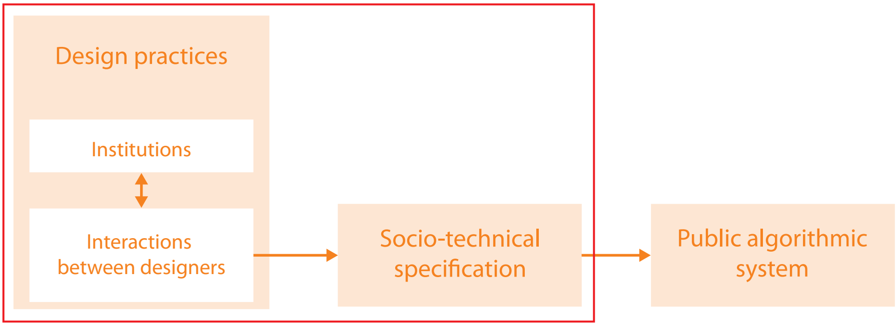


Figure 1.1 Unit-of-analysis: this research studies how interactions between designers result in a socio-technical specification of public algorithmic systems and how these interactions are shaped by institutions

The research is situated in the Dutch context. In the Netherlands, public algorithmic systems are deployed in a wide variety of policy domains (Hoekstra et al., 2021; Van Veenstra et al., 2019). Our interviewees indicated that these systems are partly developed internally, mostly by larger organisations such as executive agencies and large municipalities. But most of the systems are procured externally, or their development depends on the labour of external designers.

Public organisations push for the use of public algorithmic systems but are also confronted with the problems of the use of these systems, for example, false accusations of fraud with childcare allowances (Peeters & Widlak, 2023). In recent years, public organisations across different government tiers in the Netherlands have launched initiatives to tackle problems in their algorithmic practices and, therefore, are reflecting on their design processes (see Chapter 5). Currently, several policy efforts regarding algorithmic systems emerge at the national level. The Ministry of the Interior and Kingdom Relations has developed a policy agenda on digitalisation of government and launched a national algorithm register, and the cabinet has appointed an algorithm watchdog (Ministery of the Interior and Kingdom Relations, 2023). When this research was conducted, these policy efforts were not yet effectuated, and the consequential AI act was still in negotiation.

1.3.3 Research methodology

The research in this thesis is divided into three sequential parts that each focus on one of the three cycles of design science. Our research started with a relevance cycle, in which the current organisation of design processes in Dutch public organisations was empirically studied. In the second part of the research, we shifted towards the rigor cycle. We interpreted our empirical insights by comparing them to theoretical notions on the issues related to public algorithmic systems, and normative conceptions of democracy and the Rule of Law. Through abductive reasoning (cf. Timmermans & Tavory, 2012), we formulated a meta-theory for embedding socio-technical design practices in a democratic and Rule of Law context. The final research part focused on the design cycle. Using the insights from part II, we formulated design principles and tested instantiations of these design principles (i.e., specific institutional interventions) in a local public organisation through a simulation of a design process. The simulation enabled us to evaluate the design principles, and close both the relevance as well as the rigor cycle by bringing the gained knowledge back to practice and scientific literature. The details of the research methods are discussed in Chapter 3, here we provide a short overview of the research methods used in the three research parts.

Diagnosis: empirical study

We answered the first research question by performing two successive empirical studies. Accordingly, we addressed the gap in empirical research. Moreover, in order to do design science, the current situation needs to be analysed in order to formulate a problem that will be addressed through design (Hevner et al., 2004). We use empirical research to gain insight in current design practices.

The goal of the empirical studies was to study design practices for algorithmic systems in Dutch public organisations and the initiatives they develop to transform those practices. We started the empirical research with an exploratory study at a consortium of 13 Dutch public organisations that collaboratively developed policy instruments for algorithmic governance. We mapped design practices of consortium partners and assessed their policy instruments using our theoretical lens (see Chapter 2) to understand what institutional interventions public organisations are already using. We elaborated the results of this exploratory study in an explanatory study. In the latter study, we interviewed different types of designers in design teams in four public organisations.

Our empirical research focused on interviews with policymakers and designers. In addition, we observed meetings of the consortium and analysed documents in which the consortium elaborated their instruments. The emphasis on interviews has shortcomings, such as the dependency on the interpretations of the interviewee. However, Seaver (2017) argues that interviews can be seen as a full-fledged ethnographic method that can provide valuable information that is similar to information gathered through observations. The observations and document analysis at the consortium are used for triangulation of the results of our interviews.

Appraisal: abductive reasoning

Part I provides insight into current design practices within public organisation. The second part of the research focuses on desired design practices that follow from the presuppositions of socio-technical designing, democracy, and the Rule of Law. Our goal was to contrast current design practices with a desired situation using abductive reasoning.

Abductive reasoning is about interpreting empirical results by comparing those results to different, probably applicable but also seemingly unrelated, theories (Timmermans & Tavory, 2012). As such, surprising, remarkable, and striking empirical results are identified. This can result in the generation of new knowledge through sense-making and synthesis. We used abductive reasoning to elicit unsuitable presuppositions underlying current design practices, and to construct a grounding conceptualisation of design practices embedded in a democratic and Rule of Law context.

We started by exploring the kind of design choices that are made in creating public algorithmic systems, and which of these design choices are critical in the emergence of Kafkaesque situations in such systems. Therefore, we used documented cases of public algorithmic systems that inflicted harm on citizens to conduct a secondary analysis. Using the analyses of different scholars of these cases, we identified origins of Kafkaesque situations in the socio-technical specification of public algorithmic systems. These insights enabled us to point towards design practices that can create algorithmic Kafka.

Second, we synthesised the presupposition of socio-technical designing with the presuppositions of democracy, and the Rule of Law. The concepts of democracy and the Rule of Law are contested (Collier et al., 2006; Waldron, 2002) – i.e., their meaning and value is part of elaborate scientific debates that will never lead to a univocal conceptualisation. Accordingly, we followed interpretations of the concepts that align with the universal characteristics of design. The synthesis of the three concepts is based on juxtaposing democracy and the Rule of Law with characteristics of socio-technical designing. By synthesising the three presuppositions, we formulated the meta-theory that establishes an overview of the assumptions underlying our design principles for institutional interventions in the design process of public algorithmic systems.

Create and assess: generate-test cycle

The final part of this thesis formulates the design theory. Our goal was to formulate design principles for institutional interventions and test the institutional interventions in a simulation. The design theory was developed in three generate-test cycles (cf. Hevner et al., 2004). First, we formulated design principles by extending the abductive analysis of Part II. The design principles bridge the gap between the current design practices identified in Part I and the desired practices derived in Part II. We translated the design principles to an instantiation – i.e., institutional interventions in a design process in the form of process instructions. These interventions were tested and evaluated in a simulation of a design process of a public algorithmic system.

The first two generate-test cycles were focused on the preliminary generation and test of the institutional interventions. We translated the design principles to institutional interventions in accordance with conversations with practitioners. These initial institutional interventions were tested in two trial sessions with fellow PhD candidates. After each trial session, we refined the institutional interventions.

The final test of the interventions was performed through a simulation of a design process at a local public organisation in the Netherlands. Before the simulation, we asked participants to describe their current design practices by answering 13 open questions. In the actual simulation, politicians and public servants worked on a fictional design problem through a design process that was structured by institutional interventions based on our design theory. The simulation was observed by two researchers. Afterwards, participants answered another set of open questions to reflect on what happened in the simulation.

1.4 Thesis structure

The structure of this thesis is shown in Figure 1.2. This introduction is followed by *Chapter 2,* which discusses the background of our research. It introduces our critical realist ontology. Thereafter, the chapter elaborates how this ontological perspective influences our conception of public algorithmic systems, (socio-technical) design processes, and institutional interventions. The chapter also discusses our perspective on democracy and the Rule of Law as guiding concepts for shaping the institutional interventions. In sum, Chapter 2 discusses the theoretical lens and the analytical lens of this thesis.

Chapter 3 discusses the research methodology. It starts with discussing the constructionist epistemology that underlies our research design. Thereafter, it elaborates on our design science approach and on our interpretation of a design theory. Finally, the chapter discusses the details of the research methods used in each of the three parts of this dissertation.

Part I Diagnose answers research question 1. It starts with *Chapter 4,* which presents the results of the empirical study of design practices in Dutch public organisations. The chapter first discusses identified design practices. Thereafter, it elicits the institutions and attitudes that structure those practices. The chapter ends with identifying the presuppositions that form the basis of the institutions and attitudes.

Chapter 5 discusses the results of scrutinising the policy instruments of the consortium. First, it positions the consortium instruments in a broader trend of using instruments in AI and algorithmic governance. Thereafter, the chapter assesses the extent to which policy instruments achieved the goal of the consortium. These findings enrich our insights into the presuppositions identified in Chapter 4. Moreover, the analysis of the policy instruments results in a list of requirements for the institutional interventions in the design process of public algorithmic systems.

Part II Appraise is dedicated to research question 2. This part starts with *Chapter 6,* in which we explore Kafkaesque situations caused by public algorithmic systems. The chapter examines infamous and well-studied cases of harmful public algorithmic systems. It explores the role of design choices but mostly the role of designers in the emergence of algorithmic Kafka.

Chapter 7 juxtaposes the concepts of democracy and the Rule of Law with socio-technical designing. This juxtaposition results in symbioses, shared challenges, and contradictions between the three concepts. Eventually, this analysis enables a synthesis of the three concepts.

The synthesis forms a meta-theory that contains all assumptions that underlie the design principles for the institutional interventions.

Part III Create & Assess presents the design theory that follows from research question 3. It starts with *Chapter 8* that presents four design principles for institutional interventions in the design process of public algorithmic systems. These design principles address the gap between current and desired design practices.

Chapter 9 presents the results of the design theory evaluation. The design principles were tested in a simulation of a design process of a public algorithmic system. This chapter presents the simulation and the instantiations used to bring the design principles to practice. Thereafter, the chapter discusses the descriptive results of the simulation and the evaluation of the design principles on six measurement variables.

Chapter 10 presents the conclusions of this research and discusses its contributions and limitations. We end with recommendations for future research as well as for practice.

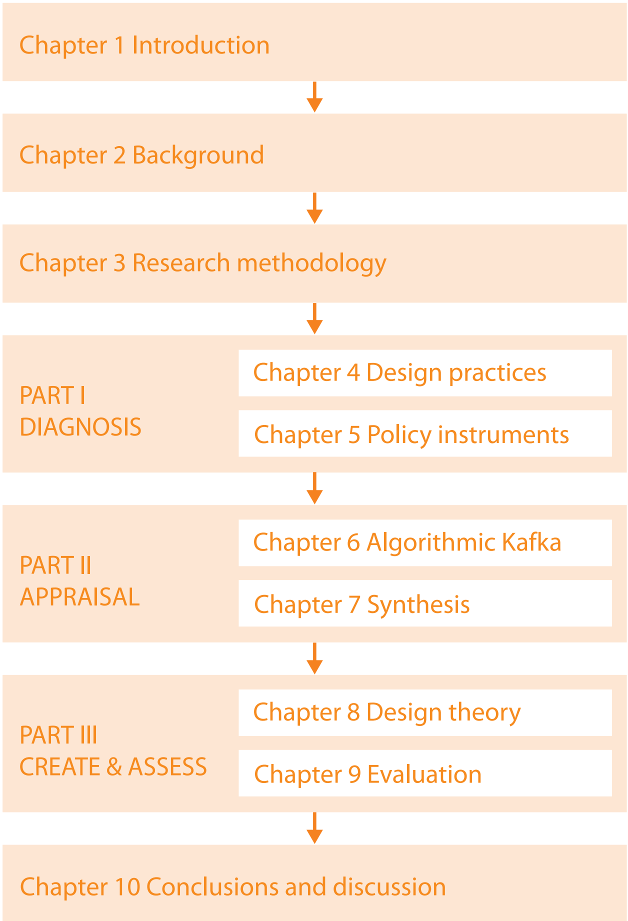


Figure 1.2 Thesis structure

Chapter 2

Background: design processes of public algorithmic systems

Central to the research presented in this dissertation is a technology – i.e., an algorithmic application – designed within the organisational context of governments. The approach to studying the behaviour and use of technology in organisational contexts has been a matter of ontological debate in socio-technical studies (Leonardi, 2013). This thesis takes on a critical realist perspective on socio-technical systems (cf. Bhaskar, 1979, 1998) that shapes our socio-technical view on public algorithmic systems, their design processes, and the role of institutions in these design processes. This chapter discusses what critical realism means for the central concepts in this thesis. First, Section 2.1 elaborates the critical realist perspective on socio-technical systems. Thereafter, Section 2.2 presents our perspective on public algorithmic systems. Subsequently, Section 2.3 discusses the consequences of a critical realist perspective on our unit of analysis: the interactions between designers of public algorithmic systems. Section 2.4 discusses the role that institutions can play in transforming these design practices. Finally, Section 2.5 discusses our perspective on democracy and the Rule of Law. The institutional interventions in our design theory will be based on these two concepts. We end this chapter with a short overview of our unit of analysis, theoretical lens, and analytical lens in Section 2.6.

2.1 A critical realist ontology

Considering the vast number of different conceptualisations of socio-technical systems (Sarker et al., 2019), researchers of these systems have to make their perspective on the nature of both the social as well as the technical, and how these two aspects interact, explicit. This is easier said than done. Researchers of technology in an organisational environment have proposed different perspectives on the interplay between the social and the material or technological. Interest in this interplay was first shown in the field of ergonomics (Mumford, 2006). Researchers started to study the interaction between humans and technology in the workplace. Thereafter, more sophisticated models were developed to study so-called socio-technical systems. For example, Bostrom & Heinen (1977) identified four interdependent components in such systems: structure, people, technology, and tasks. Another example is the Structurational model of technology by Orlikowski (1992) that describes how technology is both the product of human action and a factor that, like institutions, structures human action.

At the beginning of the century, scholars observed that, despite the efforts to model the interrelations of the social and technical discussed before, organisation studies still ignored the role of technology in organisations (Leonardi, 2013). To address this ignorance, Orlikowski (2007) introduced the notion of sociomateriality into the field.[1] The main argument behind the term is that the social and the material cannot be separated; they are mutually constitutive. According to Orlikowski (2007) this means that there is no social without the material and no material without the social. Consequently, the two also cannot be separated in analyses (Barad, 2003). So, the idea was that by introducing sociomateriality, scholars in organisation studies can no longer ignore the material in their research.

At the same time, introducing sociomateriality as an analytical lens resulted in a lively, scientific debate. Mutch (2013) argued that, despite its aim to bring technology back into the analysis of organisations, sociomateriality turns out to focus heavily on agency. Moreover, the concept is impractical in empirical research because it lacks the specificity needed to conduct analyses and it ignores temporality – i.e., material that exists before a certain practice commences (Mutch, 2013). Leonardi (2013) added to this critique that sociomateriality, considered from an agential realist perspective, does not consider the role of external factors.

To counter this critique, sociomateriality can also be considered from a critical realist perspective (Leonardi, 2013). Originally a body of thought of Roy Bhaskar (1979), critical realism is an ontology that positions itself between the two ontological extremes of realism and social constructivism (Archer et al., 2013). It recognises the social constructivist standpoint that science is a social process, and that observations of researchers influence the construction of reality. However, critical realism disagrees that this role of researchers implies that there cannot be one reality that forms the background or basis of those observations and constructions (Bhaskar, 1998). In other words, reality is not less real when it is not perceived by humans, and human experience of structures does not make these structures any less real (Fleetwood, 2005; Leonardi, 2013). Maxwell (2012) describes critical realism as acknowledging the existence of 'one reality but compatible with the idea that there are different valid perspectives on reality.' (p. 9).

Accordingly, critical realism is indifferent to strict dichotomies like those between structure and agency. It considers structure and agency as constituting one another but holds that every situation in which a human agent acts is pre-structured (e.g., by actions of preceding human agents). In other words, the world in which human agents act is always structured by constraints and possibilities that these human agents did not create themselves (Bhaskar, 1998). In order to analyse phenomena from a critical realist perspective, Archer (1995) developed the morphogenetic approach. In this approach, researchers enact 'analytical dualism': they should consider structure and agency as fundamentally distinct, and focus on how the interaction between the two results in elaboration, reproduction, or transformation of structure (Archer, 1995; Leonardi, 2013).

This thesis adopts the critical realist ontology because of its analytical pragmatism regarding conceptualising the components of socio-technical systems. Moreover, the ontology has proved itself in studies on systems similar to algorithmic systems. According to Leonardi (2013), critical realism provides a perspective that also considers software in analyses. Although software lacks the physicality of other objects that we consider as materiality, it shares the sustainability of form and function over time of physical objects (Leonardi, 2012).

1 Orlikowski got the term from Suchman (2006), who coined the term to translate Barad's (2003) agential realist thought to the field of human-machine interaction.

From the critical realist perspective, we conceive socio-technical systems following the Structurational model of technology[2] by Orlikowski (1992) that describes how technology is both the product of human action and a factor that – like institutions – structures human action. The model is based on the idea that socio-technical systems comprise three aggregated components that interact with each other: technology, human agents, and institutions (see Figure 2.1). Each of the aggregated components consists of multiple subcomponents. Orlikowski describes four important interactions between these components. First, technologies are the material or technical artefacts that mediate human action. At the same time, technology is the product of design and use activities of human agents. Third, institutions are 'systems of established and embedded social rules that structure social interactions' (Hodgson, 2006, p. 18), and therefore, mediate human behaviour. Finally, the use of technology can reinforce or transform institutions. The two remaining interactions in Figure 2.1 are about institutions. Institutions are, like technology, the product of human action. Moreover, institutions determine the design and use space of technology. Our critical realist perspective on Orlikowski's model emphasises the possible temporality in the existence of the three different socio-technical components.

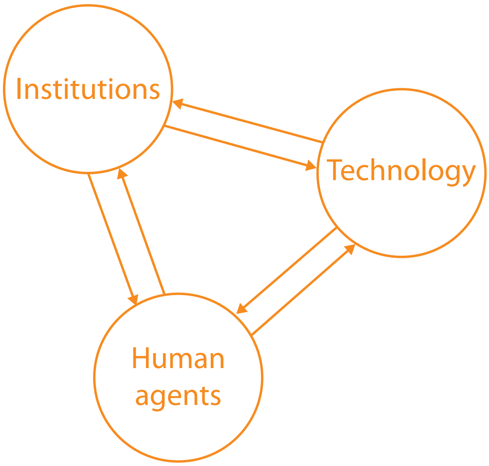


Figure 2.1 The three socio-technical components and their interactions (based on Orlikowski, 1992)

2.2 Algorithmic practices in a governmental organisation context

Following a longer trend of automation of government practices, public organisations are integrating rule-based and data-driven computational technologies in their services and work processes. Section 2.2.1 conceptualises these technologies as algorithmic applications. However, this perspective is too narrow to fully understand the full consequences of algorithmic applications in public administration or the design process of the socio-technical system that emerges. Therefore, Section 2.2.2 applies this thesis' critical realist ontology on

2 The Structurational model of technology is based on Giddens structuration theory (Giddens, 1976, 1979, 1986). This theory is criticised by critical realists such as Archer for conflating structure and agency (see Mutch, 2013). Still, the distinction that Orlikowski makes between different system components is compatible with analytical dualism in critical realism.

algorithmic applications by expanding the scope to algorithmic practices in public organisations. Accordingly, this section presents a socio-technical perspective on public algorithmic systems as the object of design in the design processes we study.

2.2.1 Algorithmic applications

The latest developments in Artificial Intelligence (AI) are part of a longer history of automating human tasks and pursuing machine intelligence. The first advancements in AI were accomplished with expert systems – i.e., business rules based on expert knowledge. When these advancements halted, the push for big data in business contexts coincided with the rise of machine learning techniques and resulted in the development of new AI applications. The latter applications can be characterised as supervised or unsupervised algorithms looking for patterns in vast amounts of data (Elish & Boyd, 2018).

The interest of markets and governments in AI has been fuelled by overselling the capabilities of the applications and framing them as possessing human-like characteristics – e.g., data are 'fed' and machines 'think' (Elish & Boyd, 2018). Still, the promise of reaching human cognition is not in reach. Current AI applications are restricted to achieving a predefined goal and miss constitutive elements of intelligence (Mitchell, 2021). Similarly, the latest advances in generative AI, such as large language models, statistically predict strings of words that fit a context instead of fully understanding a language and applying it (Bender et al., 2021). Accordingly, the term AI is a container for assumptions from early computer science, and of abstract and anthropomorphic descriptions of techniques that serve the interests of investors and policymakers (Suchman, 2023). Following Suchman (2023, p. 1), this dissertation does 'not … deny the achievements and injuries of data-intensive algorithmic practices but rather challenge[s] the misplaced concreteness that the nominalisation 'AI' effects.'

The difficulties and frames in definitions of AI technologies can also be observed in policy spheres. In the realm of public organisations, policy documents by supranational organisations are setting the standards for defining AI. The OECD (2019, p. 7) defines an AI system as 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.' The influential High-Level Expert Group on Artificial Intelligence of the European Union (AI HLEG) provides a similar, but more expansive definition of AI (AI HLEG, 2019a). Moreover, the group formulated requirements for trustworthy AI. According to the AI HLEG, AI should respect all applicable laws and regulations, respect ethical principles and values, and be technically and socially robust (AI HLEG, 2019b). The AI system definition in the European AI Act is inspired by the work of both organisations: 'a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environment'.

This thesis examines the technical artefacts that are currently used in public organisations, not the frames and imaginaries that are built around AI (cf. Elish & Boyd, 2018). Hence, this dissertation will refer to the shared element in rule-based and data-driven techniques: an algorithm. Janssen & Kuk (2016) characterise algorithms using two axes: manual to automated, and simple to complex, while Yeung (2018) uses the distinction between fixed and adaptive algorithms. Considering that simple and/or fixed algorithms can have the same detrimental consequences for citizens as complex and/or adaptive algorithms, this thesis covers both the full range of simple to complex and fixed to adaptive algorithms. However, this research focuses only on algorithms with any level of automation; meaning that a technical component (i.e., computer) is involved (Vagia et al., 2016). Manual algorithms are exclusively executed by human agents. Focusing on automated algorithms, this dissertation relates to the literature on automated decision-making. In sum, we define *algorithmic applications* as automated rule-based and/or data-driven algorithms encoded in software-based technologies that predict or generate output.

2.2.2 Public algorithmic systems

Public organisations enact algorithmic applications in their work processes. New algorithmic applications, such as machine learning, are a next step in a longer tradition of automating and augmenting these processes in public administration by using data, statistics, information systems, and algorithms (Eubanks, 2017; Veale & Brass, 2019). A practice perspective on algorithmic applications is needed to fully understand that these applications only acquire meaning and provide affordances when enacted in a specific context and for a particular goal – in this case, automation or augmentation (cf. Fountain, 2001; Seaver, 2017). Moreover, a practice perspective brings the role of algorithmic applications in politics, power relations, and decision-making into view (Beer, 2017). In this thesis, *algorithmic practices* refer to organisational and human practices in which the enactment of algorithmic applications has become an inextricable part. Hence, while algorithmic applications are a defining element in this research, the practice perspective emphasises that designing algorithmic applications to be used in public organisations cannot be addressed as a purely technological exercise. This section examines the interdependencies between algorithmic applications and other artefacts and actors that provide these applications with a function in a public organisation.

An algorithmic practice is a phenomenon that manifests over time. These practices can be studied but cannot be designed or fully controlled because of their emergent properties. Therefore, we also demarcate the object of design through which designers can influence algorithmic practices. This thesis considers *public algorithmic systems* as the object of design. Following from our critical realist perspective, we consider these systems through the analytical lens of the Structurational model of technology by Orlikowski (1992). Accordingly, public algorithmic systems are considered to be complex socio-technical systems (Dobbe et al., 2021; Elish & Boyd, 2018; Van De Poel, 2020)[3] that (aim to) automate, support, or augment public administration practices by using rule-based and/or data-driven algorithmic applications. Like other socio-technical systems, public algorithmic systems consist of interacting technological, agential and institutional components in which the algorithmic application is an elemental part. By explicating these socio-technical components in algorithmic systems, we also determine the artefact that can be shaped by designers.

Technological component

Section 2.2.1 described this thesis' main focus regarding the *technological component* of public algorithmic systems. But these algorithmic applications are embedded in a system of interrelated technological artefacts (De Bruijn & Herder, 2009). The algorithmic application

3 Ropohl (1999) even argues that all technology should be considered a socio-technical system.

is embedded in an ecosystem of computational infrastructure, software, and datasets that ensures that the application can be used in practice (Mueller, 2025). The need for data, especially in data-driven algorithmic applications, already requires a broad computational infrastructure. For example, the data that feeds the algorithm is gathered by technological artefacts on more local levels such as sensors or administrations (Nissenbaum, 2019). In public algorithmic systems, these datasets consist of data provided by citizens, for example, in benefit requests or data gathered from other public organisations (Eubanks, 2017; Peeters & Widlak, 2018). In sum, the technological component in algorithmic systems is already a complex system in itself. Accordingly, authors have argued that the use of data and information systems has transformed public organisations into system-level or infrastructure-level bureaucracies (Bovens & Zouridis, 2002; Widlak & Peeters, 2025).

Agential component

The agential component of public algorithmic systems comprises all involved and affected human actors in the system. This agential component accentuates that algorithmic applications do not operate fully autonomously but are constitutively entangled with human agents and their behaviour. Typically, four broad types of human agents can be identified in public algorithmic systems: agents *involved* in designing, *involved* in operating the system, *involved* in supervising and regulating the system, and agents *affected* the system. We will shortly discuss each of these categories:

- Agents *involved* in designing
  As will be discussed in Section 2.3, designers are decision-makers, policymakers, data-analysts and developers (both internal and external), information architects, legal officers, privacy officers, ethicists, and domain experts who intentionally shape or change the technological or institutional artefacts constituting a public algorithmic system. Giving expert advice is also considered a design activity.
- Agents *involved* in operating the system
  Operating actors are the ones that implement and use the algorithmic system in their work practices. Especially for front-line workers, work practices have changed due to the introduction of public algorithmic systems and other information systems. Their discretion has generally shifted to so-called street-level bureaucrats (Bovens & Zouridis, 2002; Zouridis et al., 2020).
- Agents *involved* in supervising and regulating the system
  These controllers form a broad category of agents that supervise other actors in the system from a distance. This category includes the judiciary, media, regulatory agencies, and internal and external auditors.
- Agents *affected* by the system
  Multiple actors in the system can be affected by its output such as the front-line worker whose work practices change. Notwithstanding, in this thesis, we focus on citizens as the main affected human agents in public algorithmic systems. It is this role of citizens in the system that makes it a *public* algorithmic system. The algorithmic system intervenes in the dyadic relationship between citizens and public organisations. Citizens might be involved in the system by providing required data but often have little influence on the design and use of algorithmic systems.

In general, the agential component can be perceived as a network of actors with their own perspectives and interests (De Bruijn & Herder, 2009). From these positions, human agents influence the constitution of the technological component. Basically, the actions of human agents in the system shape the goals pursued and affordances emerging from implementing or using the system. At the same time, the technological component also changes the behaviour of human agents. For example, people are prone to confirmation biases when confronted with the output of an algorithmic system (Green & Chen, 2019).

Institutional component

The institutional component comprises all formal and informal social rules that structure the behaviour of human agents (i.e., institutions as defined by Hodgson (2006)). Institutions can have the form of: (1) stable and robust norms, values, and culture on a societal level; (2) legal and written rules that are the product of politics in polities, judiciaries, and bureaucracies; (3) governance or agreements between actors in the form of contracts, relations, guidelines, codes, work instructions; and, (4) rules guiding the interactions between individuals such as positions, roles, and conventions (Koppenjan & Groenewegen, 2005; Williamson, 1998). Institutions also play an important role in the operation of public algorithmic systems as they are constitutively entangled with both human behaviour as well as the technological component.

An example of the interaction between institutions and human behaviour is the role of work instructions in determining how a specific algorithmic application will be used by operators. Such work instructions can, for example, constrain the situations for which an algorithmic system is used. As we will discuss in Section 2.4.2, institutions, and therefore also work instructions, cannot fully determine human behaviour. The operators will develop their own interaction with the algorithmic application and, consequently, refine or redefine (interpretation of) the work instructions. Accordingly, institutions are mostly a way to create expectations between human agents through decreasing uncertainty about behaviour (Hodgson, 2006). As such institutions steer the use of algorithmic applications.

Institutions also interact with the technological component. In the case of public algorithmic systems, these are mostly institutions related to the governmental context. The algorithmic applications are part of a bureaucratic environment of procedures, policies, laws, and regulations (Mulligan & Bamberger, 2019). The corollary institutions structure the form and function of algorithmic applications and related technological artefacts. The interaction between institutions and technological artefacts is not always direct; mostly it is mediated by human agents. Still, especially in rule-based systems, the incorporation of pre-defined rules in an algorithm can be seen as the direct application of a specific institution in a technical artefact. For example, a rule in an algorithm prevents a specific output from being generated. Similarly, the use of an algorithmic application can influence the enforcement of a law. It will partly determine how open norms in laws are interpreted.

When we refer to a public algorithmic system in this thesis, we always refer to the full set of these components and their interactions. This is the object of design for designers who want to constitute or restructure algorithmic practices. Designers can shape the institutional and technological artefacts (Koppenjan & Groenewegen, 2005). To create these individual artefacts, the designers should consider the influence of and to other components. Otherwise, the designer would disregard the interactions between components that will emerge anyway. This shows the boundaries of the influence of the designer. The behaviour of human agents cannot be fully structured by designing institutions and technical artefacts (cf. De Bruijn & Herder, 2009).

2.3 Designing public algorithmic systems

Where the previous section presented the object of design, this section discusses our perspective on design processes of public algorithmic systems. Demarcating these design processes is complicated. This partly originates from the use of the word design, which, especially in academic writing, is an overloaded term. Design, both as a process or product, can refer to the object of study, to activities that subjects of study engage in, and even to activities undertaken by researchers. Moreover, the use of the word differs between scientific domains and disciplines. The diversity in meanings complicates both conducting research on design as well as communicating such research. Therefore, demarcating the usage of the word design in this thesis is a key issue.

The core of the issue is that design is both a noun and a verb (Walls et al., 1992; Steinitz, 1995; Bots, 2007). Design-as-a-noun is the object of design discussed in the previous section. But design-as-a-verb is about 'doing design', which is this thesis' unit of analysis. Throughout this thesis, we use the term 'designing' to stress that we refer to design-as-a-verb. This section elaborates that design-as-a-verb comprises design activities, design practices, and design processes. The double use of design as a term is not only a semantic issue. When analysing designing as a phenomenon, it can be hard to distinguish the social from the material or technological. Technology or artefacts can also play a role in design processes when, for example, designers use them as tools. Moreover, users of artefacts can also change the form or function of artefacts and, therefore, are designers in their own way. Again, we consider these interactions between the social and the material from our critical realist perspective, see Section 2.1.

This section starts with elaborating on the phenomenon of designing by focusing on design activities that individual designers perform. Thereafter, we expand the activity perspective by placing it in the context of public organisations. Thereby, we shift the focus to the interaction between designers. Finally, this section lists the characteristics of socio-technical design processes which apply to the design of socio-technical systems like public algorithmic systems.

2.3.1 Design as a phenomenon

According to Simon (1969/1996, p. 111) 'everyone designs who devises courses of actions aimed at changing existing situations into preferred ones.' The action that aims to change existing situations into preferred ones – i.e., the design activity – is central to this conception of design. Simon (1969/1996) differentiated three design activities: representing the problem, listing alternatives, and evaluating those alternatives. We will discuss four aspects that characterise these design activities: 1) performed by many, 2) inherently creative, 3) explorative, and 4) resulting in a specification.

First, design activities are performed by a variety of actors. Following from Simon's (1969/1996) notion of a designer – i.e., everyone performing design activities – actors can be designers even though that might not be their formal function. In addition, designing is not restricted to technology, designers can also devise institutional artefacts (Koppenjan & Groenewegen, 2005). Moreover, design activities are also performed by users. When using an artefact or system, users give a meaning to that artefact or system (Orlikowski, 1992). The changes to the meaning of the object can have effects similar to those of deliberate design choices (Vermaas & Houkes, 2006). Consequently, in practice, many involved or affected actors can be perceived as designers.

Second, design is an inherently creative activity (Simon, 1969/1996; Van Aken, 2007). Design involves abductive reasoning to end up with new ideas. This abductive process comprises the gathering of all available information, making sense of that information, and then synthesising the information into a design (Kolko, 2010).

Third, design activities are based on inquiry. In other words, a design activity can be defined as 'a purposeful intellectual activity' (Bots, 2007). The explorative nature comes back in the logic of design, which is to identify or generate alternatives that can result in a preferred situation (Simon, 1969/1996). Subsequently, these alternatives are elaborated and tested.

Finally, design activities result in a transition of the state of an artefact or of its specification (Reymen et al., 2006). Designers need to make design choices in order to arrive at that specification. For Bots (2007) and Van Aken (2005b), design activities are only about the activities in which the specification of an artefact (or its model) is formulated. So-called realisation activities in which the artefact is actually created are considered a different type of activity (Bots, 2007). However, distinguishing between design and realisation activities is difficult in practice. Choices made in realisation activities can significantly affect the specification of form and function of an artefact. While recognising the role of design choices in realisation activities, this research focuses on design activities that produce a system specification.

The notion of design activities discussed here is too broad to be used in our research. Therefore, we scoped the concept in line with the focus of this thesis. *Design activities* include all creative actions that affect or result in design choices, constituting the socio-technical specification. A *designer* is an actor involved in or responsible for making design choices that influence at least one socio-technical component of public algorithmic systems. A designer intentionally shapes or changes the technological or institutional artefacts constituting a public algorithmic system. We only consider actors that are not directly affected by the system as designers. Users and other actors interacting with the system, changing or shaping the design output, are excluded as they are directly affected by outcomes of public algorithmic systems. In other words, in this thesis all actors formally appointed or assigned to contribute to or create the system specification are considered designers.

These actors do not necessarily have to have a functional role description of designer. If they perform design activities, this thesis considers them as designers. This broad interpretation of the role is used to identify all actors, officially appointed, that have influence on the design and design process, as well as identify actors who are not. Throughout this thesis we distinguish between representative and executive designers. Representative designers are officially appointed by the public (e.g., through elections), whereas executive designers are appointed as public servants. Chapter 4 elaborates on the difference between representative and executive designers.

2.3.2 Design practices in a design process

Simon's view on design is often characterised as rationalistic because of its problem-solving focus (Dorst, 2019b). The emphasis on problem-solving has three issues, especially in the context of public organisations. First, a design choice that is made creates a new situation that in itself can be a starting point for new design activities (Simon, 1969/1996). In this case, design activities create problems themselves, instead of only solving problems. Second, considering design as optimising current situations by translating scientific knowledge into interventions disregards the political nature of design choices (Van Buuren et al., 2020). Finally, public organisations are often confronted with wicked problems. Wicked problems are complex societal problems for which consensus about the problem is impossible and that can never be fully addressed (Rittel & Webber, 1973). Nonetheless, after Rittel and his co-authors coined the term, wicked problems became the main aim and resulted in a solutionist turn in practice. However, it is in the very nature of wicked problems that comprehensive solutions cannot be found (Pesch & Vermaas, 2020). Public algorithmic systems are often developed in response to a wicked problem (Van Krimpen et al., 2023). Consequently, the design process of these algorithmic systems cannot be approached from a problem-solving or solutionist perspective.

As an alternative to Simon's problem-solving perspective, designing can be regarded as a co-evolutionary practice. This co-evolution perspective understands design as a learning process that results from reinterpretation and adaptation of the formulated problem and solution, or from internal and external changes in the design process (Dorst, 2019a). Over time, both the problem and the solution space become more elaborate, clear, and demarcated. The co-evolution perspective on design as an iterative, reflective, and explorative practice corresponds better with how design-as-a-verb manifests in reality (Dorst, 2019b). Instead of only focusing on activities, co-evolution is a practice perspective on design. We conceive *design practices* as patterns and routines in interactions between designers that emerge over time. Within these interactions, designers perform multiple design activities.

In each context, designers will develop their own practices that they come back to when confronted with a design problem (Simon, 1969/1996; Dorst, 2008). Design practices can also be found in public organisations. Actors in public administration might not consider (most) of their activities as design activities. However, one can recognise the elements of design activities in their daily practices: problem formulation, exploring alternatives, and making choices. Although practitioners might not see public administration as a design-oriented field, public administration scholars have studied the field from a design perspective before. Vincent Ostrom (1973) argued that a theory of design is conditional to understanding systems in public administration. But it was mainly Simon (1969/1996) who positioned public administration as a design science.

When we talk about the design process of public algorithmic systems, we refer to a set of design practices triggered by signals (e.g., a problem surfacing). Consequently, a design process consists of multiple design practices and even more design activities. For socio-technical systems, Bots (2007) classifies the design process as a set of 'concurrent transformations'. Therefore, such design processes do not have clear start and end points. These processes are always situated in a context that was already partially designed by human agents – e.g., the institutional context or existing technological artefacts used. Considering these characteristics, demarcating the boundaries of a design process of socio-technical systems is inherently complex. This thesis conceives a *design process* as the complete set of design practices enacted during the whole life cycle of public algorithmic systems.

2.3.3 A socio-technical design process

The design processes of socio-technical systems are socio-technical processes themselves (Clegg, 2000). Accordingly, the design process of public algorithmic systems is also socio-technical in nature. This section elaborates on what a socio-technical design process means by listing the characteristics of such processes. This section summarises five characteristics discussed in the literature: systemic, emergent, contingent, multi-stakeholder, and political.

Systemic
First of all, public algorithmic systems consist of three socio-technical components that all need to be considered in the design process. According to Clegg (2000), this requires a systemic design approach, in which all system components (i.e., technical artefacts, institutions, and the behaviour of human agents), and the interactions and interdependencies between these components are consecutively designed and are consistent with each other. Therefore, the deliverable of a socio-technical design process comprises a technological (systems) design, a process design, and an institutional design (De Bruijn & Herder, 2009; Koppenjan & Groenewegen, 2005). All three components come with their own related perspective and expertise, which makes systemic design difficult, maybe even impossible. Baxter and Sommerville (2011) plead for integrating the different design approaches related to the different components. They propose the integration of system engineering processes with organisational change processes. On the other hand, De Bruijn and Herder (2009) argue that full integration of approaches will 'water down' the importance and affordances of the technological, agential, and institutional perspectives on socio-technical systems. They conclude that these 'competing' perspectives should be used alongside each other when designing socio-technical systems. Scholars do not seem to agree on how to combine knowledge and design associated with the three socio-technical components in a system approach towards design.

Emergent
Second, the three socio-technical components interact with each other, resulting in emergence. The literature on socio-technical designing distinguishes two forms of emergence. Emergence either means that a system is more than the sum of its elements (Checkland, 1981; Ropohl, 1999), or that the behaviour of a system will only become clear or known over time (Bauer & Herder, 2009). The latter can be the result of users (agents directly affected by the system) shaping or changing the system by, for example, giving an artefact a meaning different from what the designer intended (Kroes et al., 2006; Orlikowski, 2000; Vermaas & Houkes, 2006). Both types of emergence bring uncertainty – i.e., it will impede a full understanding of the design systems and their consequences. Moreover, emergence means that the design process never stops; a 'final' design object is not achievable. Moreover, designers cannot fully predict how a public algorithmic system will influence its environment, whether it will encourage public values, and what new problems are created when implementing and using the system (Simon, 1969/1996). Therefore, emergence implies anticipation and reaction to system consequences in design processes through co-evolution (Dorst, 2019a) and learning processes (Checkland, 1999).

Contingent
Third, the designed system does not emerge from nowhere. The design process occurs in a context of existing institutions and technology on which design choices are contingent (cf. Clegg, 2000). In other words, specifying the system components is constrained and enabled by infrastructural factors: existing technological structures and (in)formal institutional structures (cf. Leonardi, 2013). This also means that the start and end of the design process are not always clear and that it is hard to demarcate system boundaries. For example, large socio-technical systems such as energy infrastructures or computational infrastructures are not designed as an integral system but gradually evolve over time (Herder et al., 2008; Janssen et al., 2015). Similarly, when designing public algorithmic systems, designers are dependent on existing information architectures and datasets in their organisation (Gürses & Van Hoboken, 2018; Nissenbaum, 2019; Widlak & Peeters, 2025). As such, the design process is shaped by its material and organisational context. For example, most design processes are structured by the agile turn observed in software engineering. Concurrently, the idea of agility in development processes shapes the form and functions of algorithmic systems (Gürses & Van Hoboken, 2018).

Multi-stakeholder
Fourth, designing socio-technical systems is a social process (Clegg, 2000; Ropohl, 1999). The design process is a set of interactions between a wide array of heterogeneous actors involved or affected by the system to be designed. The diversity in actors follows from the roles they occupy within the system and its design process, but also from their frames of reference, attitudes, perspectives, interests, and needs regarding the system (Bostrom & Heinen, 1977; De Bruijn & Herder, 2009). All these actors can shape or change the system. As a result, the design process is a decision-making process (De Bruijn & Herder, 2009). Still, the extent of influence of an individual actor on the form and function of a system is determined by their position in the actor network. These power struggles can change over time as actors learn. At the same time, education about socio-technical designing can also make the playing-field more equal (Clegg, 2000).

Political
Finally, since socio-technical designing is a form of decision-making characterised by power struggle, the process is also a political process (Brey, 2018). Designers make design choices between alternatives that relate to different arrangements of values or are confronted with conflicting values (Clegg, 2000). Consequently, hard choices about what values will be given the most weight are to be made (Chang, 2017; Dobbe et al., 2021). The values that are prioritised will be embodied in the final system (Bowker et al., 2009; Van De Poel, 2020). This means that the role and positions of experts within the political process of designing need to be defined. For example, experts can be scrutinisers of so-called 'negotiated nonsense' that may result from power struggles between involved and affected actors (De Bruijn et al., 2010; Herder et al., 2008).

The five characteristics of socio-technical design processes are all sources of complexity. Because of this complexity, the design process is not fully controllable (see Bauer & Herder, 2009; Kroes et al., 2006). This has implications for organising socio-technical design processes. There is no one-size-fits-all institutional framework that can structure socio-technical design processes. Instead, institutions that shape the design process should mostly focus on supporting designers in responding to the idiosyncrasies and complexities of the design process.

2.4 Institutions to structure design processes

Sections 2.2 and 2.3 described our perspective on public algorithmic systems and their design processes. Within the broader scope of these concepts, this research focuses on the institutions that shape design practices of public algorithmic systems. This section presents the Institutional Analysis and Development (IAD) framework by Elinor Ostrom (2005) as the analytical lens we use to study design practices and how they are shaped by institutions. The IAD framework is a research instrument to analyse and design the institutions that structure interactions between actors in collective action situations.

This section elaborates on why the IAD framework is useful for this research, and how we will use it. Section 2.4.1 will present the framework and how it can be used for analysing design practices of public algorithmic systems. Apart from analysing design practices, this thesis also prescribes institutional interventions. To delineate institutional interventions, Section 2.4.2 discusses the literature on institutional change and institutional design, and how this relates to the IAD framework. We position the institutional interventions for which our design theory prescribes design principles within this literature.

2.4.1 Studying design practices through the IAD framework

The emphasis of this thesis on design-as-a-verb means that it focuses on activities that lead to design choices. The complexity of public algorithmic systems, resulting from the interactions between institutional, agential, and technical components, means that collaboration between different designers representing different types of expertise is needed. In other words, design practices of public algorithmic systems are collective action situations. Accordingly, studying design practices requires an analytical lens that entangles the interactions between actors and how these interactions are coming into being. The Institutional Analysis and Development (IAD) framework by Elinor Ostrom (2005) describes the most relevant variables that result in interactions between actors, see Figure 2.2. Before discussing the framework in detail, we discuss its applicability for this research.

The IAD framework captures collective action situations, and design practices are collective action situations. Designers interact with each other to arrive at design choices. The outcome of this collective action is a socio-technical specification of the public algorithmic system. Structural elements that shape such collective action situations are captured as exogeneous variables in the IAD framework. In accordance with our critical realist perspective on design processes of public algorithmic systems, these exogeneous variables capture both pre-existing structural elements but are also influenced by the outcomes of the collective action situation.

The underlying pragmatist epistemology of the IAD framework (Aligică & Boettke, 2010; Groenewegen, 2011) also corresponds with this study's research approach. Although the framework was originally developed to study collective actions related to common pool resources, the framework has been used extensively in empirical research and theory development on interactions and institutionalisation of interaction between humans in various domains (Heikkila & Andersson, 2018). The framework provides an overview and definition of variables that generally constitute action situations. Moreover, the framework enables both the analysis and design of institutions in a specific context.

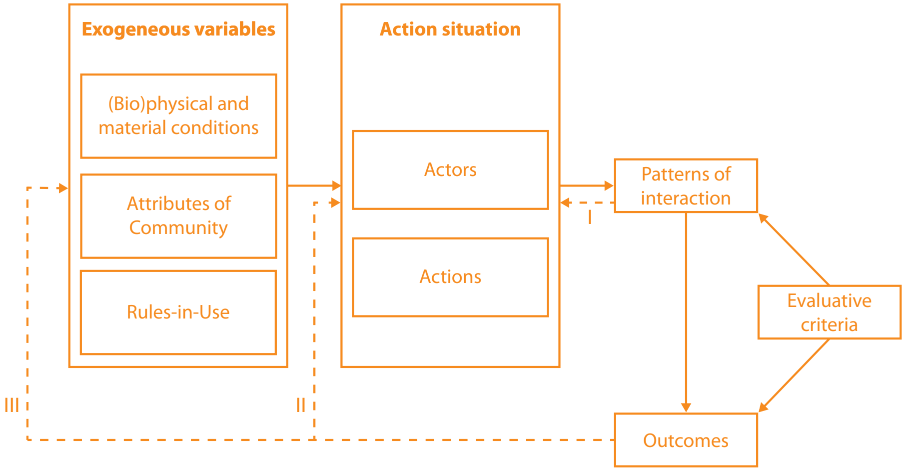


Figure 2.2 The Institutional Analysis and Development (IAD) framework by Ostrom (2005, p. 15)

The core of the IAD framework is the action situation. McGinnis (2011) classifies action situations as 'the black box' in which policy choices – in this research design choices – are made. The action situation comprises seven variables that configure the possibilities for interactions between actors, see Figure 2.3. The variables are the *actors* involved in an action situation, the *positions* these actors are assigned to, and the *actions* that actors in specific positions can perform. The actions taken by actors can result in *potential outcomes* which can be evaluated based on associated *net costs and benefits.* The link between actions and outcomes is mediated by two variables. Actors have *information about* the action situation's composition – e.g., information that actors have about other actors and how they act. This information can be based on past experiences. *Control over* signifies the influence that actors have on the outcomes of the action situation. Both the extent of control and access to information vary among actors (E. Ostrom, 2005).

Each action situation is structured by so-called exogeneous variables, see Figure 2.2. Ostrom (2005) distinguishes three variables: rules-in-use, attributes of community, and (bio) physical and material conditions. The rules-in-use are 'shared understandings by participants about enforced prescriptions concerning what actions (or outcomes) are *required*, *prohibited*, or *permitted* [emphasis in original]' (E. Ostrom, 2005, p. 18). The attributes of community are the social constructs prevalent in a community. It includes the culture in a group or organisation, shared meanings by involved actors, and the extent of homogeneity and equality in a group. (Bio)physical and material conditions are the natural and artificial resources that constrain or afford specific interactions or outcomes. Actors in a design process also use technology or materials in design activities, for example, communication tools or programming software. The qualification 'exogeneous' is a bit misleading. Ostrom (2005) stresses that the exogeneous variables are also structured by the action situation (see Section 2.4.2). However, when using the IAD framework for analysis, the researcher considers a particular action situation in a specific moment. For that specific moment, the rules-in-use, attributes-of-community, and material conditions are exogeneous.

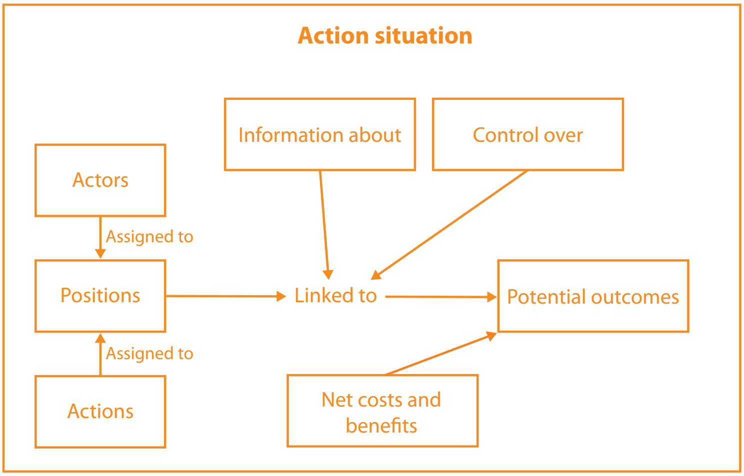


Figure 2.3 Elements in the action situation (E. Ostrom, 2005, p. 33)

Eventually, an action situation results in patterns of interactions. The IAD framework provides the universal components that underlie social interactions and social behaviour, but the specific constitution of those components is different for particular contexts. In using the framework, analysts need to identify what variables and relationships are more important than others in a particular research context (E. Ostrom, 2005). This study focuses on a specific type of actions – i.e., design activities – and outcomes – i.e., design choices about the socio-technical specification of public algorithmic systems. Regarding the exogeneous variables we focus on two specific categories. In the case of rules-in-use, we refer to institutions as defined by Hodgson (2006), whose definition we use throughout this thesis. Moreover, we have scoped down the attributes-of-community because of difficulties in observing or measuring organisational cultures. In performing our interviews, we noticed that interviewees mostly referred to a range of attitudes towards the design process among actors in public organisations. These attitudes also shape the action situation. We consider attitudes to be assumptions, values, and beliefs held by designers on the nature of the design process and public algorithmic systems. (Bio)physical and material conditions are out of the research scope, since we focus on how institutions influence design practices. Notwithstanding, the technological artefacts in public algorithmic systems are an important focus of this research, but these are considered outcomes of the patterns of interaction instead of exogeneous variables. Figure 2.4 presents the configuration of the IAD framework we use in this thesis.

As discussed in Section 2.3.2, the design process consists of multiple design practices. Similarly, it is not possible to study design processes of public algorithmic systems as a singular action situation. In public organisations, decision-making takes place in a network of adjacent action situations (McGinnis, 2011). This follows from the polycentric perspective of the Ostroms on public administration (Aligică & Tarko, 2012). In a network of adjacent action situations, multiple action situations exist that can influence, for example, the outcomes or exogeneous variables of other action situations. The network perspective on action situations is indispensable when applying the IAD framework to a public administration context (McGinnis, 2011). In sum, the design process of public algorithmic systems is such a network of adjacent action situations.

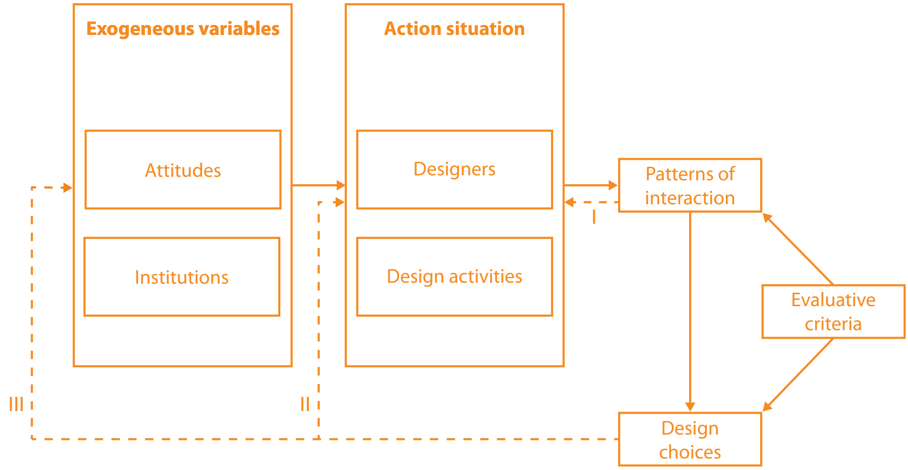


Figure 2.4 The Institutional Analysis and Development (IAD) framework by Ostrom (2005, p. 15); adapted to design practices of public algorithmic systems

2.4.2 Institutional change and institutional design

This research aims to transform design practices in the design process of public algorithmic systems through institutional interventions. Considering that institutions create stability (Peters, 2019), a focus on institutional change is not self-evident. Hodgson's (2006, p. 18) definition of institutions ('systems of established and embedded social rules that structure social interactions') also emphasises the stability of institutions. Social rules become embedded through durability and prevalence – i.e., when individuals can 'create stable expectations of the behaviour of others.' As such, institutions can constrain and enable behaviour (Hodgson, 2006). At the same time, institutions that do not change become redundant (Farjoun, 2010).

Ideas on institutional change differ among institutional theories. There are theories that consider institutions as fully external to practices and theories that argue that institutions only exist because of enactment (Peters, 2019). Following our critical realism perspective, we can distinguish two forms of institutional change (cf. Olsen, 1997; Koppenjan & Groenewegen, 2005; Klijn & Koppenjan, 2006). First, institutional change can be evolutionary and gradual. This type of change is related to enactment. In this case, actors give new interpretations to institutions, or stop adhering to institutions (Klijn & Koppenjan, 2006). In other words, by creating a routine in behaviour, actors create expectations about behaviour that become institutionalised. Second, institutions change through design. Deliberate action is the defining factor in this kind of institutional change (Klijn & Koppenjan, 2006). Institutional change through design is most prominent in legislative processes. However, the distinction between the types of institutional change is not polar (Buitelaar et al., 2007). Designers have limited influence on institutions and there are no institutional panaceas (Peters, 2020). Most of all, designers need to take into account that institutions will change organically through the learning and adaptive capabilities of individuals, notwithstanding their own interventions. Similarly, designers need to consider pre-existing institutions delineating their design space, institutions that are designed in other action situations, and the negotiations that give the design its legitimacy (Koppenjan & Groenewegen, 2005).

Both types of institutional change are incorporated in the IAD framework. First, the three learning loops in the IAD framework (see Figure 2.4) relate to organic institutional change (E. Ostrom, 2005). Milchram et al. (2019) show that the framework includes three loops of learning through which actors change action situations or the structuring exogeneous variables. The first loop connects the patterns of interaction with the action situation (dotted line I in Figure 2.4). The second loop brings knowledge and experiences about outcomes to the action situation (dotted line II). And the third loop changes the institutional context based on learnings from the outcomes (dotted line III).

Second, institutions can be designed on organisational levels higher up the hierarchy. Ostrom (2005) distinguishes three levels that are applicable to this research (see Figure 2.5). The operational level comprises the action situations in which practical decisions are made by individuals who have been appointed by the collective-choice level. Design activities for public algorithmic systems take place on the operational level. The collective-choice level represents the level on which institutions are deliberately designed and decided on. As such, this level shapes the exogeneous variables on the operational level. Our design theory is aimed at actors on this collective-choice level because it provides design principles for shaping the institutions of operational action situations. The processes on the collective-choice level mostly follow the structure that is determined on the constitutional level. For example, the constitution that is formulated on the constitutional level prescribes the procedures that the collective-choice level should follow in deciding on how to organise design processes. Apart from the top-down influence between the levels as described, the lower levels can also cause restructuration on higher levels.

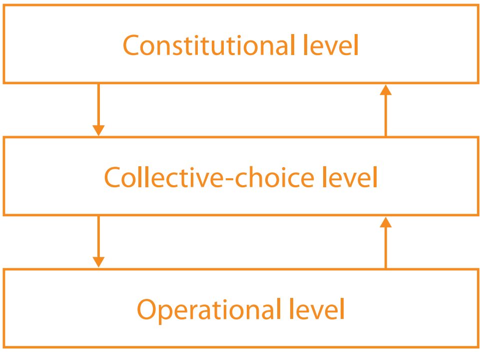


Figure 2.5 Hierarchical levels on which action situations can be situated (Ostrom, 2005)

2.4.3 Institutional interventions

We acknowledge the organic change of institutions but focus on deliberative design of the institutional context of design processes through institutional interventions. More specifically, we prescribe an institutional design (cf. Koppenjan & Groenewegen, 2005) aimed at changing the structures that shape design practices. In other words, we formulate possible changes to the design of the design process (cf. Van Aken, 2005b) in order to build a new 'character' in public organisations – i.e., new 'commitments to ways of acting and responding' (Selznick, 1984, p. 47). Accordingly, institutional interventions are deliberate changes to the structure of action situations in the design process of public algorithmic systems.

Designing institutions is a negotiation, a decision-making process. As researcher we cannot provide a legitimate institutional design (cf. Olsen, 1997). After all, 'to institutionalise is to *infuse with value* beyond the technical requirements of the task at hand' (Selznick, 1984, p. 17). We do not claim the authority to dictate institutions. Instead, this thesis prescribes design principles based on a particular set of values that can inform the negotiation process on appropriate institutional interventions at the collective choice level (see Figure 2.5). Accordingly, the design principles are heuristics for public organisations to design institutional interventions. Chapter 3 will elaborate on the form of these design principles.

We also use the IAD framework by Ostrom (2005) to identify points of intervention in the institutional context of design processes. Ostrom (2005) defines seven rule types that each structure one of the seven variables in the action situation, see Table 2.1. Together, the full set of rule types configures the action situation. In other words, the rules are interdependent (E. Ostrom, 2005). Ostrom identified the rule types from the vast amount of case studies on the self-governance of communities around common pool resources. Grimmelikhuijsen & Meijer (2022) show that the rule types can also be used as a template for institutions that strengthen the democratic legitimacy of algorithmic systems. Likewise, we use the rule types to reshape action situations in design processes of public algorithmic systems to strengthen the process' (democratic) legitimacy. More specifically, our design principles prescribe the specification of the seven rule types in order to arrive at design processes that are embedded in a democratic and Rule of Law context.

Table 2.1 Rule types related to each element in the action situation (E. Ostrom, 2005, pp. 193–210)

Rule type	Variable in action situation	Description
Position rules	Positions	Establish positions
Boundary rules	Actors/participants	Determine eligibility to position
Choice rules	Actions	Determine actions that a participant in a position must, must not, or may perform
Aggregation rules	Control	Assign decision-making power over taking actions to positions
Information rules	Information	Determine what information is available about action situations to positions
Payoff rules	Net cost and benefit	Describe external rewards or sanctions associated to particular outcomes
Scope rules	Potential outcomes	Determine outcomes that may, must, or must not result from the action situation

2.5 Democracy and the Rule of Law as fundaments for institutional design

As the basis for the institutional interventions that reshape socio-technical design practices, we used two presuppositions that are directly connected to curbing algorithmic Kafka.

Considering that democracy and the Rule of Law aim to protect citizens from situations like algorithmic Kafka, these two concepts form the basis for the design theory formulated in this thesis. In addition, the concepts of democracy and the Rule of Law reflect what is currently lacking in design practices of public algorithmic systems.

First, algorithmic practices often lack democratic legitimacy (see Grimmelikhuijsen & Meijer, 2022; König & Wenzelburger, 2021). This partly originates from democratic deficits in design processes of public algorithmic systems, for example, because of power imbalances in the design process (Zouridis et al., 2020) and a lack of (democratic) politics in the design process (Mulligan & Bamberger, 2018; Van Zoonen, 2020). The lack of political mechanisms is alarming since the development of algorithmic systems tends to present many normative trade-offs that are social and political in nature (Selbst et al., 2019; Winner, 1980). Democracy is a way to provide legitimacy to design choices.

Second, public algorithmic systems can be disruptive to society and citizens (e.g., Alkhatib, 2021; Dencik et al., 2019; König, 2020). These systems play a significant role in the relationship between citizen and government, as these algorithmic applications can be used to govern citizens' behaviour (cf. Janssen & Kuk, 2016). This increases the possibilities of arbitrary conduct from which Kafkaesque situations emerge. The Rule of Law's main focus is to provide measures and institutions that protect citizens against arbitrary use of power (Krygier, 2009; Raz, 1979). Therefore, the Rule of Law is based on the idea that people should not be ruled by man, who are inclined to use power arbitrarily, but by law (Raz, 1979; Waldron, 2002; Krygier, 2009). In this conception, laws are legitimated rules.

The two concepts are vast and, most of all, contested. Both democracy and the Rule of Law are contested concepts that are interpreted in distinguishable, and sometimes opposing, democratic theories or philosophical perspectives (Collier et al., 2006; Cunningham, 2002; Gallie, 1956; Waldron, 2002). Being contested concepts that have no univocal meaning, we need to choose a specific interpretation of both democracy as well as the Rule of Law. This section will discuss our interpretations. We based our choice for interpretations on the following. First, the understanding of both concepts has to align with our understanding of design (as discussed in Section 2.3). Second, we have considered how the concepts have been discussed in the literature on algorithmic systems. Selecting specific interpretations also comes with its limitations. Institutional interventions based on other interpretations of democracy and the Rule of Law will be different from the interventions prescribed in this thesis.

Sections 2.5.1 (democracy) and 2.5.2 (the Rule of Law) will provide a short overview of the interpretations selected for this research. Chapter 7 will elaborate on how the concepts are used in the design theory. Moreover, Chapter 10 reflects on the shortcomings and limitations of using contested concepts as the fundament for a design theory.

2.5.1 Democracy

Scholars have been calling for democratising AI and related applications in response to the intrusive and disruptive consequences of algorithmic systems for society and individuals. However, this attitude of solving problems in AI by simply introducing democracy is critiqued for being opportunistic, negligent and superficial. For example, Noorman and Swierstra (2023) observe how big tech is co-opting the democratisation language by interpreting it as providing as many people as possible access to AI technologies, which aligns closely with their commercial interests. Similarly, the scientific debate on democratising AI is critiqued for little engagement with political philosophy literature (Himmelreich, 2019; Noorman & Swierstra, 2023; Sætra et al., 2022). As a result, the literature on democratising AI has an impoverished and superficial interpretation of democracy, bases claims on democratising AI on weak grounds, provokes redundant and resource-intensive practices, and fails to acknowledge that democracy alone cannot solve injustice or oppression (Himmelreich, 2023).

In response to the critique on democratising AI, Himmelreich (2023) argues that researchers should focus on embedding algorithmic practices in prevalent democratic institutional structures and improving these existing structures instead of developing new democratic practices. Noorman and Swierstra (2023) observe a disregard of politics in design choices in Himmelreich's approach, as he assumes that political decision made on higher organisational levels will easily trickle down to design and operational levels. This thesis acknowledges both Himmelreich's (2023) argument to start from existing democratic practices as well as the need for democratic practices at operational levels as argued by Noorman and Swierstra (2023). The starting point is to embed current design practices in existing democratic and Rule of Law structures. The socio-technical nature of public algorithmic systems entails that their design process cannot be separated from other processes within (public) organisations such as policymaking processes. At the same time, we acknowledge the need for adapting design practices following democratic and Rule of Law principles. In the end, this might be accompanied with structural changes to the existing democratic and Rule of Law institutions prevalent in a public organisation. The remainder of this section discusses our perspective on democracy in designing.

Democratic fundaments for design practices

Democratic theory is a collection of various theories that represent a distinctive interpretation of democracy as a concept. Exemplary democratic theories are pluralism, participatory democracy, deliberative democracy, and representative democracy. Despite their differences, the democratic theories share three fundamental tenets that the theories elaborate on. The three fundamental tenets of the democracy concept are: liberty, participation, and self-correction (Christiano, 1996; Cunningham, 2002). Each democratic theory gives its own interpretation of the meaning, value, and conduct of these fundamental tenets (Cunningham, 2002).

In defining democracy for the purpose of this thesis (and its output, the design theory), we started from the current view on democratising designing in the literature and followed the pragmatic epistemology of this thesis (see Section 3.1). As discussed in the introduction, multiple authors have explored the possibilities of democratising design processes of technology (Feenberg, 1999; Hajer, 1995; Ozkaramanli et al., 2022; Sclove, 1995). The main challenge defined in the literature for democratising designing is to find a balance between engaging citizens and including expertise in technology development (Feenberg, 1999). Scholars mainly arrive at a division of work between the public (who defines the agenda, criteria, etc. for the system to be designed) and the expert who realises the design (Hajer, 1995; Pesch, 2021; Sclove, 1995). Hajer (1995) emphasises that this should be accompanied by a reconceptualization of current political institutions. In current institutional designs of public organisation, the expert can easily distance themself from the public.

A democratic theory that aligns with this view on democratising design and the pragmatic epistemology behind this research is Dewey's (1927/2016) interpretation of democracy. He considers democracy as an inquiry by the 'public' into actions that address societal problems. The pragmatist Dewey also argues for a division of labour between the public and the expert and provides a demarcation of the public. We will now discuss Dewey's interpretation of the three fundamental democratic tenets. In doing so, we also relate the tenets to problems in design practices discussed in Chapter 1.

The liberty (or freedom) tenet stipulates that citizens can govern themselves and, thereby, rejects illegitimate power exercised over citizens through government action, or by a few individuals over all other individuals (Christiano, 1996; Cunningham, 2002). Accordingly, citizens should be free from the arbitrary use of power that they are confronted with in algorithmic Kafka. In Dewey's idea of democracy, the liberty tenet comes back in demarcating the public (i.e., the main actor in governing society). According to Dewey, the public is formed by those affected by the indirect consequences of transactions, including harmful and unwanted consequences (Dewey, 1927/2016).

The participation tenet refers to the possibility for everyone to engage in democratic decision-making. In order to respect everyone's freedom and rights (i.e., the liberty tenet), each citizen should have equal opportunity to participate and shape self-government (Christiano, 1996; Cunningham, 2002). Adhering to the participation tenet also addresses the lack of democratic legitimacy of design practices of public algorithmic systems. In Dewey's interpretation, the public is in the lead of the inquiry into actions to address problems. Every member of the public should have the possibility to participate but this poses two challenges. First, there is a need for expertise in the inquiry. This requirement can be achieved through representative democracy. The public frames the problems and sets goals and boundaries based on their needs and interests. Representatives translate those to solutions or measures. Ultimately, the public makes trade-offs between the alternatives provided by these representatives. This division of labour should prevent elitism. However, a second problem emerges here. The public needs to be in the position to steer and control the representatives. According to Dewey, this happens through public debate in which every member of the public should be able to engage. Accordingly, the public should have access to all appropriate knowledge. It is the representatives' task to ensure this, but it also requires the education of citizens (Dewey, 1927/2016).

The self-correction tenet requires democratic processes to be in a state of continuous learning, either through retrospectively reflecting on past decisions (Olsen, 2009), or through contestation of made decisions (Spicer, 2019). This reflexivity is also necessary in design, as it inherently produces interventions in society that can create new problems (Dorst, 2019a; Simon, 1969/1996). Algorithmic Kafka is an example of a possible problem created by the introduction of an algorithmic system as intervention. The iterative practice comes back in Dewey's idea of democracy as an inquiry. For Dewey, this inquiry is similar to that in scientific research with an important role for the reflexive attitude of always questioning established orders.

2.5.2 The Rule of Law

Public algorithmic systems are considered a challenge or threat to the Rule of Law (Bayamlıoğlu & Leenes, 2018; Hildebrandt, 2016). Bayamlıoğlu and Leenes (2018) identify challenges to law as a normative, a causative, as well as a moral enterprise and even caution 'that the 'rule of law' might be exchanged for the 'rule of technology' – accompanied by Kafkaesque, Huxleyan, and Orwellian discourses of dystopia.' (p. 305). In practice, this would mean a return to the rule by man (as designer of technology) which is opposite to the Rule of Law. At the same time, algorithmic systems are subject to the Rule of Law. Several authors have argued that although algorithmic applications are of a different order compared to traditional laws and policy execution, they should fall under Rule of Law regimes (Brownsword, 2016; Hildebrandt, 2018). Consequently, public algorithmic systems fall under the Rule of Law but simultaneously undermine Rule of Law mechanisms that govern the processes in which these technologies are applied.

However, like the use of the democracy concept, the field of AI and algorithmic studies does not fully engage with legal philosophy literature when discussing the design process of public algorithmic systems. At the same time, legal philosophy literature mostly disregards the role of technology in Rule of Law practices (Nouws & Dobbe, 2024). As the role of technology in Rule of Law studies is obscure, it is also unclear how the Rule of Law can contribute to organising socio-technical design practices. This section discusses our perspective on the Rule of Law that can be applied to design practices of public algorithmic systems.

Rule of Law fundaments for design practices

Like for democracy, theoretical perspectives on how the Rule of Law reduces arbitrariness differ fundamentally. We will shortly give an overview of the different perspectives discussed in legal philosophy. In doing so, we reflect on the applicability of these perspectives for design practices of public algorithmic systems.

Traditionally, philosophy of law studies the required characteristics of rules to establish the Rule of Law. As such, it narrowly focuses on legalistic interpretations of the Rule of Law. Citizens are protected from arbitrary use of power if rules have the 'right' form of content. These 'traditional' interpretations are mainly distinguished as the formal and the substantive perspectives. The *formal* – or thin – understanding of the Rule of Law provides (lists of) requirements for the form of rules that contribute to or comply with the Rule of Law (Raz, 1979). In the end, the form of rules must ensure that the public is able to obey the law. Therefore, the rules are not necessarily related to moral rights (Raz, 1979). In reaction, Dworkin (1980) argues that the formal perspective entails that people should obey rules irrespective of the content of those rules, and stresses the importance of a *substantive* – or thick – perspective on the Rule of Law that emphasises the connection between rules and citizen's moral and political rights. The Rule of Law is a normative practice that guides, demarcates or constrains the content of rules (Dworkin, 1980). The traditional perspectives on rules do not offer much insight into designing. It can form a basis for norms that the design output (i.e., public algorithmic systems) should adhere to, but it has little to say about organising design practices.

For a practice perspective on the Rule of Law, we move to the *procedural* perspective. According to Waldron (2011), both these perspectives do not align with how ordinary people perceive the Rule of Law and, therefore, miss a key element. He argues that ordinary people have a procedural perspective on the Rule of Law. This perspective emphasises the Rule of Law as structuring argumentative practices that achieve objectivity. In other words, the Rule of Law should provide procedures to ensure that such an argumentative practice (e.g., in law-making or in court) runs properly (Waldron, 2011). The emphasis on argumentation closely aligns with the deliberative nature of designing.

Although the procedural perspective is practice-oriented, it is still mainly focused on (formal) rules that should organise procedures. As argued in this chapter, design practices cannot be captured or controlled in formal procedures. Therefore, we relate to the practice-oriented socio-legal perspective on the Rule of Law. This perspective, informed by sociology, approaches the Rule of Law as applicable to all interventions that influence the organisation and functioning of the state (Krygier, 2014) and expands its repertoire beyond legal instruments by considering the role of political, administrative, and cultural aspects in sustaining or materialising the Rule of Law (Krygier, 2009; Nonet & Selznick, 1978; Taekema, 2021). The socio-legal perspective aligns with the broad interpretation of institutional interventions discussed in Section 2.4.

Krygier's (2009) interpretation of the socio-legal perspective shows the applicability to design-as-a-verb. He distinguishes two steps. First, determine the immanent goal or purpose of the Rule of Law – i.e., reducing or addressing arbitrary use of power. Our interpretation of institutional interventions also puts emphasis on the goal that is to be achieved. Second, design and implement measures, adapted to the applicable context (Selznick, 1999), that advance the immanent goal or purpose. Like designing, the socio-legal perspective on the Rule of Law focuses on constituting specific measures for a particular goal. In the end, the measures applicable to reducing arbitrary use of power all come done to ensuring a balance of power between involved and affected actors (Mak & Taekema, 2016). Accordingly, this thesis focuses on ensuring a balance of power in design practices.

2.6 Theoretical and analytical lens

This section provides a short overview of the unit of analysis, the analytical lens, and the theoretical lens presented in this chapter. Following from our relational perspective on designing, this research's unit of analysis is the interaction between designers of public algorithmic systems. More specifically, this thesis focuses on the institutions that structure these interactions. To study these institutions, we use the IAD framework by Ostrom (2005) as analytical lens. This framework enables us to analyse the role of institutions in shaping design practices but also provides a structure for designing institutions. The IAD framework is an analytical lens and not a theoretical lens. It focuses our analysis on the specific variables relevant in collective action situations. As argued by Ostrom, researchers should adopt a theoretical lens to further contextualise the IAD framework (E. Ostrom, 2005).

Designing institutions for interactions between designers of public algorithmic systems is the central aim of this research. Therefore, we use a theoretical lens that can form the basis for the institutional interventions. The theoretical lens consists of three presuppositions and their corollary characteristics that apply to design practices of public algorithmic systems. First, the presupposition of socio-technical designing refers to the systemic, emergent, contingent, multi-stakeholder, and political characteristics of the design process of socio-technical systems. Second, the presupposition of democracy refers to considering designing as an iterative inquiry into means to achieve a particular end. This inquiry is steered by a public that consists of citizens who might be or are affected by public algorithmic systems. Finally, the presuppositions of the Rule of Law refer to a socio-legal understanding of institutional measures that reduce arbitrary use of power in the design process of public algorithmic systems through balancing power and installing argumentative practices. Apart from being a basis for Part III of this research, the theoretical lens also forms the basis for the analysis in Part I and Part II of this thesis (see Chapter 3).

Chapter 3

Research methodology

This chapter elaborates on the research methodology used for this thesis. Section 3.1 discusses the constructionist epistemology that underlies the research approach. Section 3.2 elaborates on the design science approach introduced in Chapter 1. The chapter closes with Section 3.3 by detailing the research methods we used for the three parts of this research.

3.1 A constructionist epistemology

Critical realism can be related to two epistemological perspectives that come back in our research: science as a social process, and pragmatism. First, scientific research is a social process according to critical realism. This is compatible with a constructionist epistemology that considers research as giving meaning to observations of reality (Moon & Blackman, 2014). In that process of meaning-giving, researchers bring their own ideas, experiences, and interests that they have formed during their life and training. Constructionism is close to interpretive or hermeneutic research (cf. Schwartz-Shea & Yanow, 2012), but we approach the epistemology from the perspective of abduction. Abductive reasoning combines both theories and hypotheses as well as surprising empirical insights to produce new knowledge. It enables researchers to explain and/or understand empirical results. Different from deduction it does not solely start from theory. Moreover, it rejects the inductive approach of neglecting an existing body of knowledge and fully relying on empirical data (Timmermans & Tavory, 2012). The creativity inherent to abductive reasoning also comes back in design science.

Furthermore, critical realism corresponds with a pragmatic or multimethod approach (Mingers et al., 2013; Moon & Blackman, 2014). In essence, pragmatism is about using the research method that is most effective in eliciting the knowledge one is looking for (Dewey, 1927/2016). Similarly, a pragmatic attitude enables researchers to use knowledge from different disciplines. For our empirical research, this means that we selected the ethnographic methods. The ethnographic methods are more suited and feasible for studying emerging phenomena (e.g., designing of algorithmic systems in public organisations). Moreover, we addressed the limitations of methods with triangulation – i.e., using different methods to elicit empirical data and corroborate findings. Finally, we scrutinised our own interpretation of empirical data by asking participants to reflect on our analyses.

The design science approach also aligns with a pragmatic research attitude. The choice for the design science approach stems from my training as an engineer, which has conditioned me to solve problems through design. For me, design science is a way to keep research close to practice. It enables a researcher to produce knowledge in practice together with involved actors. Through designing artefacts and testing them in practice, a researcher is able to test whether recommendations, ideas or solutions work as expected in reality. Apart from providing and scrutinising solutions, design science also facilitates problematising a current situation and proposed interventions in that situation. After all, the approach emphasises the importance of thorough problem formulation. For me, problematising prevalent ideas and practices is also a reaction to the tendency within public organisations to jump at new technologies without profoundly understanding what such technologies mean for the core processes of public administration.

3.2 Research approach

This thesis formulates a design theory on institutional interventions in the design process of public algorithmic systems. This design theory is the outcome of a design science effort. Design science is mostly associated with producing knowledge by creating artefacts, but it can also be used to formulate a design theory (see Gregor & Hevner, 2013; Kuechler & Vaishnavi, 2008). This section discusses our design science approach (Section 3.2.1) and explains our interpretation of a design theory (Section 3.2.2).

By discussing design science, we introduce yet another use of the word design. Scientific disciplines make extensive use of 'design' when discussing research methodologies or research activities. The field of 'design studies' conducts research on the act of doing design. Another discipline using the word design is that of 'design science' or 'research through design'. In these disciplines, design is used as a way to conduct research. To support this type of research, the discipline of 'design science research' theorises how design science should be executed (Van Aken, 2007). We only make use of and only refer to design science.

3.2.1 Research approach: design science

Section 1.3.1 motivated our choice for design science as our research approach. This section will elaborate on our interpretation of design science and the way we used it to gather knowledge about design practices of public algorithmic systems. Design science is a way of conducting research based on gaining knowledge through structured trial and error while creating an artefact (Gregor & Hevner, 2013; Gregor & Jones, 2007; Hevner et al., 2004; Peffers et al., 2007). In other words, knowledge is produced by manifesting a situation in practice through interventions. In our case, we manifest new design practices within public organisations. The closeness to practice is another distinctive characteristic of design science. Design science projects either focus on specific problems encountered within organisations, use co-creation with practitioners as methodology, or evaluate the designed artefacts in practice (Hevner et al., 2004; Sein et al., 2011).

Design science combines relevance, rigor, and creativity, which makes it suitable for research that addresses new phenomena or problems that emerge in practice. The combination of relevance, rigor, and creativity is coming back in the three cycles of design science (Hevner et al., 2004; Hevner, 2007), see Figure 3.1. Hevner et al. (2004) identify three fields of research in design science projects: contextual environment, knowledge base, and design activities. Each of these fields is related to one of the three cycles. In the relevance cycle, the problem space for a design is demarcated by empirically analysing the contextual environment of the studied problem. The relevance cycle embeds the research in practice. In our research, the contextual environment consists of the current design practices for algorithmic systems in public organisations.

The rigor cycle provides a scientific foundation for the research. It ensures that applicable theories and methodologies are appropriately used for creating an artefact. Moreover, using scientific insights prevents reinventing the wheel. The applicable scientific foundation for a particular design science project is called the knowledge base. It comprises the state-of-the-art knowledge related to the current and desired situations studied, but also about the artefact that is being designed. The knowledge base for this research focuses on theories on socio-technical designing, institutional theory, and theories on democracy and the Rule of Law (see the theoretical lens presented in Chapter 2).

Both the relevance as well as the rigor cycle form the input for the design cycle. This cycle follows Simon's (1969/1996) conception of design as iteratively generating alternatives and testing those alternatives. In this cycle, the design is created and justified through evaluation. The knowledge produced and lessons learned in the design cycle are brought back to the environment and knowledge base, which closes both the relevance as well as the rigor cycle.

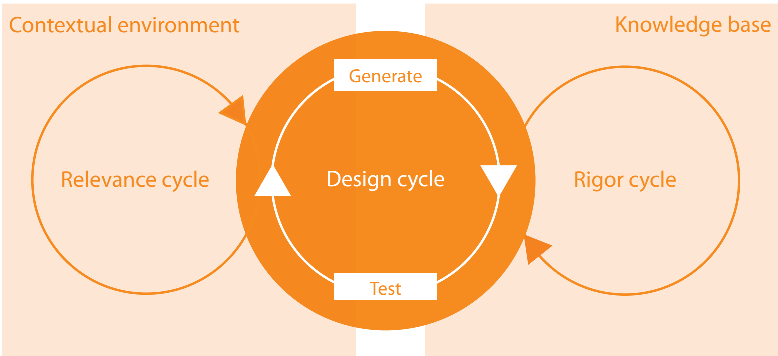


Figure 3.1 Three-cycle model of design science (based on Hevner et al., 2004; Hevner, 2007)

Apart from Hevner et al.'s (2004) conceptualisation of the design science process, others have described the course of such a process. Peffers et al. (2007) provide a step-by-step template of the process which can be seen as a more detailed description of Hevner et al.'s approach. However, Peffers et al. (2007) do not explicate the abductive reasoning fundamental to design science, instead they describe a fixed research process that obscures the connections between the three cycles. Moreover, Sein et al. (2011) prescribe a process in which design science and action research are combined. In our research, we did not have the opportunity to fully immerse ourselves in the field and, for example, design institutional interventions in a specific public organisation that would also be implemented in that organisation. The time in this PhD project was too limited for action research and public organisations were often reluctant to engage in research because of sensitivity of the topic. Therefore, the action research element could not be attained, nor was desirable as it might have limited a critical analysis of current design practices. Another reason to choose for Hevner et al.'s (2004) approach is that it explicitly refers to design theory as a possible outcome of a design science project (see Gregor & Hevner, 2013).

Hevner et al.'s (2004) approach has been criticised for focusing too much on (solving) business needs, and on designing IT artefacts or technology (Carlsson et al., 2011). Both focal points are contrary to the research aim of this thesis. Notwithstanding, design science is grounded in a pragmatic epistemology, and we approached it from a pragmatist attitude. We used the three-cycle model to structure our research approach, but within these cycles, we integrated our own research focus. Accordingly, we have shifted the focus on IT artefacts to institutional artefacts prescribed in a design theory (see Gregor and Hevner (2013) for a discussion on the broader application of design science) and emphasise the socio-technical nature of public algorithmic systems and their design practices. In addition, design science is no longer exclusively used in IS research but has been adopted by other disciplines such as organisational studies or public administration (Meijer, 2025; Romme & Meijer, 2020; Van Aken, 2005a). Moreover, instead of solely focusing on solving business needs, we complement our insights into current design practices with normative theories on socio-technical designing, democracy, and the Rule of Law. This enabled us to critically reflect on needs expressed by public organisations.

3.2.2 Design theory in design science

As discussed in Section 1.3.1, we performed our research in the context of Dutch public organisations. Similarly, institutions that structure practices are highly contingent. Nonetheless, it is possible to identify universal rules in prevalent action situations (cf. Ostrom, 2005). This thesis aims to prescribe general institutional interventions in design practices of public algorithmic systems. To arrive at generalisable research output, we adopted the following approaches in the three design science cycles. In the relevance cycle, we deliberately studied a set of public organisations situated in different government tiers. The differences between the studied public organisations enabled us to gain insight into context-independent characteristics of design practices. The rigor cycle inherently bases a design science project in theories which application transcends the studied design problem. Finally, we chose to have a design theory as a deliverable of the design cycle.

A design theory prescribes how an artefact can be designed, but does not define the artefact itself (Gregor, 2006). These artefacts can be physical (i.e., technological artefacts) but can also concern constructs, models, processes and methods (i.e., institutional artefacts) (March & Smith, 1995; Gregor & Jones, 2007; Van Aken, 2005b). A design theory has a distinctive focus on prescription. It provides actors with instructions on how to address or approach a class of problems (Walls et al., 1992). Accordingly, this research provides prescriptive instructions on how to design institutional interventions. Here it is important to stress that a design theory is not a normative theory (Gregor & Jones, 2007; Walls et al., 1992). It proposes an approach but leaves room for interpretation and the availability of other suitable approaches. At the same time, a design theory is contingent and value based. This is mostly because design science heavily depends on design choices by actors involved in a research project and by the contextual environment in which the project takes place. The formulation of a design theory through bottom-up development of concepts in a particular context and its relation to specific constellation of space and time aligns with our critical realist ontology and constructionist epistemology[1].

1 Gregor and Jones (2007) also relate design theories to critical realism. Moreover, design science is a pragmatic approach, because of its focus on problem formulation and creating interventions that address that problem (Van Aken, 2007).

Our design theory consists of design principles that prescribe institutional interventions embedded in a meta-theory on socio-technical designing in a democratic and Rule of Law context. The design principles form the focal constructs of the design theory. Design principles can be rigid and prescribe a specific solution, but they can also facilitate actors in finding an adequate solution for a specific context (Van Aken, 2005b). Our design principles fall in the latter category. Accordingly, our design principles are heuristics that public organisations can use to realise institutional interventions that bridge the gaps between existing and desired design practices.

Different forms of design principles have been proposed in the literature (Gregor et al., 2020; Reymen et al., 2006; Van Aken, 2004). Although being distinct proposals, the form of design principles has the following general structure. A principle consists of a particular goal and a mechanism to achieve that goal. Accordingly, this thesis formulates design principles as: *Establish [mechanism X] in order to achieve [goal Y]*. The design principle can be evaluated by testing to what extent the mechanism realises the stated goal.

A design principle should also be situated in a context and be provided with an underlying rationale (Gregor et al., 2020). We do this by embedding the design principles in a meta-theory (Love, 2000). Such a meta-theory elaborates the assumptions behind the design principles. The meta-theory is based on presuppositions of socio-technical designing, democracy, and the Rule of Law. Finally, the design principle is assigned to an actor that can use the principles for realising artefacts (Gregor et al., 2020). In this case, we foresee that collective choice action situations (see Section 2. 4.2) in public organisations will situate and contextualise our design principles.

3.3 Research methods

This section discusses the research methods used to answer the three research questions. Accordingly, it elaborates on Section 1.3.3.

3.3.1 Diagnosing current design practices

We used empirical research to identify current design practices and their underlying presuppositions. Through this empirical research, we answered sub research question 1:

What presuppositions underlie the design practices for public algorithmic systems that have emerged in public organisations?

First, we performed explorative research at a consortium of 13 Dutch public organisations working on policy instruments for obtaining public control over public algorithmic systems. By performing observations, interviews, and a document analysis, we elicited common design practices in the organisations and scrutinised the policy instruments developed by the consortium. Thereafter, we conducted explanatory research in which we interviewed designers of algorithmic systems within four Dutch public organisations. The results of this explanatory study complement the results of the study performed at the consortium.

Explorative empirical study

The study performed at the consortium started with observations of their meetings, and studying the documents that they produced and exchanged. Thereafter, we interviewed the project leaders of the consortium to elaborate on our insights and check our assumptions. Before discussing the observations, document analysis, and interviews in detail, we will discuss the research context of the explorative empirical study.

Research context: consortium

We performed the explorative empirical study at a consortium of 13 public organisations working on policy instruments aimed at reshaping the design processes in their organisations. The collaborative program started at the beginning of 2021 and lasted one year and three months. The consortium comprised five municipalities, three provinces, four executive agencies, and was supervised by the Ministry of the Interior and Kingdom Relations. Although the consortium included representatives of most types of governmental organisations in the Netherlands (e.g., water boards were not represented in the consortium), the member organisations were not representative of the attitude towards and advancements in governing algorithms among all Dutch public organisations. At the time, the consortium partners were considered frontrunners in addressing the problems associated with algorithmic systems compared to other public organisations. For example, smaller municipalities with fewer resources to tackle these problems were not participating in the consortium. Notwithstanding, the consortium was formed with the aim to provide first blueprints for policy instruments for public control over algorithms that would be used in all Dutch public organisations.

The policy instruments of the consortium were aimed at implementing public control over algorithms[2] and thereby contributed to filling the institutional void in design processes. They worked on four policy instruments: an algorithm register, algorithm procurement conditions, information requirements for objection procedures, and a governance document combining legal, ethical, and technical frameworks to assess developed systems. While working on the policy instruments, consortium participants were confronted with current design practices in their organisations. In the consortium meetings, the participants discussed their views and those of their colleagues on algorithmic systems, the design process, and challenges therein. The consortium participants were mainly policymakers responsible for algorithmic governance in their own public organisation. Sometimes, legal experts joined the consortium meetings.

The main aim of the consortium was to develop policy instruments that would strengthen public control over algorithmic applications. The consortium was convinced of the urgency to attain grip over the development and use of algorithmic applications and endorsed the need to approach this collaboratively. In its program plan, the consortium stated that the instruments should 'determine conditions, provide overview, and support interventions if needed; in order to provide citizens, administrators, and civil servants with more information, understanding, and influence [trans.].' The choice for the algorithmic governance framework, the algorithm register, procurement conditions, and instructions for information provision in objection procedures was mostly guided by the fact that organisations in the consortium worked on

2 The definition used by the consortium had the same scope as the definition of algorithmic applications in this dissertation: 'software that executes automated predictions, decisions or advice by using data analysis, statistics or self-learning logic which result in direct impact for citizens, companies, or physical assets in the public domain [trans.].'

these instruments before the start of the consortium. The deliverables of the consortium were planned to be 'minimal viable products', which could serve as a starting point for follow-up policymaking processes and regulating processes to establish public control over algorithmic applications on larger scales.

Observations and documents

We started our study by observing consortium meetings and analysing the documents that the consortium produced. We observed a total of ten online consortium meetings in which they were discussing the progress of the policy instruments. We observed eight meetings of the consortium's core team, consisting of policymakers responsible for both organising the design process within their own organisations as well as leading the development of policy instruments of the consortium. In these meetings, the project leaders, representing two municipalities and two provinces, shared the progress of the instruments and shared insights from within their own organizations. The other two observed meetings focused on the development of the governance framework and included participants from all involved public organisations. We collected statements by participants, related to current practices in their own organisations and how this affected the development of the new instruments, in a logbook.

The consortium described the developed policy instruments in separate documents. These documents provided information about the formal structures that public organisations had laid out for the design process. As such, a document analysis was used to complement the data from the interviews and observations, and to minimise biases (Bowen, 2009; Mackieson et al., 2019). On the other hand, as we observed in our studies, discrepancies exist between the design process described in documents and its enactment in practice. As discussed in Chapter 5, the design process as presented in policy documents did not reflect the manifestation of these processes in practice.

A document analysis is a research approach that is less influenced by participants compared to interviews or observations and can provide stable and exact data with coverage (Bowen, 2009). A rigorous selection of documents is critical, because documents available may not fit the research questions, documents needed can be hard to achieve, and there are other forms of selection bias (cf. Bowen, 2009). The risk of selection biases can be minimised through a structured selection of documents (Mackieson et al., 2019). In our case, we selected all documents produced by the consortium for analysis since we were examining the work of the consortium. The documents used were the program plan of the consortium and the four documents that presented the deliverables of the consortium. In the interviews, we dedicated approximately half of the interview time to addressing questions that were still unanswered after the observations and document analysis.

Interviews

To complement and triangulate the data gathered from observations and documents, we performed semi-structured interviews of 90 minutes with each of the four project leaders. These interviews can be considered expert interviews because we only interviewed individuals in their role as public servants. There are several perspectives on what an expert is, but we followed Pfadenhauer's (2009, p. 83) definition of an expert: '[a] person who has privileged access to information and – moreover – who can be made responsible for the planning and provision of problem solutions'. We considered the project leaders as experts, because they were responsible for the policy instruments and they were directly involved in restructuring their own organisations' approach to the design of algorithmic systems. Three interviewees were policymakers, and one interviewee was the team lead of the data science department of his public organisation. The latter was responsible for the governance of algorithmic systems within his public organisation and brought in the engineer perspective into the consortium since his team also included software engineers.

The interviews were semi-structured and based on a topic guide. The topic guide provided structure and enabled comparison between statements of different project leaders. At the same, time the topic guide provided room to explore interesting statements by interviewees that did not exactly fall within the pre-defined topics. The topic guide of the interview is presented in Appendix A2. The topic guide was structured following the IAD framework in order to cover all elements of design practices.

The interviewees were asked to reflect on the design process within their own organisation in general and reflect on the policy instruments of the consortium. In doing interviews, the interviewer is dependent on experience, perceptions, and interpretations of interviewees. The answers of interviewees can be influenced by response bias – i.e., providing answers that the interviewee expects the interviewer is looking for. Interviews will never be free from such response bias, but we tried to mitigate such response bias by telling interviewees at the start of the interview that there are no wrong answers. In addition, we asked follow-up questions when we suspected a response bias. For example, we asked interviewees to provide an example. By giving an example, interviewees have to make their answer more concrete. Moreover, comparing interviews from the same design team helps to identify response biases. Another characteristic of interviews is that reflections and evaluations by interviewees are partly captured in tacit knowledge. To elicit such tacit knowledge, we followed a strategy proposed by Meuser and Nagel (2009) to obtain knowledge that experts themselves are not aware of. We encouraged interviewees to speak in narratives, provide examples, and talk about their activities step-by-step.

Towards design practices

We coded the transcripts and logbook based on the IAD framework. First, we used the descriptions of design processes by interviewees and observants to identify action situations in these design processes and identify the institutions and attitudes that structure these action situations. Participants discussed a total of ten patterns of interaction that occur in their participating organisations. By comparing the different public organisations, we could generalise these to four patterns of interaction. Statements on common, recurrent, or habitual approaches to design activities were coded as action situations.

After eliciting action situations, we derived the institutions and attitudes that shape these action situations (see Section 2.4.1). For the institutions, we coded all established and embedded norms and rules that define permitted, prohibited, or required actions and/or outcomes discussed by research participants. These include both rules that are formally established as well as rules that lack official legitimation but are still shared and followed by designers. For attitudes, we coded all fragments in which participants and interviewees discussed attitudes, beliefs, interpretations, or values that they observed among designers in their own organisations. We used sub-codes to differentiate related but context-specific action situations, institutions, and attitudes discussed in the observed meetings or the interviews.

We derived presuppositions underlying the elicited institutions and attitudes through an abductive analysis. We compared the identified action situations, especially the institutions and attitudes structuring those action situations, with paradigms in public administration, digital government, and policy design discussed in the literature. Eventually, we used literature on technocracy and New Public Management (NPM). We considered the core elements of those paradigms and explored whether the identified institutions and attitudes matched these elements. Hultin & Mähring (2014) show that different presuppositions can coexist. Accordingly, we did not necessarily search for one all-embracing presupposition and were receptive to the possibility of multiple, coexisting presuppositions that structure the identified institutions and attitudes.

Scrutinising policy instruments

Our interview and observation data also provided insights into the policy instruments developed by the public organisations. We assessed the extent to which the instruments achieved the goal set by the consortium, i.e., public control over algorithmic systems. The analysis was executed as follows. Fragments in the observatory records and interview transcripts were assigned to one of the four instruments. These were compiled together with the documents that described the final version of the instruments. In addition, we compiled fragments including statements about the instruments in general. The fragments and documents provided insights into the contributions and challenges of policy instruments in attaining public control over algorithmic systems. The assessment of the instruments was based on the theoretical perspective presented in Chapter 2.

Explanatory empirical study

Since the exploratory empirical study mostly focused on policymakers who are responsible for setting up the governance of algorithmic systems, we wanted to expand our empirical research by studying design teams that actually performed design activities related to public algorithmic systems. Since the study of these design teams follows up on the exploratory study, we refer to it as our explanatory empirical study.

Research context: four public organisations

For the explanatory study, we selected four Dutch public organisations. In this study, we defined public organisations as organisations that execute public services and that are responsible for decision-making in public administration. The selected organisations had to be responsible for administrative or executive tasks related to public algorithmic systems. In other words, the public organisation is the actor deciding to implement or use an algorithmic system. Policies, laws, and regulations that form the basis for using the algorithmic applications could be formulated by other public organisations. Moreover, the public organisations did not have to design the whole system in-house. They could involve external parties to perform specific design activities but must lead or manage the overall design process.

In the Dutch context, institutional and algorithmic practices within public organisations vary considerably. Even public organisations at the same governmental level have different organisational cultures and structures, and different approaches to political decision-making. The public organisations in our sample represented this diversity in order to gain insight into commonalities in the approaches to designing. The selected organisations were comparable regarding their responsibility for starting and leading a design process but differed in size and type of decision-making (e.g., compared to provinces, municipalities typically make decisions that directly affect the lives of citizens).

Similarly, public organisations deploy different types of public algorithmic systems. At the time of conducting this research, larger municipalities and executive agencies used more complex algorithmic applications than smaller municipalities and provinces. We only asked public organisations that work with public algorithmic systems (as defined in Section 2.2.2) to participate in the explanatory study. Accordingly, we asked public organisations to provide a case in which a system was designed that included (1) a rule-based or data-driven algorithmic application that (partly) automates, augments, or supports decision-making; (2) that this decision-making should have impact on individual citizens; (3) that the algorithmic application is an integral part or extension of policy within the public organisation; and, (4) that the algorithmic system carries a risk of contributing to or creating Kafkaesque situations. If public organisations had designed a public algorithmic system that satisfied these four criteria, we asked the organisations to provide us with a group of executive designers who participated in the system's design process.

In the end, we selected two (large) municipalities, one province, and one national executive agency. As such, the participating public organisations reflect the organisations involved in the consortium. Like the sample in the exploratory empirical study, the set of public organisations represents three types of public organisations in the Netherlands that are working with algorithmic systems. Again, considering the size of the municipalities, provinces, and executive agencies in general, our sample of organisations does not represent smaller public organisations with limited resources, such as small municipalities. Table 3.1 provides an overview of the selected public organisations. Organisations A, B, and C were members of the consortium, public organisation D was not.

Table 3.1 Public organisations included in the explanatory empirical study and the public algorithmic systems mentioned as examples in the interviews for the specific public organisations

	Public organisation A	**Public organisation B**	**Public organisation C**	**Public organisation D**
Type of organisation	Municipality	Municipality	Province	Executive agency
Type of algorithmic application	Rule-based	Data-driven	Rule-based	Data-driven
Focus of public algorithmic system	Check address quality in public registries, to support fraud detection	Selection of recipients of benefits who will have a meeting with caseworker	IdentificatioWn of geographical areas that could be used for sustainable energy policies	Provision of information on what cases to examine/sites to visit
Technical designer	External party	Positioned in data science lab	External party	Positioned in data science lab
Policy domain	Social domain	Income benefits	Spatial planning	Inspectorate

Interviews and reflective workshop

The explanatory study started with performing one-to-one expert interviews with designers. Participating designers had to meet the following criteria (which align with our definition of a designer, see Section 2.3.1). First, the interviewee had to have contributed to the design of the technical artefacts, the institutional artefacts, or both types of artefacts in the public algorithmic system. Second, the interviewee had to have prior experience in designing public algorithmic systems. We asked public organisations to select designers who were involved in at least two design processes. Third, designers who were employed by other organisations but contributed to the design process were also invited for an interview. It is common practice within public organisations to involve external designers in designing public algorithmic systems. Moreover, these external designers provide a different perspective on practices within public organisations compared to internal designers.

We asked public organisations to provide at least one of each of the following types of designers. The types of designers represent different components in the socio-technical system. First, a designer who worked on the technological component, for example, an information architect, data analyst, or data scientist. If possible, public organisations were asked to propose both one designer working on the algorithmic application as well as one working on the broader information system. Second, a designer of institutions that form the basis of the algorithmic system, such as domain experts or policymakers in a specific domain. Third, designers who constitute frameworks and guidelines for algorithmic systems in public organisations, for example, legal advisors, security officers, or privacy officers. Finally, designers who work on institutions to guide the operation of algorithmic systems, for example, a policymaker developing policies, protocols, or processes for implementing or using the algorithmic applications. The last three types of designers represent the institutional and agential components in public algorithmic systems. Table 3.2 shows the interviewed designers within each public organisation.

Table 3.2 Overview of the interviewees in the explanatory study and the role that interviewees fulfilled in the design process

	Public organisation A	**Public organisation B**	**Public organisation C**	**Public organisation D**
Technical artefact – algorithmic application	Data engineer (external)	Data scientist	(ICT) developer (external)	Lead data scientist
Technical artefact – information architecture	N/A	Product owner (data-analyst)	Information architect	Business intelligence specialist
Institutional artefact – domain-related policy	Advisor fraud detection	Manager (business)	Advisor/ coordinator policy monitoring and evaluation (external)	Inspector (business)
Institutional artefact – domain-related policy	Policy coordinator	Policy leader (business)	N/A	Project leader (business)
Institutional artefact – compliance	Privacy officer	Project advisor (business)	N/A	Legal advisor (focus: privacy and data exchange)
Institutional artefact – algorithmic governance	N/A	Project leader data ethics	Strategic advisor technology, society and ethics	Advisor data science (data scientist)

Organisations A and C were not able to include two roles in the interviews. Notwithstanding, the designers interviewed within these organisations still cover the set of designers we wanted to interview. Public organisation A could not provide an information architect, but the technical artefact was still covered. The data engineer was also involved in an information architecture project of the municipality. We already interviewed a policymaker focusing on algorithmic governance for this public organisation in the exploratory study. For public organisation C we only interviewed one domain-related policymaker, whereas other organisation provided two of these participants. We did not interview a compliance officer of public organisation C, but the insights from these types of designers were partly covered in the interview with the strategic advisor.

Like the interview topic guide of the exploratory study, the protocol for this study was based on the IAD framework. The topic guide can be found in Appendix A3.

The interviews, observations, and documents in both empirical studies provided rich insights into the design practices of public organisations. To reduce bias in the interpretation of the empirical results, we performed a preliminary analysis of our data and presented the results to participants of the explanatory empirical study. The preliminary results shared during the workshop can be found in Appendix A5. These workshops were used to gain new insights, enrich already gained insights, and check assumptions. We did the workshops approximately half a year after the interviews were conducted. We asked interviewees from public organisations B and D to participate in the workshop; organisations A and C were not available. In the workshop with organisation B, the product owner, the project advisor, and the project leader on data ethics joined the workshop. In addition, two algorithm experts and one data advisor who did not participate in the interviews joined the workshop. The workshop at organisation D was joined by all interviewees except by the lead data scientist. In addition, we presented the results to an internal consultancy organisation within the government. This consultancy organisation consists of data scientists who support national, regional, and local Dutch public organisations in developing algorithmic systems. They work for a diverse range of public organisations and, therefore, have insight into the design processes in different organisations. Five consultants joined the workshop.

Elaborating on design practices and presuppositions

The interviews with designers were analysed through thematic analysis (Braun & Clarke, 2006). We followed the two-cycle coding approach described by Saldaña (2013). This method aligns with the IAD framework as it provides room for focusing the coding on the variables in the framework. At the same time, this coding method enabled us to focus on interactions between designers without using a pre-defined theory on these interactions. The code book can be found in Appendix A4.

The first cycle of coding focused on assigning codes to (fragments of) statements by interviewees. We used four types of coding. First, we used structural coding to assign larger pieces of text in the transcripts to general questions in the topic guide of the interviews. These structural codes provide a basis for the more detailed process, descriptive, and value codes. Second, we used process coding to indicate all actions that interviewees mentioned. Process codes are phrased as a verb and signify design activities or interactions. Third, we used descriptive coding to assign topics to transcript fragments. Descriptive codes do not assign a theme but only indicate the topic of a text fragment. Though these codes we were able to relate statements to variables in the IAD framework. Finally, we used value coding to indicate attitudes and beliefs of interviewees related to how they think about the design process, their own role, or the role of other designers.

The second coding cycle generally focuses on extracting themes from the first-cycle codes. However, in this study, we used the second cycle to connect the first-cycle codes to our findings in the exploratory study. We extended the exploratory results by looking for similarities and contradictions. Again, we focused on design practices, institutions and attitudes, and presuppositions.

3.3.2 Appraising democratic and Rule of Law practices

After gaining a more in-depth insight into the problems in design processes of public algorithmic systems, the research continued by establishing the appropriate knowledge base. This part of the research also started a first design cycle based on abductive reasoning. Accordingly, we answered research question 2:

What design practices that curb algorithmic Kafka are prescribed by the synthesis of the presuppositions of socio-technical designing, democracy, and the Rule of Law?

Compared to scientific approaches that focus on deriving explanations through induction or deduction, the iterative and creative nature of design implies a different approach to conducting scientific research (Hevner et al., 2004; Simon, 1969/1996). Instead, design science is based on abductive reasoning (Kolko, 2010; Koskela et al., 2018). Abductive reasoning is an inherently creative process in which a researcher compares empirical results with theories (Timmermans & Tavory, 2012). Abduction encompasses two steps (Kolko, 2010). First, researchers make sense of empirical results by understanding them from different theoretical perspectives. Thereby, the researcher searches for surprising anomalies, differences, and unexpected findings (Timmermans & Tavory, 2012). Second, the insights from sensemaking are synthesised to generate new knowledge and perspectives (Torraco, 2016).

Although abductive reasoning is also knowledge-intensive, the role of theories in abductive reasoning differs from that in inductive or deductive reasoning (Timmermans & Tavory, 2012). A researcher brings theory to empirical results to reframe those results. Most likely, this researcher brings theories familiar to them. This can introduce bias towards certain perspectives into the research. At the same time, bringing one's own perspective to research is inevitable (see Section 3.1). Notwithstanding, it is important in abductive reasoning to look for applicable theories that might be outside the researcher's knowledge. This was done by discussing this research and our empirical results with academic colleagues.

We reframed our empirical results in two ways. First, we examined how the design practices relate to the origins of Kafkaesque situations in socio-technical specifications of public algorithmic systems. Second, we explored how the socio-technical design practices of public algorithmic systems can be reshaped in order to align with democratic and Rule of Law presuppositions.

Secondary analysis of documented cases of harmful public algorithmic systems

We aimed to derive design practices that result in less, but preferably no, harmful public algorithmic systems. Therefore, the role of the socio-technical specification (i.e., the object of design) in creating these harms needs to be known. In order to do so, we examined how Kafkaesque situations can emerge from public algorithmic systems, and how that emergence can be related to design choices. Therefore, we performed a secondary analysis of notorious and studied cases from different jurisdictions in which public algorithmic systems inflicted harm on citizens. This secondary analysis is based on the descriptions of these cases in scientific and grey literature. We did not gather empirical data on these cases ourselves.

For the selection of cases, we mostly aimed for cases that are documented in scientific literature. We used the following selection criteria:

- The case should include a public algorithmic system that harmed citizens – i.e., algorithmic Kafka.
- At least two analyses of the public algorithmic system are available in official reports, conference proceedings and/or scientific journal.
- The sources on the case should cover the harms caused by the use of a specific public algorithmic system or even theorise the causes of these harms.
- The cases in the selection are situated in different jurisdictions.

We selected two notorious cases: the Dutch childcare allowances scandal and the Australian Robodebt scheme. These cases had wide coverage in media but also have been studied by several scholars. We included two more cases for the purpose of saturation. The British Post Office scandal is an infamous case of an ICT system that created Kafkaesque situations for sub post masters. The case has been studied in scientific literature; however, at first sight, the IT system in this case is quite far from our definition of algorithmic systems. Nevertheless, the system is a form of automation that clearly preceded the current wave of algorithmic systems. Moreover, its implementation was based on similar grounds of efficiency, etc. In addition, we study the Dutch DUO case in which students were falsely accused of fraud with study grants. This case has only been studied in grey literature.

The selection of cases also shows the limitations of a secondary analysis, because it fully depends on the analysis of other scholars. This means that we could only select cases that had already been studied, and that we relied on the perspectives of these scholars on the harmfulness of the algorithmic systems. To mitigate this bias, we used the selection criteria that at least two articles should be available. Moreover, all four selected cases have had widespread coverage by journalists and were picked up by politicians. Consequently, we did not include cases of algorithmic Kafka that did not gain the same traction as the selected cases. Considering the limitations of our approach, we restricted our analysis to examining the role of socio-technical specifications in the studied cases. Moreover, we corroborated our findings with literature on harms in algorithmic systems and on algorithmic Kafka.

The secondary analysis was done through abductive reasoning, more specifically, through sense-making. We used the lens of Kafkaesque situations to interpret how public algorithmic systems inflict harm on citizens. In other words, we focused on what is to be prevented, mitigated, or corrected within the socio-technical specification of public algorithmic systems to curb algorithmic Kafka. The lens of Kafkaesque situations was based on public administration literature that specifies these situations in classic bureaucracies. By interpreting our case descriptions through the Kafkaesque situations lens, we were also able to connect the emergence of algorithmic Kafka to our socio-legal perspective on the Rule of Law. Algorithmic

Kafka is related to arbitrary use of power, which is a central concept in the Rule of Law. From the perspective of arbitrary use of power, we were able to identify interactions between designers that impede addressing possibilities for Kafkaesque situations in the socio-technical specification of public algorithmic systems.

Synthesis of socio-technical designing with democracy, and the Rule of Law

The analysis of harmful public algorithmic systems determined points of intervention into design practices – i.e., problematic interactions between designers. The second study in Part II examined what design practices for public algorithmic systems are desired, and what institutional interventions could be used to arrive at such practices. We arrived at these desired practices through synthesis, i.e., the second step in abductive reasoning. This synthesis resulted in a meta-theory for embedding socio-technical design practices of public algorithmic systems in a democracy and Rule of Law context.

Chapters 1 and 2 already provided the motivation for synthesising the presuppositions of socio-technical designing, democracy, and the Rule of Law. This synthesis is hampered by the fact that all three presuppositions are contested and dissimilar concepts. Contested concepts are concepts that are heavily used in a scientific discipline, but, at the same time, their meaning is heavily debated by the community related to that discipline (Collier et al., 2006; Gallie, 1956). Accordingly, when using contested concepts such as our presuppositions, a researcher needs to clarify their position regarding those concepts. Chapter 2 discussed our position towards the three presuppositions.

The synthesis approach is inspired by Hendriks (2022) who integrates two broad concepts: democracy and governance. Hendriks (2022, p. 7) stresses that to arrive at a comprehensive understanding of two value-laden concepts, the complementarities and frictions need to be considered: 'complementarities as well as frictions need to be acknowledged in any sensible synthesis, and are not to be erased by aggregation or abstraction.' Accordingly, before moving to the synthesis of the three presuppositions grounded in different scientific disciplines, we made the complementarities and frictions through juxtaposing the three concepts. We understand juxtaposing as comparing and contrasting dissimilar concepts. We distinguished symbioses and shared challenges in the complementarities between the three concepts. A *symbiosis* of presuppositions occurs when characteristics of one presupposition are shared and compatible with other presuppositions, and this intersection reinforces the functioning of the different presuppositions. A *shared challenge* is defined as characteristics, shared by the presuppositions, which are difficult to translate to practice. This also relates to the theoretical nature of the presuppositions. Finally, when characteristics of presuppositions are incompatible or result in tensions, we refer to a *contradiction*. Contradictions result in trade-offs that have to be made when synthesising presuppositions.

Using insights from the symbioses, shared challenges, and contradictions, we synthesised the three presuppositions into one meta-theory (see Section 3.2.2). We translated the symbioses, shared challenges, and contradictions to the practice of designing public algorithmic systems by combining them with the results from the preceding studies in Part I and II. Using the meta-theory, we could specify the desired interactions between designers. We did this by contrasting the interactions identified in Part I with the prescriptions in the meta-theory.

3.3.3 Designing institutional interventions

The last step of the research synthesised the findings from the diagnosis (Part I) and appraisal (Part II) into design principles for institutional interventions in design practices of public algorithmic systems. This step answered research question 3:

What institutional interventions engender interactions between designers of public algorithmic systems that align with democratic and Rule of Law principles?

We developed the design theory by following the generate-test cycle described in design science (cf. Hevner et al., 2004; Cash, 2018). The generation of design principles already started in the first two parts of this research. The answers to the first two research questions provided building blocks for the design principles. In Part III, we finished the design cycle. Thereby, we focused on making the design principles explicit and translating them to institutional interventions. We tested the interventions in a simulation of a design process. Before conducting the simulation at a public organisation, we conducted two trial simulations with academic colleagues.

Generation of design principles and instantiations

To formulate the design principles, we started by bringing all the learnings from the preceding two research questions together. We generalised current and desired design practices from Part I and Part II, respectively. For each gap between current and desired practices, we formulated a design principle for institutional interventions. Each design principle has the following form: *Use [mechanism] in order to achieve [goal]*. We used the IAD framework of Ostrom (2005) to identify what kind of institutions were needed to create the desired design practices (see Chapter 2).

After formulating the design principles, we translated them into practice by generating institutional interventions in the form of process instructions. This was an iterative process of making the design principles specific to the context in which the interventions would be applied (i.e., the public organisation at which we performed the evaluation of the design theory). As will be discussed below, the instantiations of our design principles comprised instructions and templates for designers. The interventions had to fulfil the following criteria: 1) enable the testing of the underlying design principles; 2) be comprehensible for simulation participants. We reiterated the interventions by conducting two trial simulations with fellow PhD candidates. Appendix B1 describes the changes made to the interventions in response to the trial sessions.

Explorative evaluation of the design theory

Evaluation is an important step in design science. Through evaluation, the choice for a certain design to address a formulated problem is justified, but it also provides feedback for new iterations in the generate-test cycle (Hevner et al., 2004). Scholars list different evaluation methods in design science. Hevner et al. (2004) distinguish observational, analytical, experimental, testing, and descriptive methods. Peffers et al. (2012) identify logical argument, expert evaluation, subject-based experiment, action research, prototype, case study, and illustrative scenario as possible evaluation methods. To select a suitable method for the evaluation of our design theory, we follow the four steps by Venable et al. (2016) for determining a specific evaluation process: 1) explicate the goal; 2) choose a strategy; 3) determine properties; and 4) design the individual evaluation episode.

Evaluation goal

First, we explicate the evaluation goal. Like evaluation methods, scholars have provided elaborate lists of evaluation criteria that relate to goals for evaluation. The list by Prat et al. (2015) is most detailed and comprises criteria such as efficacy, feasibility, usability, completeness, simplicity, consistency, and robustness. The goal of our evaluation was to test whether the institutional interventions in design processes result in the expected interactions between designers. Accordingly, we evaluate the efficacy of the institutional interventions that were generated based on the design principles.

Evaluation strategy

The second step is about choosing an evaluation strategy. Our considerations for choosing a strategy were:

- Following the goal of the evaluation, we wanted to examine the effects of institutional interventions on the interaction between designers. Therefore, these interactions had to occur and had to be observable in the evaluation.
- The evaluation could not happen in an actual design process because such design processes take several months or years. Both the time of the researcher (PhD trajectory) and the participating public organisation were limited. The evaluation strategy should fit the limited time of both parties.
- The circumstances in the evaluation had to be similar to the circumstances in a real design process. Accordingly, the circumstances in the evaluation needed to be controllable as far as possible.
- At least one public organisation had to be available that designs public algorithmic systems within their organisation.
- It had to be possible to upscale the evaluation to multiple public organisations.

Performing a case study in which public organisations are asked to implement the institutional interventions in their daily work was not possible because of the time restrictions. Therefore, we had to simulate the design process. The main idea behind the simulation was to ask practitioners (i.e., politicians and public servants performing design activities) to perform a design process for a public algorithmic system. The setting in the simulation resembled a real design process of a public algorithmic system as closely as possible but was adapted in accordance with the institutional interventions that follow from the design principles of Chapter 8.

We were only able to conduct the evaluation at one public organisation. Public organisations were reluctant to participate in the evaluation or were not able to join in the time frame of this PhD project. In the end, we collaborated with a Dutch medium-sized municipality (around 110,000 inhabitants). This means that an experimental design of the evaluation was not possible. Instead, we approached it as an explorative evaluation of our design principles. The results of this explorative evaluation mostly provide a first indication of the working of institutional interventions (i.e., process instructions) following from the design principles and can be used to improve our design theory. In conducting the simulation, we gathered qualitative data through surveys and observations. We used this data to gain insight into the extent that the logics behind the design principles can be observed in the simulation. The disadvantage of this approach is that the results are not generalisable to other public organisations. However, the evaluation provides rich insights on the institutional interventions. Moreover, the simulation can be conducted at other public organisations in the future; a comparison would be possible in that case.

Evaluation properties
The properties for evaluation are determined in the third step. Properties are aims that are to be achieved by an artefact. Consequently, properties are inherently dependent on the artefact that is designed. In this thesis, the evaluation properties are derived from the design principles in the design theory. For each design principle, we formulated expected interactions between designers. The properties that we used in our evaluation are based on these expected interactions. The evaluation properties are elaborated in Chapter 9.

Individual evaluation episode
The last step is about designing the individual evaluation episode. In general, our evaluation episode consisted of a questionnaire preceding the simulation of a design process, the actual simulation, and a questionnaire after the simulation. We asked participants to fill in the questionnaires before and after the simulation to gain insight into the effects of the institutional interventions.

The evaluation started with a questionnaire in which participants answered thirteen open questions concerning the current situation of designing algorithmic systems in their public organisations, see Appendix B2 for the questionnaire. We asked three types of questions. The questionnaire started with introductory questions about the role and experience of participants. Thereafter, participants answered questions related to the evaluation properties. The questionnaire ended with questions on what should be done differently or improved in current design practices according to participants. Since two types of designers – i.e., political and official designers – participated in the simulation, we adapted the questions to their own position within their public organisation. The introduction to the questionnaire included definitions of 'designing' and 'public algorithmic systems' to ensure that participants would use the same broad interpretations of these concepts as used in this thesis. The results of the questionnaire were compared to the results in Part I to identify biases in the answers of participants.

After the first questionnaire, we conducted the simulation of a design process of a public algorithmic system. For the simulation, we formulated a fictional design problem (see below) for which the participants had to design a public algorithmic system following the process dictated by our institutional interventions. We steered the designers as little as possible during the simulation in an attempt to only influence their interactions through our institutional interventions.

While the participants were running the simulation, two researchers observed their interactions. One researcher provided the participants with the instructions of the simulation and collected general observations concerning the interactions. The other researcher was fully committed to observing the participants and recorded detailed observations. As such, we used two observers to address researcher bias. The observer was not involved in the rest of the research project of this thesis.

As will be discussed in Chapter 9, the simulation consisted of four sessions. All sessions took place in person in order to reflect the conditions in real design processes. Moreover, in-person sessions enable the observers to observe interactions between participants that are not structured or disturbed by digital technology. The political designers came to the building of our faculty; the sessions with the official designers took place in the office of the public organisation. In all sessions, the participants were sitting at a table and were positioned in a circle. Both the instructor and the observer were present in the same room and took a more peripheral position towards the participants. They were able to see all participants. Despite having some distance from the participants, having the instructor and observer in the room can influence the behaviour and responses of participants. The observation of participants in a simulation introduces another form of bias. Participants are aware that they are observed and might change their actions and interactions accordingly. For example, during the simulation, we noticed that participants often asked the observers whether they were doing the assignment as intended. This form of bias cannot be prevented and is difficult to reduce. Again, this underlines that our evaluation has an explorative nature.

After the simulation, participants answered questions in a second questionnaire, see Appendix B4. The 12 questions in this questionnaire resembled those in the first questionnaire but focused on what happened in the simulation. By keeping the questions similar, we were able to make a comparison between design processes in practice and the design process in the simulation. Apart from reflecting on what happened in the simulation, we asked participants to reflect on what the interventions would engender if they were actually implemented in practice, and to assess the interventions. As the participants were confronted with the design output of other actors, we also asked participants to reflect on their own role in the simulation, and the role of other designers. Similar to the first questionnaire, this questionnaire is susceptible to response bias. We have tried to reduce bias by also asking participants to reflect on the simulation at the end of sessions 3 and 4. Comparing those reflections with the answers to the questionnaire can reveal biases.

Fictional design exercise

The fictional design exercise that participants in the simulation had to perform consisted of a problem description and a call to action. The design exercise had to meet the following three criteria:

- The exercise had to focus on a problem with both societal as well as administrative aspects to it, and for which its solutions will have consequences for the situation of individual citizens.
- The exercise had to incorporate data- or information-related issues, or issues that can be interpreted as such.
- The exercise could not prescribe goals for the system to be designed (such as efficiency); that is the task of the politicians.

To enable upscaling the simulation to multiple public organisations, we constructed a general structure for the problem that was to be addressed by participants. The general structure ensures that design exercises that are specified to a specific organisational context, can still be used to compare simulations in different public organisations. We used the following general structure for the problem description: *the public organisation needs to allocate support (in nature or in money) to citizens based on an indication of those who are eligible for the support.* First, this kind of problem is familiar to a wide range of public organisations that are active in different domains or topics. Specifying the problem to a specific topic (such as student grants, social housing, or elderly care) opens up the opportunity to conduct the simulation at other public organisations. Second, public organisations tend to associate this problem with solutions in the field of algorithmic applications. Finally, this kind of problem is less controversial compared to other common problems that are often addressed with public algorithmic systems, such as checking the legitimacy of allocating a benefit or calculating the distribution of resources. Our aim was to examine interactions between designers in the case that a problem or solution is not considered controversial from the outset.

The problem formulation in the performed simulation was geared toward the context and competencies of a Dutch municipality:

The national government has asked municipalities to execute the allocation of energy cost allowances. This energy cost allowance aims to support households that are no longer able to pay their energy bills because of increasing energy prices. It is the municipality's task to ensure that all eligible households receive the allowance. However, the municipality lacks an overview of all eligible households.

In addition to this problem formulation, participants were given available resources that they could consider in designing the public algorithmic system:

- *Financial resources to cover the costs of identifying eligible households*
- *Financial resources to execute the allowances program*
- *Data on income, household size, age, use of other allowances and/or benefits*
- *Insight into average energy use by households, trends of energy prices, trends in demand and supply of energy*
- *A policy document drafted by the national government that lists what types of households are eligible and the amount that these households are granted*

In practice, such a problem formulation would be drafted by the politicians or designers and be iterated in the dialectic between politicians and designers (see design principles in Chapter 8). However, for reasons of replicability, focus and time, we decided not to include the problem formulation in the simulation. The advantage was that participants concentrated on the activities related to making design choices. On the other hand, this meant that the simulation might deviate from the interactions that designers normally have in a design process.

Part I

Diagnosis

Chapter 4

Practices and presuppositions in current design processes of public algorithmic systems[1]

4

> … in calling for a government that 'steers more, rows less', it fundamentally failed to recognize how these two functions are related. The less a government rows, the less it learns, the less productive it becomes: the less it can steer. And when governments cease to deliver a function that still needs to be delivered, they struggle to govern its delivery. This view of government also ignored the shift of power that emerge when the government stops rowing, and hands over the oars to other actors.
>
> – *Mazzucato & Collington; The Big Con, p. 252*

Chapter 1 already listed several problems in current design processes of public algorithmic systems. System engineers have obtained a strong position in design practices within system-level bureaucracies (Bovens & Zouridis, 2002). Moreover, democratic control over and democratic legitimacy of design practices of public algorithmic systems is falling short (Grimmelikhuijsen & Meijer, 2022; König, 2020; König & Wenzelburger, 2021). At the same time, there is little empirical research on design practices for public algorithmic systems. This chapter aims to gain further understanding of the current design practices for algorithmic systems in public organisations. Therefore, it answers the following question:

What design practices for public algorithmic systems have evolved in public organisations?

This chapter presents the results of the exploratory and explanatory empirical studies described in Section 3.3.1. We used the IAD framework by Ostrom (2005) to structure the study of current design practices in public organisations. Accordingly, we focused on action situations in which designers interact with each other. We identified two action situations based on interview and observation data: one action situation in which public servants from different disciplines act as

1 This chapter draws on Nouws, S., Janssen, M., & Dobbe, R. (2022). Dismantling Digital Cages: Examining Design Practices for Public Algorithmic Systems. In M. Janssen, C. Csáki, I. Lindgren, E. Loukis, U. Melin, G. Viale Pereira, M. P. Rodríguez Bolívar, & E. Tambouris (Eds.), *Electronic Government* (Vol. 13391, pp. 307–322). Springer International Publishing. https://doi.org/10.1007/978-3-031-15086-9_20.

designers and coordinate design choices; and one action situation in which political steering on design choices takes place. Sections 4.1 and 4.2 describe both action situations respectively. These two sections follow the same structure. First, they discuss the positions within action situations. Subsequently, the patterns of interactions that emerge from the two action situations are discussed. We end both sections with listing the identified institutions and attitudes that structure the action situation.

The institutions and attitudes are points of intervention for the design theory formulated in Chapter 8. Therefore, Section 4.3 discusses four impediments to conducive interactions between designers found in institutions and attitudes that structure current design practices. Following up on the impediments, we abductively reason towards two presuppositions that form the basis for prevailing institutions and attitudes: a technocratic and a businesslike presupposition. Addressing these presuppositions provides an opportunity to reshape design practices on a more structural level. Section 4.4 presents the conclusion of the chapter.

4.1 Collaboration between disciplinary designers

When interviewees discussed the roles of and interactions between public servants performing design activities, they tended to distinguish three aggregate positions – each with their own 'jargon' name: the business, ICT, and compliance. Within these aggregate positions, different specific positions are 'clustered'. In general, these three positions interact as follows. The domain-specific departments of public organisations, which interviewees referred to as 'the business', commission a public algorithmic system which they will use in executing their tasks. Next, the algorithmic system is developed by an 'ICT' function or department, or by an external party. Finally, 'compliance' functions or departments ensure that the algorithmic systems are in line with rules and regulations. This section focuses on the action situation in which these three aggregate positions interact, as shown in Figure 4.1. First, we discuss the three positions – and the related actors, actions, information, and control – in the action situation. This is followed by an overview of three patterns of interaction in Section 4.1.2. This section concludes with examining the institutions and attitudes that structure the action situation. We will refer to *executive designers* when discussing public servants who perform design activities. The adjective 'executive' is chosen because public servants are positioned in the executive branch of government.

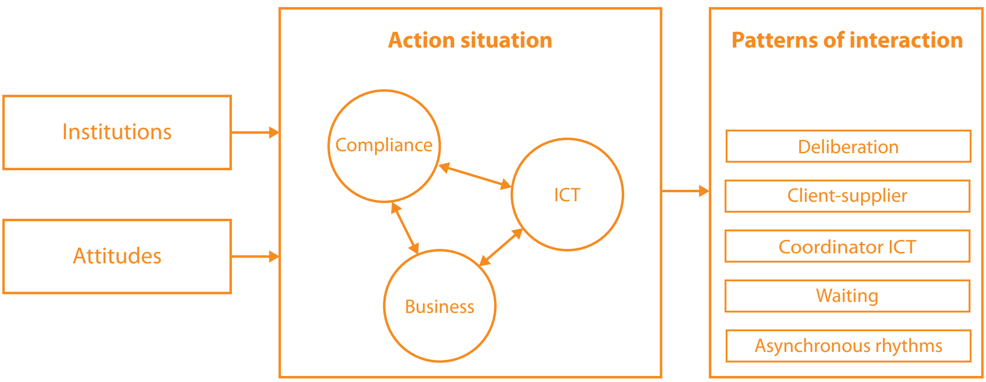


Figure 4.1 Action situation in design processes of public algorithmic systems that comprises the interaction between different disciplines of executive designers

4.1.1 Positions

'The business'

In most studied public organisations, 'the business', also called 'the client', comprises the departments responsible for the core tasks of the public organisation, such as policymaking or execution of public services. Typically, this position is filled by domain experts, often policymakers. These policymakers can commission a system and provide their situated knowledge about the context in which the algorithmic system is (to be) embedded. Although actors in this position have in-depth knowledge about their own domain, they largely depend on the knowledge of technical designers to create the information architecture for their executive tasks. Consequently, while the business theoretically is in control of defining the goals and boundaries of an algorithmic application, their influence on the algorithmic system's form and function is constrained by their lack of knowledge about what is technically possible and what is not. As one interview stated: 'a subset of people working with data often do have the knowledge. But, yeah, those people only form a small percentage of all people working at the organisation. All the others, yes, they do not have that awareness [trans.].'

'ICT'

The public organisations in this study developed algorithmic applications both internally and externally. When developed internally, this is done by traditional ICT departments, data analysis or business information teams, or more experimental data science teams (see also Lorenz, 2023). Actors within these teams vary from data analysts, data scientists, to information architects, business analysts, project leaders and product owners. In the studied cases, external developers were still organisationally embedded within the ICT 'silo' of the public organisations.

Officially, the ICT position executes the commission by the business. Generally, this starts with an intake in which technical designers refine the question or problem formulation brought in by the business. Apart from demand-driven development of applications or systems, some public organisations have set up more independent data science teams or labs to explore possibilities that algorithmic applications can provide for the organisations – i.e., supply-driven development. Technical knowledge on algorithmic applications is concentrated within the ICT position. In addition, through their project work, technical designers accumulate information about what happens in different parts of the organisation. Overall, technical experts have become more involved in policymaking in traditional policy domains over the years.

'Compliance'

The actors not concerned with either commissioning the system or developing the technical artifact are generally referred to as compliance. This position comprises privacy officers, security officers, legal advisors, and other actors who assess whether the (to be) designed algorithmic system complies with laws or regulations, or who evaluate the external effects of such systems. The actors within the compliance position bring in their own knowledge on privacy, law, ethics, safety, and security. Compliance officers increasingly have a say in developing and using public algorithmic systems, especially privacy officers, after the introduction of the GDPR. However, apart from the increase in policy instruments that bring risks of algorithmic systems into view, these actors often have few means to intervene and relatively little control over design choices in the design process.

4.1.2 Patterns of interaction

In terms of interactions, executive designers mostly collaborate, within and between positions, by deliberating, aligning, and seeking agreement on design choices with each other. Interviewees in the explanatory study characterise designing public algorithmic systems as a continuous deliberation between designers from different disciplines. The designers from different positions bring their own knowledge, expertise, perspectives, and experiences. Consequently, executive designers must inform each other, explore design alternatives, and agree on design choices by deliberating with each other. As one interviewee stated: 'you also notice in that kind of conversations, in which the ones responsible for the technology, execution of policy, and the policy itself really start a conversation … it shows the different perspectives of each party, or other perspectives than they had before [trans.]'.

Behind the consensus-oriented activities of deliberation, a hierarchy emerges between the different positions. First of all, most public organisations have structured the interaction between the business and ICT as a client-supplier relationship. The business – in its role as the 'client' – commissions an algorithmic application from ICT – being the 'supplier'. Although the 'client' makes the final decisions on the algorithmic system, they lack the knowledge about developing, assessing, implementing, and governing these systems. As a result, many (normative) choices are made at the developers' level, as we will elaborate in the next paragraphs. The 'client' often also lacks knowledge about the responsible use of algorithmic systems but can ignore compliance officers in their role as the main decision-maker.

Within this client-supplier relationship, ICT naturally takes up a coordinating role. Interviewees provided examples of the leading role of ICT. For example, ICT provides products for several domain-related departments that are strictly separated. ICT is naturally put in a position in which they become aware of possible connections between different departments. One interviewee stated concerning the interactions between departments and their role as developer: 'The product owners [i.e., domain department] mostly do not know each other, they do not actively exchange [information about their projects and insights]. Encouraging and facilitating these exchanges, using insights more broadly within the organization, and educating product owners is really our role [trans.].' Technology seems to become a vehicle for making public organisations work more comprehensively; however, not in an interdisciplinary way. ICT is the leading force in linking departments, bringing in its own specific perspective. The assignment of this role is not a deliberate choice but a gradual and natural process. Moreover, interviewees discussed the differences in power positions or influence in the design process between actors. For example, compliance officers are involved too late, or their input is subordinated to other considerations.

Accordingly, the question or problem addressed in the design process of public algorithmic systems is mostly formulated or shaped by technical designers. However, these designers approach the client's question from a specific – i.e., technical – perspective. This can result in attuning the needs and wishes of the client to what the technical developer can provide. Other departments consider ICT as a facilitative element to which the development of public algorithmic systems is 'thrown over the fence'. In this way, the development of algorithmic applications attains a sense of neutrality that also disguises the central role of ICT in the design process. Naturally, the technical artefact attains most attention, neglecting institutional and agential components of public algorithmic systems. Compliance departments cannot correct this due to their weak position.

Besides the patterns of deliberating and of a coordinating role of ICT in client-supplier relationship, we identified a pattern of interaction related to time. Interviewees in the explanatory study stated that they often have to wait on crucial input from or decisions by other designers. They are confronted with difficulties of bringing experts from other parts of the organisations into the design process at the right time. But slack in the design process is also caused by frequent changes in staff within design projects. Since the design process is a lengthy process, it will probably not be finished by the same group of designers that started the project. One interviewee provided an example: '[obtaining authorisation to use particular data] takes a lot of time. That was a discussion between the project leader and the department head. And then the department head leaves for another function. It takes another month before the next department head comes in, who needs to settle in and be informed about the project [trans.].' Finally, interviewees discussed how different disciplines within the organisation work in asynchronous rhythms. Teams of developers, the technical designers, often work following scrum or agile methods. Every few weeks, they complete a sprint in which a new iteration of an application is delivered. However, legal advisors and domain specialists indicated that their activities follow different timelines, which makes it harder for them to be and stay involved in the design process.

4.1.3 Institutions and attitudes structuring the action situation

This section discusses the institutions (I) and attitudes (A) that structure the first action situation. Table 4.1 provides an overview of the institutions and attitudes. The rest of this subsection discusses the institutions and attitudes in more detail. We highlight the individual institutions and attitudes in bold. Important to note here is that the described attitudes are generalised. Within organisations, attitudes can differ between individuals. In this study we have identified attitudes that are shared by the majority of the interviewees, or attitudes that contradict each other.

Table 4.1 Institutions and attitudes structuring the collaboration between disciplinary designers

Institutions	I1	Public organisation is structured in silos
	I2	Obligation to involve compliance officers
	I3	Ethics as discipline or expertise included as focus point in design process
	I4	Project-based approach of the design process
	I5	Documentation of design choices in, for example, impact assessment
	I6	Informal deliberation between public servants about design choices
	I7	Formalised deliberation between public servants of different disciplines
	I8	Procurement of IT products from external parties
	I9	Public organisation is structured following a client-supplier relationship
	I10	Use of corporate language in English
	I11	Compliance has less definite position in phases of design process compared to the business and ICT
	I12	Developers start with intake as soon as question for algorithmic systems comes in (part of phasing of design process)
	I13	Multiple established teams that develop algorithmic applications
	I14	Agile and/or scrum practices at developers' side
	I15	A strict phasing of the design process in policy documents
Attitudes	A1	Expectation that critical reflection emerges from interdisciplinary approach
	A2	Impression that most public servants lack knowledge, awareness, and expertise about algorithmic systems
	A3	Differences in routines and expertise between departments
	A4	Impression that the business lacks knowledge and capabilities
	A5	Algorithmic systems are mere or simple automation
	A6	Government should not disrupt markets
	A7	Public organisations are not able to design algorithmic systems efficiently
	A8	ICT is a facilitative element in the public organisation
	A9	Public organisations should innovate
	A10	Compliance is burdensome and a hindrance
	A11	Acknowledgement that a design process is not linear, but needs flexibility
	A12	Governance of design processes is workable when simple and straightforward

A defining characteristic of this action situation is the distinction between three aggregated positions: business, ICT, and compliance. These positions mirror the **organisational silos** (and the independence of data science teams) in the studied public organisations **(I1)**. Each silo represents a specific expertise within a content-oriented domain or a service-oriented discipline. The design of algorithmic systems requires expertise within the different silos (see also Van Der Voort et al., 2019). **Involvement of specific compliance officers**, such as the privacy officer, can even be obligatory by law **(I2)**. The public organisations in this study also expected that organising **interdisciplinary practices will result in critical reflections** on ethical challenges **(A1)**. At the same time, interviewees stressed the importance of **awareness among involved designers of the socio-technical character of public algorithmic systems** to arrive at a more comprehensive or interdisciplinary design process **(A2)**. One interviewee stressed the character trades of actors that often result in apparent successful projects: 'The specific people involved in this project were already thinking about these themes [e.g., ethics or the impact on citizens]. So, not just developing an application, but to think, okay, why do I program these choices in this model? [trans.]' This might explain why several public organisations have **established a new position for ethicists** in the design process **(I3)**.

Second, the need for interdisciplinary collaboration between silos has resulted in deliberating being the main design activity. Deliberation is needed since designers, working in the **project-based design processes,** are dispersed over the whole public organisations **(I4)**. These designers need to share information, explain their own expertise to their project colleagues, arrive at design choices, and **document** all this to ensure continuity **(I5)**. But each organisational **silo or position has a different view on practices and routines in design processes (A3)**, which hampers communication and collaboration between disciplines. Developers shared the experience that questions by 'clients' are often not specific or clear enough to start developing an application. As the data science team lead put it: 'we have to educate our client [the business] [trans.]', implying that the **business is not yet capable of posing the right questions** for developing such systems **(A4)**. In reaction, **informal communication between designers** has become a core practice in designing public algorithmic systems **(I6)**. In other public organisations, the **deliberation** between disciplines has already been **formalised (I7)**. One interview described the situation: 'we have a whole procedure that we have to go through before we can start an experiment [with an algorithmic system]. That is called the special table [a session in which different experts deliberate about a system] at which everything needs to be approved; legal and, you name it, all these aspects.'

While the prevalence of deliberation in design practices may signify equality between designers from different disciplines, officially the business and ICT are in a client-supplier relationship. According to our interviewees, most of their colleagues **consider algorithms as mere or simple automation (A5)**. In this logic, an algorithmic system is a product that can be procured. Mulligan and Bamberger (2019) have called this the 'procurement mind-set' of public organisations, also when a system is developed internally. Although the studied public organisations do act according to this procurement mindset, various interviewees themselves stressed that this approach underestimates and misunderstands how automation is ingrained in and structures the execution of public policy.

Although some organisations develop algorithmic systems internally, most organisations, especially those that are smaller and with limited resources, **procure systems from external parties (I8)**. Some of the participating public organisations embrace the principle that **government should not disrupt markets (A6)** and should **outsource any public tasks that can be executed more efficiently by commercial parties (A7)**. These public organisations do not regard the acts of design as a core role for public administration. One interviewee stated: 'It is also the starting point of the government. People say we do not want to disrupt markets. We are the coordinator; we outsource what we can outsource [to external parties] … Well, that is also the direction that the government has taken in recent years [trans.].' The development of algorithmic systems often stems from the perceived need of public organisations to use limited resources efficiently. Public organisations start using algorithmic systems because they lack the capacity to execute their core tasks. Similar perceptions are observed considering capacities and capabilities to develop AI systems. Participants argue that smaller public organisations do not have the resources, capabilities, or capacity for development. That said, bigger organisations similarly argue that they do not have the knowledge or expertise in-house.

Still, some algorithmic systems are developed internally. But this also happens within a **client-supplier relationship** between the business and ICT, which has become the formal approach in most public organisations **(I9)**. The domain experts are the 'clients' that commission a product supplied by a developing department (e.g., a data science team). Within this relationship, **ICT is considered to be a facilitative element** within the organisation **(A8)**. This is also manifested in the usage of a **specific vocabulary of English terms** used in a Dutch context **(I10)**. The domain-oriented part of the public organisations is considered to be the core of the organisation and, therefore, referred to as the 'business', which is serviced by internal IT departments of external suppliers. The development of algorithmic systems is mostly outsourced to these suppliers. Interestingly, the English term 'business', when used in the Dutch language, mostly refers to the private sector.

The client-supplier relationship brings a clear division of roles on paper, but interviewees stated that roles and responsibilities are not always clear in practice. The coordinating role of ICT has emerged from this situation. Section 4.1.2 already pointed towards the institutions and attitudes that form the basis for this shift. First, **compliance officers have a less definite position** in the governance of design processes **(I11)**. These officers have few means and a relatively weak position to intervene in the design process. Discussing their position as compliance department, one interviewee stated: 'The tricky thing is, we cannot control the money. The money is at 'the business', we cannot stop them [trans.].' Second, the **intake performed by ICT** puts the technical designers in a crucial role **(I12)**. The developers clarify the question of the 'client' and derive their needs and requirements in this intake. In our study, interviewed developers justify their role in the intake with their experience in previous projects that non-technical designers often are not able to formulate specific or clear questions **(A4)**.

ICT can also take up this coordinating role because of the perspectives on algorithmic applications, and technology in general, within public organisations. The attitude towards algorithmic systems as being mere automation **(A5)** makes the algorithmic application in algorithmic systems the focal point in the design process. Executive designers disregard the fact that an algorithm can also influence the interpretation of those policies and regulations. Furthermore, the lack of awareness was observed in the difficulty of demarcating system boundaries. Interviewees stated that the business is not always aware of the practices that can be affected by an algorithmic application, and that they often do not recognize the products they use as algorithmic systems. The focus on technology can also be observed in the attitude of most interviewees and their colleagues that **public organisations should innovate (A9)**. Interviewees discussed that, within their organisations, algorithmic systems are considered innovations in service provision (apart from efficiency improvements). In this, innovation is often considered equal to creating and implementing new technologies. Public organisations often have **multiple teams** that they assign the (sometimes competing) responsibility for exploring what technologies might be fruitful for the organisation's purposes **(I13)**. They have a classic IT department and teams that are dedicated to data analysis or business information. To stimulate innovation, the studied public organisations created independent data science teams. Such teams are often positioned at a distance from core practices to give them the possibility to experiment with emerging technologies (see also Lorenz, 2023).

Considering the interactions related to the pattern of waiting, we see several causes for slack in the design process of public algorithmic systems. This slack can be related to the ad hoc, project-based approach towards designing algorithmic systems within public organisations **(I4)**. Designers involved in a project often do this besides their normal work activities and have to divide their attention between these two responsibilities. Moreover, designers in a compliance position tend to be involved to late, making it hard to change or stop a project with high sunk costs. As a result, interviewees in the exploratory study stressed how **compliance is considered burdensome (A10)**. One of the interviewees stated: '… as compliance, you are considered as hindering. Whereas if we would be involved early on, there are lots of possibilities to ensure compliance with frameworks and guidelines. At present, we are considered burdensome at the end of a project and, well, there is a continuous battle between innovating and formulating compliance frameworks and guidelines within the organization. Yes, that is an area of tension [trans.].' This poses a paradox. The consortium was formed because of the need for frameworks and guidelines for public algorithmic systems. As the data science team lead put it: 'if you tell what I can and what I can't do, and what my boundaries are; than I can be very creative within those boundaries and will not cross them [trans.].' At the same time, existing frameworks and guidelines are considered as hindrances.

The project-based approach also brings designers together who have different work rhythms. Teams of technical designers have often adopted agile or scrum practices common in software development **(I14)**. They develop algorithmic applications in sprints (i.e., delimited periods of time in which new iterations of the technical artefact are produced). In between sprints, the 'client' is informed about the progress of the project, and priorities and requirements are aligned. The sprints are typically strictly embedded within the boundaries of the different phases in the design process. In policy documents and during discussions in the consortium, the design process is often conceptualised as a linear and discrete process of subsequent steps – e.g., visualised as blocks connected by arrows presenting a clearly demarcated and chronological process **(I15)**. Policymakers did recognise the shortcomings in structuring and standardising an inherently deliberative and creative process. They **acknowledged that the design process is iterative and idiosyncratic in practice (A11)** but returned to their linear conception to **make these policies** – as stated in the consortium meetings – **'workable' (A12)**. In other words, the design process of public algorithmic systems is iterative in practice but is not understood as such in policy documents. Public organizations have created design processes that do not align with how public algorithmic systems are developed in reality. This is mostly because the starting point of the design process – i.e., linearity – conflicts with the nature of development – i.e., iterative.

4.2 Political steering

This section discusses the second action situation identified in our exploratory and explanatory studies. In this action situation, representative designers (being politicians) determine the normative boundaries that define the design space of executive designers, see Figure 4.2. Subsequently, executive designers conduct the work that follows from the instruction formulated by the representative designers. The position of executive designers consists of the three types of designers discussed in Section 4.1. The action situation presented in this section focuses on the interaction between politicians from public servants. Again, the section discusses the two positions in the action situation. Thereafter, Section 4.2.2 discusses the patterns of interaction between these two positions. The section ends with the identification of institutions and attitudes that structure the action situation.

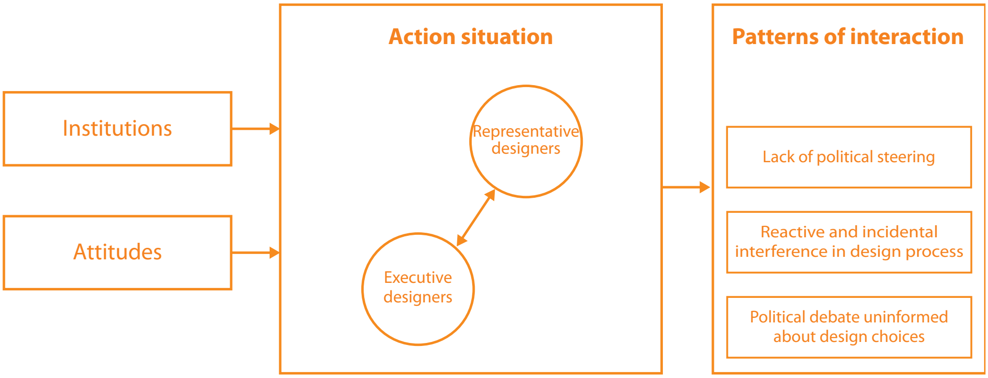


Figure 4.2 Action situation in design processes of public algorithmic systems that comprises the interaction between representative designers and executive designers

4.2.1 Action situation

Representative designers
Representative designers are actors assigned to this position through elections or other forms of democratic mandates. So, for example, a legislative branch consisting of a representative body that translates the interests of the public into policy- and decision-making is a representative designer. But also the political side of the executive branch, such as an Alderman or a minister, can have the mandate to make political trade-offs. Through their actions, they provide democratic legitimacy to design choices. The actions of representative designers are related to the legislative, controlling, and decision-making tasks assigned to this position. Supposedly, representative designers have the best information position related to the interests they represent and the political conflicts and/or consensus that arise from those interests. In theory, the representative designers have primacy over the executive designers. In practice, they heavily depend on the expert knowledge of executive designers.

Executive designers
Representative designers give a mandate or instruction to executive designers, who formulate the socio-technical specification or create the actual algorithmic systems. The position of executive designers is filled by actors employed as public servants. Representative designers heavily depend on the in-depth expertise of and information provided by executive designers in their advices and reports. In the studied organisations, the business is generally responsible for and in the lead in the interaction between executive designers and representative designers.

4.2.2 Patterns of interaction

We observed three patterns of interaction between the representative and executive designers. First, representative designers refrain from determining a normative or political instruction that adequately steers the problem-formulating and problem-solving activities of executive designers. Politicians either do not engage in steering design choices of executive designers, or they interfere in the design process in a reactive and incidental way. The unclarity about normative boundaries of public algorithmic systems impedes executive designers in making design choices and, ultimately, results in prioritising 'political sensitivity' as rationale behind design choices.

Second, if representative designers do provide a political instruction, this happens in reaction to incidents that are prominent in public debate. Interviewees discussed how the introduction of laws and regulations – e.g., the GDPR boosting an emphasis on privacy – or dramatic incidents – e.g., the childcare allowances scandal or the SyRI case – became leading considerations in making design choices. The reactive and incidental nature of political debate results in two dynamics: 1) increased attention for specific 'high-risk' or 'critical' algorithmic systems, and 2) political pressure on executive designers to prevent negative publicity related to the implementation and/or operation of public algorithmic systems. In response, representative designers focus on specific 'critical' algorithmic systems. In addition, for executive designers, 'political sensitive' becomes a prominent argument in making design choices. In sum, the political discussion is not focused on the design choices that have to be made, but the problems that publicity might bring.

Third, interviewees either argued that the awareness of political trade-offs in a system specification is often low among executive designers or questioned their own agency in working on these trade-offs. The lack of awareness, and of agency results in inadequate communication of normative trade-offs or hard choices to politicians. Consequently, representative designers are insufficiently informed about the alternatives in design choices.

4.2.3 Institutions and attitudes

Like Section 4.1.3, this section discusses the institutions (I) and attitudes (A) that structured the action situation discussed before. Table 4.2 provides an overview of all institutions and attitudes specific to political steering.

Table 4.2 Institutions and attitudes structuring political steering of design choices

Institutions	I16	Design process is organised as a traditional policymaking process
	I17	Laws applicable to algorithmic systems such as the GDPR
Attitudes	A13	Politicians show low interest in design process
	A14	Politicians lack knowledge to engage in design process
	A15	Avoiding negative publicity regarding algorithmic systems is important
	A16	Political sensitivity is an important rationale behind design choices
	A17	Low awareness among executive designers of normative trade-offs
	A18	Low feeling of agency to make normative trade-offs among executive designers

Several, contradicting, institutions and attitudes can be related to the lack of political instruction by politicians. In principle, the design process is **structured following a traditional policy-making process logic (I16)**. Design choices and outputs of designers on levels lower in the organisational hierarchy are assessed by designers on higher levels to guarantee and embed values in the specification of algorithmic systems. This bureaucratic approach should ensure that domain-specific laws and regulations, high-level (digitalisation) policies, and political consensus captured in coalition agreements trickle down to the design level and the final algorithmic system. Despite having this policymaking structure in place, the translation of the political debate to design levels falls short. Interviewees stated two attitudes that they observe among politicians. First, politicians show **low interest in design processes** of algorithmic systems **(A13)**. In the explanatory study, executive designers mentioned how they often need to lobby for or attract attention towards algorithmic systems among politicians. Second, others observe a **lack of knowledge about algorithmic systems among politicians (A14)**.

The low interest and lack of knowledge among politicians contradicts the reactive and incidental interference of politicians in the design process, the second interaction identified in this action situation. Interviewees stated that politicians develop an interest in the design process in response to enforced laws or major incidents caused by algorithmic systems. Interviewees mentioned **the GDPR as an example of a law** that has had a considerable impact on the deliberation of design choices **(I17)**. Since the enactment of this law, the political debate on algorithmic systems has concentrated on privacy concerns. In terms of incidents, interviewees mentioned the childcare allowances scandal as a turning point in the interest among politicians in algorithmic systems. The interactions observed in one of the four organisations in the explanatory study show what dynamic emerges from a similar incident. The auditor of the public organisation deemed one of their algorithmic systems to have discriminatory outcomes. The auditor's report resulted in negative publicity. Our interviewees observed an attitude among politicians to **avoid such publicity (A15)**. One interviewee stated: 'this [project] is so sensitive and so closely monitored, also with different [information retrieval requests], but also, yes, the childcare allowances scandal … made this political'. In response, executive designers stated that **political sensitivity** of the algorithmic systems becomes **an important rationale** behind design choices **(A16)**.

A similar lack of awareness about political implications of algorithmic systems among executive designers explains the impeded information provision from executive designers to representative designers. First of all, the attitude towards algorithmic systems as mere automation disregards the politics in designing such systems **(A5)**. In developing and implementing the policy instruments in their own organisations, consortium participants were confronted with **low awareness among their colleagues about the challenges that algorithmic systems pose (A17)**. One interviewee stated that their colleagues working with algorithmic systems are 'not aware of ethical choices [trans.].' Executive designers seem to be heedless of the risks that public algorithmic systems may have for individual citizens. One interviewee described the attitude in their organisation as: 'it is just a system that does what it must do and there are no risks related to it [trans.].' Similarly, several interviewees in the explanatory study stated that not all algorithmic systems do have a political element to it. This interpretation of simple automation seems to make a strict distinction between algorithmic systems and the policies that underlie these systems. An algorithmic application is seen as a tool that supports public policy and other activities of public organisations. The application rationally and neutrally executes the policy it is associated with and does not influence the policy itself. Thereby, the political and normative aspects of these systems are disregarded, misinterpreted, or underestimated. Consequently, these political and normative aspects are also not shared with the representative designers.

And in the case that executive designers are aware of the normativity of design choices, their **feeling of agency to make or discuss these trade-offs is often low (A18)**. For example, the consortium often discussed the need to consider values in the design process. However, they argued that it was up to the individual public organisations to specify the relevant values. In the explanatory study, part of the executive designers discussed being uncomfortable in making normative trade-offs. When discussing their role in performing an ethics impact assessment, a legal advisor stated: 'the question [concerning ethics] is assigned to me, but I cannot answer it.' The advisor felt they did not have the authority to give definite answers in the assessment. Policy instruments such as ethics impact assessments do play an ambiguous role in design. Ethics has become a main consideration in designing algorithmic systems with the implementation of ethics assessments and ethics boards **(I3)**. These instruments can be seen as a way to depoliticise design choices (cf. Andrejevic, 2020). In this case, performing an ethical impact assessment transfers a political mandate to executive designers. On the other hand, participants in the follow-up workshop stated that the assessment used in their organisation provides individual executive designers with a means to voice their objections and doubts. As such, assessments are a way to address impeded information flows between political and executive designers.

4.3 Presuppositions structuring design practices

The identified action situations indicate four impediments in design practices. First, the patterns of interactions regarding political steering exemplify a *political disconnect* between representative and executive designers. The vertical information flows between the two groups of designers are structurally and culturally disconnected. This impedes deliberation on normative aspects of public algorithmic systems. Second, the coordinating role of technical designers that emerges within the client-supplier relationship between the business and ICT reduces the integration of knowledge and experience of non-technical designers. Accordingly, the siloes in public organisations impede interdisciplinary collaboration between executive designers as *non-technical expertise is disregarded*. Third, a *procurement mindset* is manifested in the client-supplier relationship. This mindset reinforces the perception of algorithmic applications as products that can be bought. Again, this impedes a systemic approach towards the design process of public algorithmic systems. Finally, public organisations adopt *a project-based and phased approach* to the design process. This approach is a way to coordinate the different disciplines involved in designers. However, the rigidity of depictions of design processes impedes an iterative design process.

In this section, we abductively derive the presuppositions behind the institutions and attitudes identified in the action situation (see Section 3.3.1 for the research method). We start from the four impediments to elicit the presuppositions within the studied public organisations. Presuppositions are ideas, ideologies or institutional logics prevalent within organisations that form the basis of institutions and attitudes that shape the design practices of public algorithmic systems. The identified institutions and attitudes identified in Sections 4.1.3 and 4.2.3 show close parallels with fundamental principles of technocracy and New Public Management (NPM) as discussed in the literature. Therefore, we distinguish two presuppositions that structure design practices: technocratic and businesslike.

4.3.1 Technocratic

The attitudes and institutions shaping the design practices are similar to the notion of technocracy. Centeno (1993, p. 314) defines *technocracy* as: 'the administrative and political domination of a society by a state elite and allied institutions that seek to impose a single, exclusive policy paradigm based on application of instrumentally rational techniques.' Although literature on technocracy describes an ideal type that focuses mostly on the organisation of a polity or state, it is still useful to analyse empirical cases such as public algorithmic system design processes. The characteristics of technocracy are mostly coming back in the prevailing political disconnect and the disregard of non-technical expertise in design processes. Moreover, the project-based and phased approach towards design can also be related to the technocratic presupposition.

In a technocracy, politics is considered to be stable (Centeno, 1993). This is reflected in the shared meaning among public servants that some algorithmic systems can be considered simple automation and, therefore, normative and political deliberation on design choices is not needed **(A5).** In addition, the shift from politics to ethics – where public servants are responsible for assessing ethical compliance of algorithmic systems – also shows that public organisations undervalue the political trade-offs that design choices pose **(I3; A17)**. Accordingly, the design practices can be compared to another characteristic of technocracy: a method or tool is considered suitable for addressing the main issues facing society. Algorithmic applications are seen as appropriate tools to address societal or organisational issues and, therefore, legitimate technocratic governance (cf. Janssen & Kuk, 2016) **(A5; A9)**.

The disregard of non-technical expertise is in line with the considerable institutional autonomy or freedom that is assigned to experts in a technocracy (Centeno, 1993). In the design process, technical experts and developers have a particular form of autonomy in the client-supplier relationship ingrained in design practices **(I9; I11; I12)**. In addition, technical designers are provided some autonomy by not being embedded in the core of public organisations but in so-called data science labs **(I13; A8)**. As such, data-analysts and engineers obtain more prominent roles and positions in organisations and in designing policies. A coordinating role of technical actors and engineers emerges in the design process (see also Zouridis et al., 2020). A focus on technical expertise is also observed in prevailing views of public organisations on algorithmic systems. These systems are considered efficient, accurate and neutral **(A5)**. Consequently, they are increasingly used in all sorts of policy areas (Endacott & Leonardi, 2024; Reutter, 2022). Here we see technocracy being reflected in both the ideology obtained by an organisation and the kind of personnel it hires and assigns to the design process (cf. Centeno, 1993).

Another characteristic of technocracy is that it develops when an organisation is confronted with a complex task (Centeno, 1993). Regarding public algorithmic systems, public organisations are confronted with a technology with the potential to improve their work practices. Designing such systems is a complex task, see Section 2.3.3. The project-based and phased approach can be considered a technocratic response to the complexity of the design process. Public organisations depict the design process as a straightforward process with clear predefined outcomes that only involves limited politics **(I15; A13; A17)**. Finally, Centeno (1993) describes a 'bias towards technocratic methods and interpretations' as a core element of technocracy. Accordingly, the project-based and phased approach within design practices suggests an attempt to standardise an inherently complex design process.

4.3.2 Businesslike

The second presupposition indicates that public organisations integrate strategies and practices from the private sector in their design processes. This businesslike presupposition stems from the New Public Management (NPM) paradigm that has shaped public administration for decades (Dunleavy et al., 2005). Two defining themes of NPM identified by Dunleavy et al. (2005) – disaggregation and competition – can be observed in the procurement mindset and the project-based phasing of design processes.

Competition, as defining theme of NPM, refers to splitting up purchaser and provider in order to instigate competition at the side of the provider (Dunleavy et al., 2005). This theme can be recognised in the procurement-mindset observed in design practices **(I4).** The idea behind competition is to arrive at a more efficient allocation of resources. Similarly, public organisations bring in a competition element as they consider themselves to be inefficient designers of algorithmic systems, to stimulate innovation, or because they do not consider designing algorithmic systems to be their task **(A6; A7; A9)**. In NPM, competition is encouraged by 'introducing purchaser/provider separation' (Dunleavy et al., 2005, p. 470). This is reflected in the prevalence of client-supplier relationships between designers of public algorithmic systems **(I9)**.

Disaggregation in organisational structure refers to transforming organisational hierarchies in public organisations into flatter hierarchies that can be found in the private sector (Dunleavy et al., 2005). Disaggregation is most prominent in the project-based approach towards designing public algorithmic systems. The project serves as a vehicle to bring executive designers from different siloes in the organisation together. The siloes are an effect of disaggregation **(I1)**. Another example of disaggregation is the autonomy of data science teams mentioned before. In reaction, interviewees stressed the need for co-creation practices in which different siloes work together **(A2)**. However, the attempts to arrive at these practices are currently impeded by disaggregation within public organisations. For example, this is observed in the approach towards system user involvement. The organisations in this study mostly focus on training users to work with a finished system and do not engage them in the design of the system. Public organisations seem to be unable to transcend one-way approaches of engaging users such as training users in using algorithmic systems. A paradox arises here: public organisations have high expectations of making the design process more interdisciplinary, but at the same time it can be argued that, because of the many disciplines involved, the design process is already interdisciplinary. It seems that problems related to collaboration and communication between disciplines need to be addressed before interdisciplinary approaches will work appropriately.

In addition to NPM-related characteristics, the adoption of work practices from the private sector observed in design practices also reflects a businesslike presupposition. For example, public organisations embrace agile and scrum practices **(I14)**. Similarly, adopting jargon or vocabulary from the private domain, e.g., 'business' for the core of the public organisations and 'product owner' for those responsible for the algorithmic system, is a manifestation of a public organisation that strives to become more businesslike **(I10)**.

4.3.3 Implications of technocratic and businesslike presuppositions

This section reflects on the implications of the two identified presuppositions. We discuss how the presuppositions relate to the output of the design process, the extent to which these presuppositions are new, and what the identified presuppositions mean for designing institutional interventions in design practices.

The presuppositions are embodied in and will manifest in the output of the design process – i.e., the algorithmic system. Simon (1969/1996) already described how the organisation of design processes influences the output of that design process. Accordingly, the technocratic and businesslike presuppositions can also be observed in public algorithmic systems. The technocratic presupposition corresponds with the tendency of public algorithmic systems to result in technocratic governance (see Janssen & Kuk, 2016). Most of all, the disregard of politics in the design process is also found in public algorithmic systems themselves. The presupposition shows that algorithmic systems are considered as another technical artefact that supports work processes in public organisations, which disregards the fact that automation, especially with more advanced forms of AI systems, is integral to what public organisations deliver and may reshape core processes in public administration (see also Mulligan & Bamberger, 2019). Moreover, austerity has been a core motive behind making processes in public administration more efficient by implementing algorithmic systems (Dencik et al., 2019), which can be related to the businesslike presupposition.

The presuppositions are not necessarily a response to the increase in possibilities of algorithmic systems for public administration. The adopted presuppositions have been associated to designing in public administration for a longer time, for example, in the field of policy design. The two elicited presuppositions are similar to two streams of thought on policy design. The discipline started with a structured, technocratic, and expert-driven perspective on design in public administration (Van Buuren et al., 2020). Currently, policy design is mostly focused on design thinking, co-creation, and other NPM-related concepts (Van Buuren et al., 2020). We observed both approaches to design: a clear inclination or tendency towards expert-led design, and procedural design processes infused by private sector practices. However, policy design scholars also have acknowledged that these two streams of thought disregard political characteristics of public administration (Van Buuren et al., 2020). Our empirical studies also observed a neglect of political factors in the design practices of public algorithmic systems in the Dutch context.

Despite the renewed enactment of technocratic and businesslike presuppositions, these presuppositions have been dismissed as a suitable basis for design processes of socio-technical systems in a public context. Bostrom and Heinen (1977) already discussed how practices comparable to those found in our analysis are incompatible with designing socio-technical systems. Similarly, Dunleavy (2005) already observed that the client-supplier relationship was being phased out in public organisations. Our study on the presuppositions behind design practices suggests that public organisations' attitudes and approaches towards digital technologies have not changed, while the role of digital technology in public administration has transformed considerably.

The introduction of more sophisticated algorithmic systems in core processes of public administration raises the need to consider different presuppositions to base the design process on. It is well known that ensuring algorithmic tools and software-based forms of automation are built safe, responsible, and adhere to other relevant values is a steep task that requires going well beyond technical specifications (Lindgren, 2023). As discussed before, the fact that these presuppositions and the lack of political mechanisms are still found in public organisations is an indication that the many normative trade-offs presented by algorithmic systems (Selbst et al., 2019) are disregarded. As Dobbe et al. (2021) show, the lack of prior standards or forms of consensus for the socio-technical context of a given system renders the specification of that system problematic or incomplete. In fact, values as well as hazards in algorithmic systems are emergent properties which can only be understood and designed for across the technical artefacts, its social context and the institutional environment (Nouws et al., 2022). Consequently, current design practices do not enable public organisations to prevent, mitigate, or correct hazardous situations in public algorithmic systems.

This study has underlined the importance of a broader perspective on the design process of algorithmic system and the role that presuppositions play in shaping design practices. This does not mean that a change in presuppositions will certainly result in design processes that prevent, correct or mitigate Kafkaesque situations in public algorithmic systems. For example, Siffels et al. (2022) and Fest et al. (2022) describe problems related to capacities of and capabilities in public organisation, such as digital literacy, that also impede adequate design practices to emerge. Moreover, this research's sample is quite specific, as all involved public organisations are located in the Netherlands. Different design practices – and underlying presuppositions – may be present in other contexts.

4.4 Conclusion

This chapter set out to identify design practices of public algorithmic systems and to study presuppositions that structure the institutions and designers' attitudes in these practices. The study consisted of explorative interviews and observations at a consortium of 13 Dutch public organisations, followed by explanatory interviews with designers at four Dutch public organisations. Using the IAD framework to elicit design practices, we found patterns of interactions in two action situations: collaboration between disciplinary designers, and political steering of design.

The study identified four impediments to the interactions between designers in the two action situations. First, there is a disconnection between representative designers and executive designers. Second, the input or expertise of non-technical executive designers is disregarded by other designers. Third, following a procurement-based dynamic, the executive designers are divided in 'clients' and 'suppliers'. Finally, the different interacting types of designers work in asynchronous rhythms.

From these four impediments, we were able to reconstruct the presuppositions that underlie the institutions and attitudes constitutive to the identified interactions. The institutions and attitudes shaping interactions within design practices are characterised by a focus on client-supplier relationships, a procurement mindset, too little knowledge and experience among public servants, and organising the design process as traditional policymaking combined with agile and scrum practices borrowed from private sector parties. In general, these characteristics of the institutions and attitudes can be attributed to two presuppositions prevalent in the studied public organisations: technocracy and businesslike. The technocratic presupposition can be observed in the procedural approach towards design processes and the emphasis on expertise. The businesslike presupposition can be recognised in the practices copied from private sector parties.

The results of this study indicate that current design practices are both inconsistent with socio-technical designing characteristics, as well as increase the possibilities for Kafkaesque situations to emerge. Inconsistency with socio-technical designing characteristics can be observed in the leading role of technical, executive designers that comes with an emphasis on technical artefacts in the design process. As such, institutional and agential components of public algorithmic systems are neglected. Apart from neglecting this systemic characteristic of socio-technical designing, current design practices also disregard the other characteristics. First, lacking an overview of the whole system (to be designed), designers do not bring the emergent properties of algorithmic systems into view. Moreover, the contingency of algorithmic systems is disregarded by ignoring non-technical expertise. Although a diversity of designers is involved in the design practices, public organisations tend to distinguish two groups of designers – clients and suppliers – that are in a simplified bilateral relationship. This division does not do justice to the complexity of a multi-stakeholder network. Finally, in line with the technocratic approach that inherently ignores the politics behind design choices, we observed designers falling short in translating public or political debates to design choices underlying public algorithmic systems.

In addition, the protection of citizens is not a core concern within the technocratic and businesslike presuppositions. Consequently, designers overlook the effects on those affected by the algorithmic systems they design. Designing socio-technical systems as a public organisation also requires designers to consider and respect the critical relationship between citizens and government. Again, this is a reason to facilitate public debate about design choices and guarantee the protection of citizen rights and the prevention of citizen harms.

The results of this empirical study emphasise the importance of presuppositions in forming design practices. The identified presuppositions based on technocracy and New Public Management have shaped practices within executive branches of government for a longer time. Unsurprisingly, they are coming back in the algorithmic practices of public organisations. Algorithmic systems are part of a longer tradition of automation in public administration and align with the technocratic and businesslike presuppositions. The design practices of algorithmic systems have a significant influence on the algorithmic practices that will eventually emerge. This empirical study indicates that public organisations can change their design practices by shifting the underlying presuppositions and, accordingly, constitute alternative algorithmic practices. The next chapter will study what public organisations already do to transform the design practices of public algorithmic systems through institutional interventions.

Chapter 5

Policy instruments as institutional interventions in design processes

> Once one adopts the view that one cannot create the perfect set of rules and that all efforts at reform must be viewed as experiments, one recognises that policy analysis can never find "the" answer.
>
> – *Elinor Ostrom, Understanding Institutional Diversity, p. 254*

5

Chapter 4 focused on identifying design practices for public algorithmic systems currently found in public organisations. At the same time, these organisations are also exploring means to transform their algorithmic practices in general. We are interested in the institutional interventions developed and implemented by public organisations to reshape algorithmic practices and in the goals these organisations aim to achieve with these interventions. Insight into the initiatives of public organisations can provide an understanding of the requirements for design practices as perceived by public organisations. This chapter examines initiatives of public organisations by answering the following questions:

What institutional interventions do public organisations develop to transform their algorithmic practices? To what extent do these interventions achieve the pursued goal?

In addition to studying design practices within the consortium (see explorative empirical study), we examined their efforts to develop policy instruments for public control over algorithmic systems. The consortium of 13 Dutch public organisations developed an algorithm register, a governance framework, procurement conditions, and instructions for objection procedures. We consider these policy instruments to be a specific type of institutional interventions in algorithmic practices, as they aim to alter the behaviour of human agents (i.e., public servants) within these practices. Policy instruments are a common approach to respond to challenges that public algorithmic systems pose to institutional practices in Dutch public organisations.

We used the interviews and observations at the consortium, in combination with the document analysis, to assess the four policy instruments of the consortium (see Section 3.3.1). We mostly focused on the extent to which the policy instruments contribute to the consortium's goal of establishing public control. The assessment of the policy instruments also informs our design theory. The consortium's instruments implicitly reflect the requirements that public organisations have identified for their algorithmic practices. Considering that design practices of algorithmic systems are a specific form of algorithmic practices, the assessment provides insight into the requirements that public organisations have – explicitly or implicitly – formulated for design practices. Accordingly, this chapter also answers the following question:

What requirements for design practices for public algorithmic systems follow from the institutional interventions developed by public organisations?

This chapter is structured as follows. Section 5.1 discusses the four instruments developed by the consortium and relates them to institutional interventions discussed in scientific literature. Section 5.2 presents the results of the analysis of the instruments by discussing the affordances and challenges of each individual policy instrument. Section 5.3 reflects on these affordances and challenges by comparing them to the consortium's goal behind the instruments. Section 5.4 concludes this chapter by presenting requirements for institutional interventions in the design process of public algorithmic systems that follow from the consortium efforts in developing and implementing policy instruments to reshape algorithmic practices.

5.1 Policy instruments for public control over algorithmic systems

The consortium studied in the previous chapter developed policy instruments with which the public organisations aimed to establish public control over public algorithmic systems. At the start, the consortium defined public control as: 'For us, public control means that the society – citizens, businesses, government, and all other actors playing a role in public control – has insight in and has a say about algorithms, even if they are complex. … If we fail to facilitate this for citizens and businesses, trust in the government will disappear [trans.].' As such, public control is achieved by providing relevant actors with meaningful information about public algorithmic systems and enabling them to engage in the design and deployment of those systems. For the consortium, public control was explicitly intended to refer to control over algorithmic systems, rather than using algorithmic systems to control processes or human agents, such as for surveillance purposes.

In its program plan, the consortium stated that the instruments should 'determine conditions, provide overview, and support interventions if needed; in order to provide citizens, administrators, and civil servants with more information, understanding, and influence [trans.].' The consortium developed an algorithm governance framework, an algorithm register, procurement conditions, and instructions for information provision in objection procedures. Table 5.1 provides an overview of these four instruments and how they evolved over time. The deliverables of the consortium were planned to be 'minimal viable products' which could serve as a starting point for follow-up policymaking processes and regulating processes to establish public control on a larger scale.

Table 5.1 The four instruments defined and elaborated by the consortium in their program plan and final deliverables of the four instruments

Instru-ment	Rationale at the start of the program	Description of the instrument	Deliverable at end of the program	After the program
Algorithm governance framework	Fulfilling the need for control over the full life cycle of public algorithmic systems. Directive and guiding framework applicable to Dutch public organisations. Supporting organisations in organising their working methods, inspection and audits related to algorithms.	Synthesis of three different frameworks (technical, legal, and ethical) in one governance model.	Guide for organising governance, including: definitions; a process description which highlights key moments at which organisations should intervene in the system; and a description of tasks and responsibilities to be assigned.	Used in different public organisations as basis for establishing algorithmic governance.
Algorithm register	Increase transparency and inform citizens, journalists, members of parliament/council, inspectorates, suppliers and others about algorithms in use. Concerns in local politics, resolutions in Dutch Parliament and the European AI act have put political pressure on developing an algorithm register.	Development of a standard to harmonise registers of different Dutch public organisations. Furthermore, citizen interests and needs regarding information provision on algorithms were researched.	A metadata standard consisting of a list of attributes (with descriptions) that should be explicated when including an algorithm in the register.	The register is further developed as part of a national program on digital government.
Procurement conditions	Public organisations mainly depend on external or private developers for their algorithms. In order to prevent obscuring the inner working of algorithms (e.g., through trade secrets) clear contracts about information provision by external parties should be concluded when systems are procured.	One of the participating municipalities already had developed a list of procurement conditions. The consortium tested these conditions in real-life cases.	Document with procurement conditions that public organisations can adopt in their own standard (IT) procurement agreements.	In the Netherlands, the conditions are planned to be adopted in local, provincial, and possibly national IT procurement regulations. Furthermore, the conditions are promoted internationally.
Objection procedures	Algorithms can form the basis for decisions on individual cases. When citizen seek recourse, they (as well as the organisations receiving the objection) need to have information about the role of algorithms in the decision. Often, it is unclear what information is needed and how to provide it.	In a real-life test case, the consortium explored the need of objection handlers regarding information on the role of algorithms in government decisions.	Instructions for objection handlers on what information they need and where they can find it.	Unknown but compared to the other three instruments, the instruction is not mentioned in policy documents published by, for example, national government after the end of the consortium.

5

Using the descriptions in Table 5.1, the consortium's instruments can be compared with approaches for attaining control over public algorithmic systems or with algorithmic governance approaches developed, described, and assessed in scientific literature. Situating the policy instruments in the literature discussing interventions in algorithmic practices shows that the instruments 1) focus on an organisational perspective, 2) prioritise policy instruments over technical instruments, 3) attempt to incorporate legalistic approaches, and 4) pay little attention to value-based approaches. We will discuss these four elements individually.

First, the instruments adopt an organisational perspective that seems to divert from notions of human-in-the-loop. Human-in-the-loop interventions heavily focus on human operators responsible for controlling or improving the working of an algorithmic system through interacting with the system (Rahwan, 2018). But fully confiding in the capability of individual people to intervene in algorithmic applications has its drawbacks (Green, 2022). For example, humans can be incorporated into the loop to ensure that autonomous systems fall under less stringent legal regimes, often leaving the human agent with limited agency but full responsibility for the system (Wagner, 2019). Furthermore, phenomena such as automation bias or algorithm aversion are hard to mitigate in practice (Green & Chen, 2019), and the effects of human behaviour on the working and quality of algorithms are still unclear (Peeters, 2020). The policy instruments respond to these drawbacks by establishing organisational control over algorithmic applications through the implementation of information flows and the assignment of responsibilities to actors, as shown in Table 5.1.

Second, the consortium focused on policy instruments rather than on technical instruments, such as bias detection tools, because they focused on policymakers. These technical tools prove to be challenging to implement in practice, and often fall short in providing adequate control when used in isolation (Balayn, 2023; Hutiri, 2023). The consortium explicitly chose to focus on policy instruments. The policy instruments are associated with initiatives such as legal and ethical frameworks, and impact assessments (Fest et al., 2022; Siffels et al., 2022).

Third, the instruments align with the legalistic approach attained by most public organisations. Especially on the European Union level, the discourse on governing algorithmic systems strongly emphasises regulating AI through legal instruments (Smuha, 2021). The AI Act[1] is the most prominent example of this approach. The algorithm governance framework of the consortium aimed to support public organisations in translating applicable laws to their own organisation and, thereby, ensure compliance with these laws.

Finally, the instruments have a procedural focus in which the role of values is unclear. In communities such as FAccT, Responsible AI, or Value-Sensitive Design, values are starting points to come to workable or practical algorithmic systems without harming humans. For example, debiasing tools are coming from these communities, but they also advance ways to implement ethical deliberation in AI systems (Dignum, 2019), and the translation of values to design requirements (Aizenberg & Van Den Hoven, 2020). In contrast, the policy instruments in Table 5.1 mostly prescribe procedures, leaving the responsibility for embodying values in algorithms to individual public organisations that will implement the policy instruments. For incorporating values, the consortium provided some starting points. For example, the consor-

1 At the time, the AI Act was already being drafted, but not final yet. The consortium tried to anticipate the implications of the act.

tium linked the algorithm governance framework to the principles for good digital governance formulated by the Ministry of Interior Affairs and Kingdom Relations.

5.2 Results: affordances and challenges in policy instruments

This section presents the results of our empirical study on the efforts of the consortium to develop the four policy instruments. We assessed the extent to which the instruments support in achieving the consortium's aim of attaining public control. The assessment is based on the theoretical lens presented in Chapter 2, i.e., institutional interventions in socio-technical systems grounded in principles of democracy and the Rule of Law. From this perspective, Tables 5.2 to 5.5 present the affordances and challenges of each individual policy instrument. *Affordances* reflect the potential of the instruments in strengthening public control mechanisms when the instrument is enacted in the prescribed way. *Challenges* are the shortcomings of the instruments that need to be overcome to establish public control. The affordances and challenges are based on our observations, our interviews, and the documents describing the instruments (see Section 3.3.1).

Algorithm governance framework

With the algorithm governance framework, the consortium aimed to bring different technical, ethical and legal frameworks together in one overarching framework, and to assign responsibilities and accountabilities to specific organisational roles. As Table 5.2 shows, these goals are reflected in the affordances of the instrument. However, the challenges listed in Table 5.2 also indicate that the usability of the framework is compromised, as it does not fully align with the idiosyncratic practice of design. The framework emphasises the organisational dimensions of control but disregards the technical aspects of public algorithmic systems. Moreover, the assumption that the design process can be captured in step-by-step routines is evident in the linear conception of the algorithmic system life cycle. Finally, the framework does not question established practices in public organisation that might be unfit for designing or using public algorithmic systems.

Algorithm register

Where the final algorithm governance framework partly reflects the goals set at the start of the consortium's project, the lack of a clear goal for the algorithm register undermined the register's development. The affordances presented in Table 5.3 show the potential of the register to serve multiple, distinct goals. However, the challenges show that, in practice, the register is only a tool for creating information flows between actors. As such, the register is mostly a means that does not achieve the end in itself. For the register to be effective to actors, it needs to be embedded in work practices and combined with other instruments. This requires convincing actors, within and outside of the organisation that implements a register, of the register's potential. However, before implementing the register, public organisations must first determine who will be served by it and, consequently, in what control mechanism it will play a role.

Table 5.2 Affordances and challenges of the algorithm governance framework

Algorithm governance framework	
Affordances	**Challenges**
Standardises efforts to control algorithmic systems The framework indicates what the legal and ethical boundaries are on the development of public algorithmic systems. One interviewee stated: 'If you tell me [data science team lead] what is possible or not, where the boundaries are, I can be creative within those boundaries. [trans.]' The framework prevents a recurring discussion in every development trajectory on what boundaries are to be considered. **Offers overview on all available means for control** Public organisations are confronted with an increasing supply of frameworks, assessments, and best practices to address harms coming from algorithmic systems. A governance framework can provide insight in the connections between instruments. However, since it was still unclear what means of control would be obligatory and to what extent they could be useful, creating an overview of all means turned out to be too ambitious for the consortium. **Provides smaller organisations with essential knowledge** Small public organisations often lack the capacity to organise control over the algorithms they use. A framework that is shared by several public organisations may compensate for this lack of capacity, for example, through knowledge sharing. **Supports assignment of tasks and responsibilities** The framework formulated by the consortium focuses on describing tasks and responsibility that are needed when algorithmic systems are developed, implemented, and used. This list of tasks and responsibilities (instead of roles) provides flexibility for every public organisation to operationalise the framework in their own processes and context.	**Technical aspects are unpronounced and unclear** The current framework mostly focuses on institutional and organisational aspects of algorithmic systems, for example, by emphasising legal, democratic, and societal legitimacy in combination with risk management. Although algorithms are the main topic, the framework misses a technical component and does not provide leads for ensuring technical quality. The institutional and the technical should be addressed in a systemic fashion (see Section 2.3.3). **Linear conception of life cycle** Although not reflecting practice within consortium organisations, the framework conceives the design process as linear. Upfront, the consortium intended to incorporate concepts such as agile and scrum in the framework, but this has only surfaced as arrows signifying feedback loops in the linear process. The linear conception contradicts with what is known about both design processes and policy-making processes within the consortium organisations. Still, the framework was given this basis to make it 'feasible'. **Formalisation of the design process is ambiguous and unattainable** In practice, it turned out to be difficult to define the design process of public algorithmic systems upfront. It appeared hard to find shared practices among public organisations. Moreover, the framework shows large variety in the level of detail of different elements. For example, the description of obtaining democratic and societal legitimacy is highly abstract, whereas the instructions to gain legal legitimacy are detailed and refer to legal instruments. **Based on conventional ideas** The framework reflects current practices in public organisations. It is still based on a separation between the 'client' and the 'supplier', although both may be part of the same organisation. Similarly, it follows the common distinction within organisations between development and deployment, rather than an iterative design and maintenance practice.

Table 5.3 Affordances and challenges of the algorithm register

Algorithm register	
Affordances	**Challenges**
Potential to have a pivotal role in a broader governance Considering the reliance on information flows in gaining control, public organisations can use the algorithm register to support the use of other instruments. Serving different goals (e.g., providing transparency, information provision, internal overview of algorithms) the register can address information asymmetries between actors. **Can steer the conception of algorithmic systems** The metadata standard formulated by the consortium presents public algorithmic systems as socio-technical. The register does not only include information about the technology but also refers to rules and regulations and the goal of the system. Accordingly, it may institute a socio-technical lens on algorithmic systems within public organisations. As the standard was developed, there was a continuous debate about whether the register should mostly focus on technical specifications or provide a broader lens. **Has a function during the whole life cycle** The register can serve different goals throughout the design and use phases of public algorithmic systems. During development, the meta-data standard can be used by designers or other public servants as a design guide for or to identify blind spots in a system. Thereafter, the register can serve archiving purposes, information provision (e.g. to journalists or representatives), or even stimulate learning from failed design processes when discarded systems are included in the register.	**Heterogeneity in need for information** There is great diversity in potential user groups of the register. For each group, the register serves a different goal. For example, a citizen can use the register for dissent purposes, where a public servants can use it for an internal overview of algorithms used in their organisation. The consortium struggled to decide on the final user group and goal but also did not explore possibilities for differentiation of information within the register. **Reservations among public servants** Within public organisations, public servants confronted with the register often asked questions about workability or feasibility. Furthermore, some actors expected superfluous commotion and questions when presenting algorithms used by the organisation. **The register is a means for attaining transparency, not a guarantee** A register is one way of information provision, and other forms may serve specific needs, for example of citizens, better. Still, in potential, an algorithm register can facilitate a variety of means for transparency. Moreover, the extent to which transparency is achieved, greatly depends on how the register is used and maintained in practice. **A register needs to be firmly embedded in organisations** It turned out to be quite a challenge for consortium partners to collect information on all algorithms used within their own organisations. Besides, changes in algorithms should also be processed in the register. Considering the scale of public organisations and the different departments that algorithms are used in, a governance approach for developing and maintaining the register is critical to its effectiveness.

Procurement conditions

Like the algorithm register, the procurement conditions focus on creating information flows. However, the conditions are specifically focused on the flow of information from external developers to public organisations in the case of an audit of externally developed algorithmic applications. This specific focus is also a pitfall of the instrument, see Table 5.4. The merit of the instrument is signified by the uptake of the procurement conditions by several public organisations in their general procurement policies. This uptake can be explained by the conditions' alignment with pre-existing procurement practices. On the other hand, this will also reinforce the existing procurement practice and its associated procurement mindset discussed in Section 4.2.

Table 5.4 Affordances and challenges of the procurement conditions

Procurement conditions	
Affordances	**Challenges**
Strengthens position of public organisations The procurement conditions are likely to be adopted in ICT procurement conditions of local and regional umbrella organisations. This may increase the use of the conditions and, therefore, result in a stronger position of public organisations towards external parties. Especially for public organisations that lack the experience or resources to negotiate with external parties. **A means to obtain information about externally developed systems** The conditions oblige suppliers of algorithmic applications or systems to provide mostly technical information when public organisations need such information for, for example, audits. Formerly, public organisations often could not retrieve such information. **Aligns with current procurement procedures** The rationale of this instrument is to align with current procedures in public administration. Public organisations are familiar with procurement of ICT technology and have standardised that process. Moreover, most algorithmic applications are procured by public organisations. Therefore, including specific conditions for algorithmic systems procurement conditions is a quick win for establishing public control.	**Assumes that external parties only provide a technical artefact** Using procurement conditions re-emphasises the idea of the external developer only providing a technical artefact. It does not address the shift of normative design choices towards private parties. This may be problematic since public algorithmic systems are intricately entangled with public policy and the execution thereof and have implications for how processes are organized. **Too much focus on information provision** The focus on obtaining information about externally developed systems, overshadows quality criteria for the algorithmic systems. One interviewee stated that including strict quality criteria is unattainable for the large diversity in types of algorithmic systems used. Furthermore, procurement procedures already contain quality checks. **Relation to life-cycle unclear** The conditions mostly focus on the use phase of algorithmic systems. Therefore, the dynamic between public organisation and external developer during design phases is not explicitly formalised through the conditions. Furthermore, the conditions are limited in describing what happens or who is responsible when undesirable and/or unintended consequences emerge caused by or because of the algorithmic system.

Instructions for objection procedures

The instructions for objection procedures did not make the expected impact in practice. Working on the instructions mostly made the consortium aware that the settled practice of objection procedures is hard to change. So instead of changing those objection procedures, the consortium developed an instruction to assist objection handlers in asking the right questions about algorithmic systems and to raise awareness among handlers about the role of algorithmic applications in decision-making. As reflected in the affordances and challenges in Table 5.5, the efforts related to the objection procedures underline the already acknowledged need for a more structural transformation of these procedures. Currently, flaws on a system level are hard to identify in individual-based objection procedures.

Table 5.5 Affordances and challenges of the instruction for objection procedures

Instruction for objection procedures	
Affordances	**Challenges**
Shows that structures for control are available, but that algorithmic systems should be situated within those structures Like the procurement procedures, objection procedures are established practices in public administration. However, the role of algorithmic systems gives a new dimension to the legalistic procedures currently in place. Therefore, the attempt to instruct objection handlers on how to address algorithmic systems involved in decision-making might improve objection procedures. **Indicates the need for organisational or system change** While providing objection handlers with tools, the explorations and test cases performed for this instrument show the need for a change in the organisational structure of public organisations. The use of algorithmic systems changes decision-making in such a way that current objection procedures do not suffice in providing citizens with possibilities for recourse anymore.	**Focus on later stages in life cycle** This instrument focuses on the situation that a decision in which an algorithmic system was involved is taken and the affected citizen seeks recourse. It does not allow citizens to voice their questions and/or criticism about the development or use of a specific algorithmic system before decisions are made. **Focus on detecting and correcting errors in specific individual case** The deliverable of the consortium aligns with current administrative and judicial procedures that aim to correct incorrect decisions in individual cases. However, this does not provide a procedure to tackle structural errors, requiring maintenance or redesign of the algorithmic systems, that may not be revealed when an individual citizen seeks recourse.

5.3 Analysis of policy instruments

The main goal of the consortium was to develop policy instruments that would establish public control over algorithmic systems. To restate the consortium's interpretation of public control: 'For us, public control means that the society – citizens, businesses, government, and all other actors playing a role in public control – has insight in and has a say about algorithms, even if they are complex. […] If we fail to facilitate this for citizens and businesses, trust in the government will disappear. [trans.]' Our analysis in Tables 5.2 to 5.5 shows that the consortium operationalised this definition by mostly focusing on establishing information flows and assigning responsibilities and accountability through the four policy instruments. The algorithm register, the procurement conditions, and the objection procedures mostly provide insight into the working of public algorithmic systems by facilitating information flows between actors who design and operate the system. In assigning tasks and responsibilities, the consortium prioritised institutionalising frameworks and impact assessments executed by public servants (be they policymakers, data scientists, or other positions) to ensure that algorithmic systems are developed and used more conscientiously. Furthermore, several consortium participants were considering establishing a new role (the algorithm expert), aiming to fill accountability gaps.

Tables 5.2 to 5.5 also show that the observed Dutch public organisation were still at the beginning of establishing public control over algorithmic systems. The consortium stated that it wanted to facilitate public control, but the instruments indicate that the consortium did not fully accomplish this at the end of their program. This section discusses three impediments to establishing public control that can be generalised from the instruments: 1) ambiguity in the concept of public control over algorithmic systems, 2) the socio-technical nature of control over (public) algorithmic systems is disregarded, and 3) control mechanisms are not fully embedded a public administration context. These three impediments indicate that public organisations are still acting in accordance with the technocratic and businesslike presuppositions identified in current design practices (see Chapter 4).

Ambiguity in the concept of public control over algorithmic systems

Three forms of ambiguity are prevalent in the four policy instruments: unclear goals due to fragmentation, vagueness in definitions and instructions, and discrepancy between plans and practice. First, the consortium chose to conceive public control through four policy instruments with diverging aims, and effects on different levels and processes. This fragmentation was reinforced by the lack of reflection on how the particular instruments contributed to the starting conception of public control. The goals of the instruments were often unclear or disputed by participants, which may be explained by the consortium's emphasis on incremental exploration and implementation of instruments. Although the instruments increased awareness and indicated gaps in knowledge on control over algorithmic systems, they will probably not suffice in achieving the intended control over these systems. This is most notable in the efforts on the objection procedures that identified the need to adapt public organisations on a more structural level instead of implementing a single policy instrument.

Second, the use of instruments was obscured for the end user because the consortium left different forms of vagueness in the instruments. The consortium listed values for good governance on algorithmic public administration in their algorithm governance framework but did not address the trade-offs that need to be made between these values. The consortium considered this to be the authority of political actors. As a result, the value trade-offs remain unclear in the policy instruments, and the consortium's insights into these trade-offs are not well-documented. Apart from political reasons, vagueness arose from uncertainty or limited knowledge. For example, in all four instruments, the consortium considered 'fundamental changes [trans.]' in algorithmic systems as a key element to determine the moments when public control efforts need to be executed. However, the individual public organisations were assigned the responsibility for defining fundamental changes, and the consortium did not provide an elaboration of the term fundamental changes. Finally, the consortium often used vague concepts in their documents because they were still waiting for official support from their own organisations. Mostly, they could not settle discussions on the availability of capacity for deploying the instrument within their organisations yet. All participants acknowledged the importance of public control but were keen on keeping the instruments, or control in general, 'feasible'. Colleagues not directly involved in the consortium considered most alternatives for the instruments to require extensive organisational capacity.

Third, there is a discrepancy between the ideas reflected in the instruments and actual algorithmic practices. The consortium provides two examples. First, the policy instruments present the life cycle of algorithmic systems as a linear process. Second, the instruments address sub-problems and focus on specific steps in the life cycle. These choices within the instruments contradict with the idiosyncratic nature of the actual life cycle.

Ambiguity in the instruments is inevitable, considering the great variety in understanding of (public) control across scientific disciplines. Notions of control in general vary from purely technical to comprehensively socio-technical (e.g., including social, organisational, and institutional elements) interpretations. These interpretations have implications on what needs to be controlled: either individual accidents or system hazards; either system components or emergent properties (Leveson, 2012). Concentrating on the democratic element in public control as a concept, terms such as democratic, public, political, or popular are used interchangeably as qualifications for control. Despite their similarities, the different terms have different applications. For example, political control is often used to classify the control that bureaucrats have over political decision-making (Moe, 2006). The perspective used in this thesis aligns most with democratic control, which emphasises the central position of citizens in controlling algorithmic systems, but also the role elected representatives and other democratic institutions play in control. Following Hajer (1995), democratic control over designing technological systems is mostly about bringing politics into these processes.

As discussed in Section 5.1, the literature on AI and algorithmic systems can also be characterised by its variety in perspectives on control over algorithmic systems. Perspectives vary from emphasising human control, institutional mechanisms, to value-based approaches. The ongoing discussion on how these different approaches of control are to be conceptualised shows that a common definition has not crystallised yet. The ambiguity in the control concept not only is an intellectual challenge. It also hinders public organisations in developing and implementing effective policy instruments.

Public control over algorithmic systems is socio-technical

The analysis of the instruments shows an oversight of the socio-technical nature of public algorithmic systems and, therefore, the instruments do not approach control as a socio-technical exercise. The fragmentation of the instruments into tech-focused or legal-focused policy instruments disregards the emergent properties of algorithmic systems. The policy instruments either focus on technical aspects – e.g., the procurement conditions and objection procedures – or on legal and policy aspects – e.g., the algorithm governance framework. The algorithm register shows a first step in bridging the strict separation between institutional and technical components, as the meta-standard clearly incorporates both components.

Considered from a socio-technical perspective, emergence and uncertainty are properties of public algorithmic systems. Emergence means that system behaviour cannot fully be explained by the behaviour of its individual components (Checkland, 1999). Consequently, the forms and functions of algorithmic systems evolve through the various stages of their life cycle. In contrast, the policy instruments present the design process or life cycle as linear and assume that moments in which a system changes fundamentally can be identified. The idiosyncratic nature of the life cycle is ignored, and the need for a continuous process of anticipation and reaction is not manifested in the instruments.

Finally, socio-technical systems are complex and control over them is too. Public algorithmic systems are situated in a multi-stakeholder context, which makes both designing and using algorithmic systems a complex, multidisciplinary effort (De Bruijn & Herder, 2009). Such a multidisciplinary approach conflicts with the siloed structure often found in public organisations and their focus on instruments singling out specific dimensions rather than building an integral picture of the system. The separation of ICT and policy-heavy departments further impedes effective control over these systems (see Chapter 4). More general, public control over algorithmic systems is context dependent. Workable and proper system boundaries have to be drawn for every specific system (Leveson, 2012). Moreover, given the complexity of socio-technical systems, there is not a particular one-size-fits-all control mechanism; control is a contextual exercise for which a coherent set of control mechanisms is needed. The case of cancelling the illegitimate System Risk Indication (SyRI) system in the Netherlands illustrates how socio-technical complexity and emergence render accountability mechanisms multi-faceted – both complicating and strengthening public control (Wieringa, 2023). At first, political actors did not recognise the hazardous implications of SyRI. A civil society coalition starting strategic litigation was needed to increase political interest (Wieringa, 2023).

Situatedness in public administration context

The consortium's conception of public control comprises two aspects that are fundamental to public administration, which is organised based on democracy and the Rule of Law. The conception emphasises the importance of informing citizens and giving them a say in designing and using algorithmic systems. Considering Tables 5.2 to 5.5, it is debatable whether the four instruments support this goal of the consortium. Ensuring a say for citizens can be observed in the improvements of objection procedures and notions such as democratic and societal legitimacy in the algorithm governance framework. Nonetheless, the efforts of the consortium showed the difficulties of involving citizens. During the year that the consortium worked on public control, it often discussed the information position of citizens and performed a study on citizens' information needs about algorithmic systems. Informing citizens seemed to be a tough nut to crack because of the different wishes and needs among citizens. But the observation that citizens often do not even know when or how algorithms are used as part of policies or decision-making indicates that low awareness and distance to the topic are impeding citizens in participating in the public debate regarding public algorithmic systems and their possibilities to seek recourse. Therefore, the consortium also considered strengthening institutions that can protect the position of citizens. For example, the algorithm register can also be used by judges, journalists, or researchers to scrutinise algorithmic systems. It remains to be seen whether and how these developments will truly contribute to increasing control over algorithmic systems by the public.

Remarkably, while the citizen was the primary motivation for the consortium to explore means for public control, the year their attention gradually shifted to internal issues regarding civil servants and their daily practices during the year. The consortium mostly adopted a procedural interpretation of public control by making it a predominantly administrative and operational exercise. This internal focus clashes with their aim to give the citizen a central role in public control.

The procedural focus at the end of the project might have been the result of internal struggles within the public organisations. Reservations within the organisations about the instruments were impeding the development and implementation of the instruments. Consortium members were confronted with reservations of colleagues when working on the algorithm register (e.g., while searching for algorithms to include in the register). Their colleagues worried about the extra work and financial investments that might follow from implementing the instruments. Similarly, there was reluctance or resistance to structural changes within the organisations. In order to be effective, approaches to public control need to be accepted and shared within organisations, supported by evidence of their efficacy, as well as accompanied with financial resources, skills and knowledge.

So, new forms of public control are needed but also need to be embedded in existing practices within public administration. Currently, it appears that public organisations struggle to find a balance between conventional practices and routines, and the effects of incorporating new data science practices. When innovating policy instruments, public organisations should make their role of serving and protecting citizens central. Without this focus, the lack of democratic legitimacy of algorithmic systems (see Grimmelikhuijsen & Meijer, 2022) will remain unresolved.

5.4 Conclusion

This thesis aims to prescribe institutional interventions that embed the design processes of public algorithmic systems within established democratic and Rule of Law practices. Public organisations already undertake similar initiatives. This chapter examined the development of policy instruments in Dutch public organisations to learn from their efforts in exploring institutional interventions in algorithmic practices. Our analysis focused on four policy instruments developed by the consortium studied in Chapter 4.

The consortium developed an algorithm register, an algorithm governance framework, procurement conditions for algorithmic applications, and instructions for objection procedures in which public algorithmic systems are involved. As such, the consortium deliberately focused on institutional interventions instead of technical instruments. The policy instruments of the consortium aim to transform general algorithmic practices within public organisations. More specifically, their goal was to establish public control over algorithmic applications.

To achieve the goal of public control over algorithms, our analysis shows that public organisations focused on three elements throughout the four policy instruments. All instruments contribute to establishing structural information flows between designers, demarcate tasks and responsibilities of actors in the design process, and aim to raise awareness among public servants about the need to control risks of public algorithmic systems. On the other hand, policy instruments suffer from shortcomings that impede the establishment of public control over public algorithmic systems. The policy instruments contain critical yet vague elements, provide a fragmented approach to public control, disregard the emergent properties of algorithmic systems, and do not specify the role of citizens or politics in public control. Moreover, while the organisations in the consortium explicitly pursued public control, they lacked a clear understanding of what public control means. This chapter shows that defaulting to policy instruments as institutional interventions without a comprehensive vision on the instruments' goal, neglects the need for structural organisational change and disregards the socio-technical nature of algorithmic practices in public organisations.

Since design practices of public algorithmic systems are a specific category of algorithmic practice, the set of policy instruments also relates to the interactions between designers of public algorithmic systems. The assessment of the policy instruments provides insight into requirements for institutional interventions as perceived by practitioners.

Considering the three elements that the consortium focused on in developing policy instruments, we can formulate three requirements for institutional interventions:

- *Establish structural information flows between designers*
- *Demarcate tasks and responsibilities of actors in the design process*
- *Raise awareness among public servants about the need to control the risk of public algorithmic systems*

Similarly, the shortcomings in the consortium's approach provide requirements for institutional interventions in design practices:

- *Establish an unambiguous goal for institutional interventions*
  Our analysis shows that there is no quick fix for design practices. Reshaping these practices is a complex endeavour that needs to be preceded by resolving or reducing the ambiguity of the goals behind institutional interventions.
- *Realise a mix of interventions to reshape design practices (cf. Hutiri, 2023)*
  The algorithm register, for example, shows that it lacks added value when used as the only intervention. It should be embedded in or combined with other instruments and practices. This mix of instruments and practices must embody both procedural, normative, and substantive instruments. In other words, institutional design should not only emphasise procedural interventions.
- *Acknowledge that institutional interventions in design practices are not a panacea (cf. Ostrom, 2007)*
  Practices cannot be fully controlled but are dependent on the behaviour of involved actors. A way to deal with this phenomenon is to organise design practices as reflexive or adaptive exercises in which designers anticipate and react to emergent properties of algorithmic systems. Accordingly, institutional design should not aim for formalising the design process or imposing rigid methods such as agile. Instead, institutions should stimulate professionalism among designers of public algorithmic systems.
- *Increase the democratic legitimacy of design practices by specifying the role of citizens and political debates*
  The interpretation of public control by the consortium does not account for the complexity arising from societal and organisational processes and arising from trade-offs related to citizen rights which are fundamental to public administration. Instead, institutional interventions should aim to increase the democratic legitimacy of design practices by specifying the role of citizens and political debate.

This chapter also closes the first part of this thesis. Part I aimed to identify the presuppositions that underlie current design practices for public algorithmic systems. Chapter 4 showed that technocratic and businesslike presuppositions explain the institutions and attitudes that have emerged in design practices. These presuppositions are also related to impediments to interactions between designers. Accordingly, Chapter 4 suggests that public organisations should base their design practices, and, therefore, their institutional interventions, on different presuppositions. The findings of this chapter indicate that public organisations still fall back on the technocratic and businesslike presuppositions of current design practices when developing institutional interventions. The policy instruments focus on procedures, and centralise the expert (i.e., public servant) in decision-making regarding design choices. Like the institutions and attitudes within current design practices, the institutional interventions currently developed do not correspond with the type of systems that are being designed. At the same time, the consortium did not have the mandate to fundamentally transform practices within their public organisations. In the following chapters, we use the lessons learned by and from the consortium to take a next step in improving design practices of public algorithmic systems.

Part II

Appraisal

Chapter 6

Algorithmic Kafka[1]

> And life is like a pipe
> And I'm a tiny penny rollin' up the walls inside
> – *Amy Winehouse; Back To Black*

Algorithmic systems are prone to errors, misconduct, and unintended consequences that can eventually harm citizens and other actors involved and/or affected by the use of the applications (e.g., front-line workers) (Eubanks, 2017). As such, public algorithmic systems are no different than other automated systems. They are likely to fail at some point (Lindgren, 2023), and this failure can inflict harm on citizens. Infamous examples such as the Robodebt scheme in Australia and the childcare allowances scandal in the Netherlands show how public algorithmic systems are prone to inflicting psychological, physical, and emotional harm on citizens.

To understand the failures in public algorithmic systems that lead to these harms, we examine the emergence of such systems that put citizens in Kafkaesque situations. Other authors have already indicated the likelihood of Kafkaesque situations in public algorithmic systems (see Bayamlıoğlu & Leenes, 2018; De Laat, 2019; Ossewaarde, 2023; Susser, 2016). However, these authors only indicate the phenomenon of algorithmic Kafka and refrain from specifying the characteristics of the phenomenon. The lack of insight into the origins of Kafkaesque situations hinders the adequate and timely anticipation and reaction to causes for algorithmic Kafka in the design process of public algorithmic systems. Therefore, this chapter explores the role of designers in the creation of algorithmic Kafka (through the design choices that designers make) by answering the following question:

What is the role of designers in the emergence of Kafkaesque situations in public algorithmic systems?

This chapter concentrates on the influence that designers have on the manifestation of public algorithmic systems by formulating the socio-technical specification of these systems. In other words, we relate the socio-technical specification to the behaviour and interactions of designers. Being a building block for the design theory in Chapter 8, this chapter describes the

1 This chapter draws on Nouws, S., & Dobbe, R. (2024). The Rule of Law for Artificial Intelligence in Public Administration: A System Safety Perspective. In K. Prifti, E. Demir, J. Krämer, K. Heine, & E. Stamhuis (Eds.), *Digital Governance* (Vol. 39, pp. 183–208). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-639-0_9

interactions that designers should avoid in formulating socio-technical specifications of public algorithmic systems.

The research methods behind the analysis in this chapter are discussed in Section 3.3.2. This chapter starts with describing four well-studied cases of public algorithmic systems that inflicted harm on citizens in Section 6.1. These cases provide insight into the manifestation of harms inflicted by public algorithmic systems on citizens in practice. Section 6.2 analyses the algorithmic systems in the four cases from the perspective of Kafkaesque situations and arbitrary use of power. Thereafter, Section 6.3 examines the role of designers and design practices in creating the possibilities for arbitrary conduct in the socio-technical specification of algorithmic systems. Section 6.4 concludes this chapter by comparing the design practices identified in Section 6.3 with the current design practices in Dutch public organisations discussed in Part I.

6.1 Harmful public algorithmic systems

Scholars have described a transformation in paradigms within public organisations instigated by the possibilities that algorithmic applications provide to analyse data. Dencik et al. (2019, p. 18), for example, discuss 'the pursuit of a 'golden views' of citizens in which citizens are positioned … not as participants or co-creators, but primarily as (potential) risks, unable to engage with or challenge decisions that govern their lives.' Other scholars write about the anticipation of behaviour (Peeters & Schuilenburg, 2018) or new public analytics (Yeung, 2023) as emerging paradigms in public sector administration driven by the implementation of algorithmic applications. As a result, the newest applications of ICT systems in government have become more intrusive, more rigid, and more obscure (Alkhatib, 2021).

Unsurprisingly, using systems that include algorithmic applications based on the paradigms discussed above can result in harm inflicted on citizens. Well-studied cases of algorithmic systems in different jurisdictions show the detrimental impact on citizens. This section describes four of these cases to understand how different components and the interactions between components of algorithmic systems are involved in creating such harms. The four cases are selected based on the criteria listed in Section 3.3.2. The case descriptions are based on a secondary analysis of available scientific and grey literature. First, Section 6.1.1 discusses the Dutch childcare allowances scandal in which false accusations of fraud have pushed a large group of citizens into debt, resulting in the tragic destruction of people's lives (Peeters & Widlak, 2023). Thereafter, Section 6.1.2 discusses the Robodebt scheme. This Australian algorithmic system – a data-matching algorithm that automated debt detection among welfare recipients – was discontinued after a Federal Court ruling that followed years of criticism on the algorithm by, amongst others, victims and national inquiries (Braithwaite, 2020). Section 6.1.3 highlights two similar cases: the British Post Office scandal, and the Dutch DUO case. Section 6.1.4 presents a comparison between the four cases based on our descriptions.

6.1.1 Childcare allowances scandal

In the Dutch childcare allowances scandal, a flawed and discriminatory risk indication model was used to identify fraudulent recipients of childcare allowances. The model produced many false positive risk scores for recipients. These risk scores were not adequately scrutinised by the tax office, leading to false accusations of many households. In line with harsh anti-fraud policies, accused recipients had to pay back large amounts of money to the tax office. Initially, this put these recipients into high debt, which eventually affected other aspects of their own personal lives and the lives of those related to the recipients. Konaté and Pali (2023) interviewed 5 victims of the scandal, and distinguish five ways in which harms were inflicted on citizens: (1) consequences for the physical, mental, emotional, and material wellbeing of the victims, (2) taking away chances for their children to grow up in a safe and quiet environment, (3) impact on their relationships with others, (4) impact on their attitudes towards state institutions)especially due to obscureness, responsibility gaps, etc), and (5) impact on their hopes and needs for the future. These harms are similar to those experienced by unguilty but imprisoned people, and by parents whose children are falsely placed into custody (Konaté & Pali, 2023). Only after 10 years, the problem of false accusations was acknowledged by the government (Bouwmeester, 2023), but today, most victims are still waiting for compensation.

The infamous Dutch childcare allowances scandal shows how both the technical and institutional components of public algorithmic systems have a role in the emergence of algorithmic harms. From a technical perspective, the risk indication model used in the scandal can be considered an algorithmic application. Investigation reports identify the model as a self-learning model, but the type and level of learning are not fully clear. What is clear is that the model largely relied on attributes like foreign-sounding names and, therefore, harmed already vulnerable groups. Moreover, the model was trained on a dataset in which cases of previous allowance requests – including cases with errors, omissions, and mistakes in the request – were flagged as fraud (Peeters & Widlak, 2023). Consequently, unintentional flaws in allowance request forms were interpreted as cases of deliberate fraud.

Apart from the condemned risk indication model, Peeters & Widlak (2023) identify flaws in the information architecture in which the model was embedded. They argue that before the tax office considered the use of a risk indication model, the information infrastructure in which the model was to be embedded already constrained the agency of human decision-makers and front-line workers. Moreover, data ownership was separated from the actors responsible for decision-making. Consequently, 'the tax authority failed … to understand the reasoning underlying its own administrative decisions to terminate daycare benefits' (Peeters & Widlak, 2023, p. 871). Finally, the interaction between the algorithmic system and the bureaucrats in the organisation, changed the work practice of those bureaucrats. They were presented with risk scores related to cases, without information about that score or the model that resulted in the risk score. Consequently, the front-line workers were not able to scrutinise and adjust the 'decision' (i.e., risk score) of the algorithmic application (Giest & Klievink, 2024; Oldenhof et al., 2024).

Apart from technical issues, the origins of the scandal partly go back to institutional and political factors. Two moments are often indicated as pivotal: (1) a major tax reform, in 2005, in which the allowances system was created and the responsibility for executing the system was assigned to the tax office; and (2) the introduction of rigid anti-fraud laws and policies in reaction to what is known as the 'Bulgarian fraud incident' around 2013 (Fenger & Simonse, 2024). These and other events created a political push to tackle fraud and resulted in a rigid interpretation of policy that put a heavy burden on citizens accused of fraud. Suspected citizens had to provide detailed information about their personal situation, and, once accused, had to pay back all received allowances, even when only small mistakes in the allowance request were made (Peeters & Widlak, 2023).

Over time, the public algorithmic systems (i.e., the combination of the risk indication model, information infrastructure, and anti-fraud policies) affected more and more citizens who were truly eligible for childcare allowances, instead of detecting and addressing actual fraud (Fenger & Simonse, 2024). However, it took several years before the scandal was widely acknowledged by politicians and the media. The full extent of the scandal was also obscured because of failing institutional mechanisms that were in place to protect citizens (Bouwmeester, 2023). Parliament neglected its control task since it was too focused on reducing fraud and overly optimistic about digital and automated approaches in public administration. On the other hand, requests for information about the effects of the fraud policy by members of Parliament were frustrated by the government. Lawcourts, performing judicial control, followed the rigid interpretation of anti-fraud laws of the tax office and, thereby, lost sight of the interests of citizens (Bouwmeester, 2023).

6.1.2 Robodebt scheme

In the Robodebt scheme, the Australian government collected overpayments of unemployment benefits through the Online Compliance Intervention (OCI) program executed by Centrelink, the national income support agency. The main aim of the OCI was to cut social welfare spendings by increasing efficiency through digitalisation and automation (Whiteford, 2021). However, the calculations of overpayments (i.e., debts) were full of mistakes (Galloway, 2017; Carney, 2019; Braithwaite, 2020). In the end, the OCI program was comparatively expensive, as the government had to compensate victims of falsely assigned debts after a successful class action (Whiteford, 2021). Similar to the childcare allowances scandal, many citizens were harmed by the Robodebt scheme. Recipients of unemployment benefits fell into debt, which resulted in emotional and psychological harms, and losing trust in the government (Clarke et al., 2024).

Regarding the technical component, the government framed OCI as a big data project. Where early descriptions of the scheme followed this frame (Galloway, 2017), over time scholars observed that the algorithmic decision-making solution was very limited (Rinta-Kahila et al., 2022). The algorithmic application calculated the fortnightly income by dividing the yearly income by 26 and compared that to the amount of benefit received. Considering that fortnight income can fluctuate considerably over a year, it should have been clear from the start that discrepancies between the calculated and real fortnight income would emerge (Braithwaite, 2020; Whelan, 2020; Whiteford, 2021). This design choice was a break with standing social welfare policies, and, therefore, Whiteford (2021) concludes that designers purposefully used yearly income in the calculations. In hindsight, it is clear that most overpayments were the consequence of mistakes by recipients (e.g., errors in filled out forms), not of intentional fraud (Whiteford, 2021).

Another break with earlier approaches in social welfare was the lifting of human oversight over the system. This enabled Centrelink to increase the scale of the program since less oversight provides room for an increase in the number of calculations (Whiteford, 2021). The oversight was transferred to citizens as they had to check whether the OCI calculations were correct (Whelan, 2020). In this case, technical systems related to the OCI caused more problems. First of all, recipients had to provide information about their income for the past 7 years, which raised problems such as the inability to obtain income data from former employers (Whelan, 2020). Moreover, citizens were supposed to communicate through the myGov portal.

At the start of OCI, less than half of the Australian population was using the portal (Whiteford, 2021). Recipients who did not use myGov could not be reached but were still registered as debtors.

Considering the scheme from an institutional perspective, what stands out are the barriers to finding recourse. Citizens were not informed where they needed to go to appeal the overpayment decision (Braithwaite, 2020). Being ill-informed, in combination with a reputation that it was impossible to reverse the decision, many accused recipients choose to pay the money and not use the possibility of appeal (Rinta-Kahila et al., 2022). In cases where people used their right to appeal, Centrelink settled the case with the individual recipient to halt the legal procedure (Whiteford, 2021). Apart from this barrier, Carney (2019, p. 4) lists seven additional 'serious structural deficiencies and oversights in the design and operation of accountability and remedial avenues'.

Other institutional drivers behind the Robodebt scheme are the 'welfare ideology and economic imperative' behind the OCI program (Rinta-Kahila et al., 2022), and the changes in work instructions. The former is an important argument for Whiteford (2021) to conclude that the system was intentionally designed as such; after all, money needed to be saved. The latter refers to a shift of the responsibility of checking anomalies from public servants to the accused recipient (Whelan, 2020). This new interpretation of standing social welfare policy, meant a transfer of temporal burdens to recipients as they had to provide information, check calculations, and provide proof (Whelan, 2020).

6.1.3 Other Kafkaesque algorithmic systems

The childcare allowances scandal and Robodebt scheme are not isolated or unique cases. This section discusses the emergence of Kafkaesque situations in public algorithmic systems in two other cases: the British Post Office Scandal and the DUO case in the Netherlands. We added these two cases for the aim of saturation. They confirm the findings of the more extensively studied childcare allowances and Robodebt cases. The Post Office scandal does not revolve around an algorithmic application, but an IT system for accounting. The implementation of this IT system is grounded on similar grounds (e.g., efficiency) as the algorithmic systems in the Dutch and Australian cases. Moreover, the implemented IT system is a form of automation that preceded the current focus on algorithmic systems. Finally, the harms inflicted on citizens, the political inaction, and the public indignation are similar to the childcare allowances scandal and the Robodebt scheme. The DUO case came into the public eye after the childcare allowances scandal and shows that other Dutch public organisations deployed similar algorithmic systems. Being a recent case, there are no scientific analyses of the case, but several investigations have been published in grey literature. We included this case as it shows the shortcomings of individual recourse regarding false accusations by algorithmic applications.

Post Office Scandal

In 1999, the British Post Office implemented an IT system called Horizon to automate administrative and accounting tasks in Post Office branches. These post offices, especially those in more peripheral areas, are typically run by so-called sub post masters. Often, these are small business owners who run a post office in addition to their main business. As soon as Horizon was implemented in local post offices, discrepancies in bookkeeping (e.g., between money registered and money received) were observed. Eventually, the IT systems turned out to be plagued by bugs, which both the software vendor and the Post Office were aware of from 2011 onwards (Christie, 2020).

Several sub post masters complained about the errors in the system, but their complaints were filed in isolation. As a result, the sub post masters were not aware of complaints by their colleagues, and, therefore, lacked an overview of all complaints (Augustine et al., 2024; Georgiadou et al., 2024). Moreover, complaining sub post masters were not believed, since the Post Office had a high reputation of being a reliable organisation (Augustine et al., 2024; Marshall, 2022). Therefore, the complaints did not lead to political action. A miniseries on the story of sub post master Bates (one of the first complainers) in 2024 resulted in the problems taken up by politicians (Georgiadou et al., 2024).

Apart from dismissing the complaints, the Post Office started prosecuting sub post masters because of the discrepancies in their records. The discrepancies were interpreted as fraud, not as failures of the Horizon system. Sub post masters were accused of theft and false accounting (Christie, 2020). Remarkably, the Post Office had a special status; it could prosecute fraudsters themselves (Christie, 2020). In addition, they conducted the investigations leading to the accusation themselves; an uncommon concentration of powers in one organisation (Christie, 2020). Sub post masters were often convicted. This can partly be explained by the presumption in British law that IT systems are always right and reliable (Christie, 2020). Apart from being a problematic presumption (the presumption was renounced in a court appeal in 2019), judges and juries lacked the expertise or knowledge to assess whether IT systems are reliable (Marshall, 2022). The convicted sub post masters often fell into high debts, saw their reputation in the community damaged, and suffered from mental health issues (Growns et al., 2024).

DUO

The DUO case in the Netherlands is another example in which the complexity of a public algorithmic system resulted in harm to citizens. Until 2015, DUO (an executive agency of the Ministry of Education, Culture and Science) paid Dutch students monthly study grants to cover costs of living. Students living with their parents received a lower amount than students who lived on their own, as the cost of living of the latter group was deemed higher. Whether students lived with their parents or not was determined based on the address registration of the students. A public algorithmic system was implemented to detect fraud with the grants. A rule-based algorithm made a first selection of potential fraudsters. Thereafter, public servants of DUO refined this selection. External bureaus checked the selected cases by paying house visits. The decision that followed could be contested, in the first instance, at DUO. If the complaint was denied, students could challenge the decision in court (Algorithm Audit, 2024a; PWC, 2024)

In the summer of 2023, the discriminatory working of the public algorithmic systems and the unlawfulness of accusations surfaced. Lawyers were noticing that a high number of accused students had non-Dutch ethnic backgrounds. Moreover, most accusations of fraud were annulled in court. Although there were several flaws in the institutional design, such as the illegitimacy of the house visits by external parties, the DUO case mostly shows that the algorithm was biased and that this bias was not resolved by putting a human-in-the-loop – i.e., the public servants of DUO. More surprisingly, DUO did not test its system on bias (Algorithm Audit, 2024b; Autoriteit Persoonsgegevens, 2024). It took several years and considerable effort from lawyers and journalists to detect this system flaw. Accordingly, the DUO case shows the lack of system-level correction by only detecting and correcting individual, wrongful cases.

6.1.4 Similarities between harmful public algorithmic systems

The four cases show how algorithmic applications, in interaction with the technical and institutional environment they are embedded in, can result in detrimental harms inflicted on individual citizens. This section elaborates on the similarities between the cases as a starting point for increasing our understanding of algorithmic harms and their causes. The cases are similar in the type of harm inflicted on citizens, the time it took for these harms to become universally acknowledged, and the interactions between socio-technical components as causes of the algorithmic harms.

First, the output of the algorithmic applications in the described cases ultimately resulted in false accusations of citizens. These citizens had to pay money and saw their reputation within their community destroyed. As victims were often not able to meaningfully object to decisions made by the public algorithmic systems, they saw their personal lives changed drastically. Ultimately, this resulted in physical, mental, and emotional harm to the falsely accused citizens and the people in their environment.

Second, the recognition of the devastating effects of the harmful systems on citizens by the general public took several years or even decades. Moreover, the cases are similar in the political inaction by parliaments, and the lack of responsibility taken by executive agencies. In the end, three of the four cases resulted in national inquiries. However, the inertia in reacting to the issues in the different public algorithmic systems has amplified the harms already inflicted on citizens.

Finally, the four cases show that the cause for algorithmic harms cannot be explained by only focusing on the algorithmic application (cf. Clarke et al., 2024; Rinta-Kahila et al., 2022). For example, the childcare allowances scandal shows how political framing and strategic behaviour, ill-functioning institutional mechanisms, and the underlying information infrastructure, in combination with the irresponsible deployment of algorithmic applications, resulted in the harms inflicted on citizens. In other words, the emergence of the harms can only be understood by considering the interactions between technical, institutional, and agential components of the algorithmic system.

6.2 Analysis: Kafka in public algorithmic systems

This section analyses the four cases discussed in Section 6.1 from the perspective of Kafkaesque situations and arbitrary use of power. The analysis provides a better understanding of what algorithmic harms are and disentangles the harms from their causes. In general, algorithmic harms are inflicted on citizens who find themselves in Kafkaesque situations that originate from manifestations of arbitrary conduct, see Figure 6.1.

Figure 6.1 Algorithmic harms and their causes

This section is structured following Figure 6.1 in the reversed order. It starts with demarcating the concept of algorithmic harms in Section 6.2.1. Thereafter, Section 6.2.2 explores the Kafkaesque situations that citizens are in when harm is inflicted on them by public algorithmic systems. Finally, Section 6.2.3 discusses how these Kafkaesque situations originate in possibilities for arbitrary conduct in the socio-technical specification of algorithmic systems.

6.2.1 Algorithmic harms inflicted on citizens...

Several studies have initiated the definition of harms that algorithmic applications inflict on individuals. Authors such as Weidinger et al. (2022) and Shelby et al. (2023) have focused on formulating taxonomies of harms. Weidinger et al. (2022) focus on the risks of ethical and social harms in the operation of language models. Their taxonomy mainly emphasises the potential risks related to information content and information sharing in language models. Shelby et al. (2023) have developed a taxonomy of harms that mostly focuses on harms related to identity and inequality. They distinguish representational, allocative, quality-of-service, interpersonal, and societal harms. Their focus on identity and inequality leads them to conclude that most harms are the result of inequities in society and historical biases. Harms related to discrimination and inequality are also notable in cases like the childcare allowances and DUO cases discussed in Section 6.1.

Most harm taxonomies are based on the idea that algorithmic harms are socio-technical by nature (Balayn, 2023; Hutiri et al., 2024; Shelby et al., 2023). The cases in Section 6.1 confirm that algorithmic harms do not only emerge from flaws, hazards, or errors in technical artefacts – the algorithmic application – but can only be understood from the interaction between institutional components, technical components, and human behaviour. Balayn (2023) demonstrates how a technical focus on harms falls short. She examines the limitations of considering harms from a bias and algorithmic fairness view. She argues that harms related to, for example, discrimination, social inequities, or the production of algorithmic systems are not captured within this focus on bias and algorithmic fairness.

A weakness of the aforementioned taxonomies is that they conflate consequences (i.e., harms) and their causes (see also Hutiri et al., 2024). Consequently, they obscure the role or responsibility of designers and/or owners of algorithmic systems in creating harm. Hutiri et al. (2024) address this conflation by basing their taxonomy (for harms related to speech generation) on an explicit conceptual framework. They distinguish harms from their causes as follows:

'AI harms are caused by responsible entities that create or deploy AI, and result in negative outcomes to affected entities. On the one hand, AI harms are a consequence of an affected entity's exposure to AI. This exposure can be of different kinds: harm can arise when an affected entity is the *subject of, interacts with, suffers due to,* or is *excluded from* AI. … On the other hand, AI harms are also a consequence of a responsible entity's [i.e., creator or deployer] *intent* [emphasis in original].' (Hutiri et al., 2024, p. 363)

The conceptual framework underlying Hutiri et al.'s (2024) taxonomy explicitly distinguishes the effects on individuals exposed to algorithmic systems from the causes of these harms. The harms are caused by failures in algorithmic systems that eventually can be traced back to the creation (the focus of this research) or deployment of algorithmic systems. The conceptual framework results in a more detailed understanding of harm, including economic, cultural, and physical harms (Hutiri et al., 2024).

This thesis defines algorithmic harms as a loss suffered by a citizen who is exposed to the output of an algorithmic system. A loss is a consequence of unplanned or undesired events that are unacceptable to citizens and can be of material, physical, emotional, and/or mental nature (cf. Leveson, 2012). Focusing on the citizen as affected individual and governmental organisations as creators (or deployers) of public algorithmic systems situates algorithmic harms in the delicate relationship between citizens and government. This relationship can be characterised by power imbalances, which are likely to be increased by algorithmic systems (Galloway, 2017). Section 6.2.2 will elaborate on how these power imbalances are emerging in Kafkaesque situations. Thereafter, Section 6.2.3 explicates the causes of algorithmic harms. In Section 6.3, we will be coming back to the socio-technical nature of Kafkaesque situations and arbitrary conduct in public algorithmic systems.

6.2.2 ...who find themselves in Kafkaesque situations...

The citizens in the cases discussed in Section 6.1 experience the losses typical of algorithmic harms. Before these harms manifest, these citizens already find themselves cornered in an impenetrable and inscrutable situation. The algorithmic system that they are confronted with is impenetrable because the responsibilities within the system and the means for recourse are unclear. The systems are inscrutable because the means for recourse, if they are available, are frustrated – see, for example, the presumption of right and reliable IT systems, or the rigid interpretation of fraud policy in court rulings. As a result, citizens were cornered. They had nowhere to go; politicians and the media were also not taking up their unjust situations. In sum, the citizens in the four cases were not capable of addressing the incorrectness and injustices in their individual situations. Although appraising the cases in Section 6.1 runs the risk of hindsight bias, the situation of citizens can be perceived as a Kafkaesque situation.

As mentioned before, authors have related algorithmic systems to the work of Franz Kafka (Bayamlıoğlu & Leenes, 2018; De Laat, 2019; Ossewaarde, 2023; Susser, 2016). They refer to Kafkaesque situations in which citizens are caught up in a digital bureaucratic system without knowing how to solve their problems. Thereby, the authors position themselves in a tradition of using the Kafka metaphor in organisation studies. Scholars in this scientific discipline have specified the usefulness of the literary metaphor of Kafka to describe the citizen perspective in bureaucratic, organisational dystopia (S. Clegg et al., 2016; Gratton et al., 2021; Hodson et al., 2013; Huber & Munro, 2014; Munro & Huber, 2012; Warner, 2007; Yang, 2022).

The strength of the Kafka metaphor is its bottom-up perspective on the relationship between government and citizen. It starts from the affected individual (i.e., citizen) rather than the common top-down perspective in the study of public and bureaucratic organisations (Warner, 2007). Clegg et al. (2016) empirically studied what a Kafkaesque situation entails for citizens who interact with bureaucratic systems. This situation starts with the meaninglessness of the bureaucratic system to citizens, because individuals do not have the cognitive capacity to make sense of processes or scrutinise practices (S. Clegg et al., 2016). Individuals feel like they are entangled in a confusing, irrational, and unending labyrinth (cf. Munro & Huber, 2012; Huber & Munro, 2014). Moreover, the bureaucratic system can give individuals the feeling of being impotent. As a result, citizens refrain from taking actions and their agency is reduced (S. Clegg et al., 2016). Eventually, individuals feel helpless or powerless as they lack control over

6

determining the outcomes of bureaucratic procedures (S. Clegg et al., 2016; Munro & Huber, 2012). This empirical description of Kafkaesque situations corresponds with the impenetrability and inscrutability of the algorithmic systems in the childcare allowances scandal, the Robodebt scheme, and the Post Office scandal.

The discipline of organisation studies suggests that Kafkaesque situations emerge from 'organisational dystopia' in which the power imbalance between government and citizen deteriorates (Warner, 2007, p. 131). Organisational dystopia is characterised by an opaque and overly complex system of rules and procedures that hinders individuals in understanding the working of the system or creating an overview of system elements and responsibilities (Munro & Huber, 2012; Warner, 2007). In addition, in an organisational dystopia, reality is expected to conform to the system of rules and procedures, and, therefore, leaves no room to question the system (Munro & Huber, 2012). Finally, in this dystopia, organisations do not acknowledge the idiosyncrasies possible in a system of rules and procedures, such as divergent goals, chaos, patrimonialism, and unwritten rules that can result in conflicts (Hodson et al., 2013).

Like the empirical description of Kafkaesque situations, the characterisation of organisational dystopia can be recognised in the cases in Section 6.1. The algorithmic applications in these cases were central parts of opaque and overly complex systems of rules and procedures that increased the power imbalance between government and citizen. Concentrating on this relationship between citizens and public organisations in more detail, we turn to how arbitrary use of power by public organisations is enabled, mediated, or engendered by algorithmic systems. The concept of arbitrary use of power makes the conduct of public organisations in organisational dystopia explicit.

6.2.3 ...that originate from arbitrary conduct

The previous section suggests that public organisations should reduce possibilities for arbitrary conduct in public algorithmic systems to prevent, mitigate, or correct algorithmic Kafka. However, therefore, these organisations should have an understanding of what exactly counts as arbitrary conduct. Moreover, the manifestations of arbitrary conduct in public algorithmic systems need to be understood. We turn to legal philosophy to understand arbitrary conduct.

Although arbitrary use of power is a central concept in legal philosophy, it is under-theorised (Krygier, 2016). Mak and Taekema (2016), and Krygier (2016) have provided first attempts to categorise and define arbitrary conduct. We base our discussion of arbitrary conduct in public algorithmic systems on four, non-exclusive and non-exhaustive, manifestations of arbitrariness listed by these authors: (1) reasoning by public servants based on their 'own will or pleasure'; (2) inability of citizens to engage in or contest decision-making; (3) unpredictability and incomprehensibility of conduct for those affected; and (4) unfair decision-making in concrete situations. The first two manifestations are focused on the exercise of power, whereas the other two are about the output of those who exercise power (cf. Mak & Taekema, 2016). This section exemplifies how these four manifestations may emerge in public algorithmic systems.

Reasoning based on own will or pleasure

Conduct is arbitrary if it is based on the will or pleasure of an individual exercising power. This means that decisions made lack a rational basis or are not supported by sound arguments (Mak & Taekema, 2016). Such a rational basis is typically based on formal rules and regulations, procedures, or mandates that steer or limit the power exercised by decision-makers (Krygier, 2016). Public algorithmic systems enable two types of actors to impose their own will or pleasure on others: the public servants deploying the system or using its output (i.e., the operator), and the designers of the system (cf. Hutiri et al., 2024).

The system's operator can interact with the system's output based on their own will or pleasure. The DUO, childcare allowances, and Post Office cases show how operators can intentionally use the system to 'rationalise' their decisions based on their own will or pleasure, thereby possibly overriding other rational arguments. Algorithmic systems may also be used selectively to confirm one's own biases (Young et al., 2021). Another issue arises when operators blindly follow algorithms' faulty output, not using their own critical reasoning to prevent an undesirable outcome. This phenomenon is sometimes referred to as automation bias, and hints at the risk of deskilling and loss of operational discretion (Green & Chen, 2019). However, the operational actors are not always to blame (Green, 2022). Often, an error is a function of the environment in which an operator is acting (Leveson, 2012). In other words, operators often lack the right information or resources to oversee or work with an algorithmic system. In such cases, discretion of the system's operator at the operational level is effectively reduced or distorted by design choices made earlier in the system's life cycle (cf. Leveson, 2012; Peeters & Schuilenburg, 2018; Zouridis et al., 2020).

The design choices that form the basis of algorithmic systems can also be based on the designers' own will or pleasure (König & Wenzelburger, 2021). In the Robodebt case, Whelan (2020) concludes that the debacle was a consequence of bad design, whereas Whiteford (2021) argues that design choices were intentionally going against standing social welfare policies. Possibilities for arbitrary use of power by designers are increasing as the use of algorithmic systems in public administration is shifting power dynamics in public organisations, for example, by shifting discretion from street-level bureaucrats to system designers (Bovens & Zouridis, 2002; Zouridis et al., 2020), or by strengthening the relative position of executive branches in governments (Passchier, 2020). Hence, assigned responsibilities may not reflect the real influence of actors on the algorithmic system. A mounting problem in this context is the handover of the development of algorithmic systems – that intimately mediate public services – to external parties. Thereby, both public accountability as well as autonomy over the quality of the public services are shifted from public to private actors (Whittaker, 2021).

No space or means to engage in or contest decision-making

An important feature of exercising power non-arbitrarily is that the interests of individuals affected by decision-making are considered (Mak & Taekema, 2016). This can be guaranteed by providing affected individuals with the means to control and question those in power (Krygier, 2016). Power is not arbitrary when citizens are able to engage in decision-making through possibilities for voicing arguments, contesting decisions, and raising complaints. An important element here is that engaging and contesting individuals are heard by those in power (Krygier, 2016).

Considering the cases in Section 6.1, the engagement of citizens in making design choices for the algorithmic systems is not fully clear. Notwithstanding, the descriptions suggest that the design and development of the systems mostly took place outside the public eye. Cases like SyRI in the Netherlands and AMS in Austria show how a public debate can have an important role in scrutinising public algorithmic systems (Allhutter et al., 2020; Grill et al., 2023; Wieringa, 2023).

Contrarily, the approach towards contestation is much clearer in the analysed cases. In all cases, possibilities to contest were unavailable or frustrated. This was mostly a matter of institutional design. Robodebt victims were not informed about possibilities for appeal, Dutch courts adopted an unlawful interpretation of anti-fraud law, and the Post Office was protected by the presumption of reliability on IT systems. Students in the DUO case often won court cases, but these individual cases did not result in an evaluation of the entire system.

In general, citizens are unable to engage in design processes of or contest algorithmic systems when the responsibilities for such a system are poorly specified. Algorithmic applications are known for the responsibility gaps they create (Santoni De Sio & Mecacci, 2021). For example, algorithmic systems are dependent on datasets and information architectures that might be situated in other organisational units (Sculley et al., 2015), see the childcare allowances scandal. In fact, algorithmic systems often rely on vast and potentially global supply chains (Cobbe et al., 2023), with inherent complexities that contribute to developers or users not experiencing or taking responsibility (Widder & Nafus, 2023).

Unpredictable and incomprehensible conduct

For conduct not to be arbitrary, citizens should be able to comprehend the relevant rules. After all, citizens cannot comply with rules they do not understand (Krygier, 2016). This also means that rules and their enforcement should be predictable (Mak & Taekema, 2016). The requirements of predictability and comprehensibility also support citizens in challenging arbitrary conduct (see previous manifestation).

The opaqueness and complexity inherent to algorithmic systems impedes comprehensibility and predictability of power exercised through public algorithmic systems (Burrell, 2016). For example, actors lack the expertise to understand the working of algorithmic systems (De Bruijn et al., 2022). Complexity mostly created a problem in the cases presented in Section 6.1 because public organisations lacked an overview of their own algorithmic systems. They did not fully understand the technical and institutional functioning of the system. At the same time, complexity also obscures the consequences of a system beyond the output of an algorithmic application. Software-based automation systems can become so complex that operators also run into the limits of cognitive capacity to properly understand how the system functions and what behaviours might emerge under particular circumstances (Leveson, 2012). For example, semi-automated systems may look like the only effective and efficient way to compute and administer eligibility for and the height of social welfare policies. Similarly, algorithmic applications may emerge to address the lack of predictability of complex social welfare systems. However, such semi-automated systems may quickly add their own complexity, or their behaviour may turn unpredictable.

Unfair decision-making in concrete situations

Finally, Mak & Taekema (2016) stress the fact that unfair decisions can also be arbitrary. To prevent or reduce arbitrariness, decision-makers need to make contextual assessments of concrete situations that will be affected by their decisions. This is related to considering the voices and interests of citizens in exercising power (Krygier, 2016).

Algorithmic systems can contribute to unfair decisions in concrete cases (Barocas & Selbst, 2016; Dobbe et al., 2018). First of all, discrimination and biases – especially prevalent in data-driven algorithmic applications (Hildebrandt, 2019) – may result in unfair decision-making. For example, when algorithmic systems are used for the allocation of benefits, biases may lead to allocation to citizens based on irrelevant characteristics. These biases may find their roots in historical conduct, as well as in the design choices made by developers, or in the ways in which the algorithmic application is used (Dobbe et al., 2018). The childcare allowances and DUO cases are exemplary as the involved algorithmic systems were deemed discriminatory.

In addition, errors, flaws, and the statistical or correlational nature of decision-making in algorithmic systems can exclude citizens. As discussed before, the competences of algorithmic applications are often overstated (Elish & Boyd, 2018; Suchman, 2023). The Post Office and the Robodebt cases show how mistakes and bugs can be a significant cause of false accusations against citizens. Related to this is the diminished discretion of front-line or case workers. These actors are often considered to be a human-in-the-loop that filters out incorrect or unjust decisions. At the same time, these street-level bureaucrats see their discretion curtailed (Zouridis et al., 2020), and their agency limited (Peeters, 2020) since algorithmic decision-making based on code is less flexible compared to the deliberative practice of a legal system based on speech and written word (Hildebrandt, 2019). Consequently, the front-line or case worker is unable to correct unfair outcomes in concrete situations.

6.3 Arbitrary conduct and the design process

Algorithmic Kafka resulting from the use of public algorithmic systems can be traced back to the arbitrary use of power by governments towards citizens. Such arbitrary conduct is mediated, enabled, or engendered by a particular constitution of the socio-technical specification of the public algorithmic systems. Where the actual use of algorithmic systems by public organisations eventually results in manifestations of arbitrary conduct, designers have an important role in realising possibilities of arbitrary conduct in public algorithmic systems through the design choices that these designers make. As such, possibilities for arbitrary conduct can partly be traced back to design activities and interactions between designers.

This section elaborates on two ways in which possibilities for arbitrary conduct can be related to design activities and interactions between designers. First, designers can use their own power arbitrarily in formulating the socio-technical specification, undermining the legitimacy of the public algorithmic system (Section 6.3.1). Second, a lack of coordination between designers from different disciplines can result in those designers creating possibilities for arbitrary conduct in the algorithmic system (Section 6.3.2). Both the lack of legitimacy of design choices as well as the lack of coordination between designers are related to the two action situations identified in Chapter 4. This section will elaborate on the two causes of arbitrary conduct possibilities in public algorithmic systems.

6.3.1 Arbitrary use of designerly power

Designers of public algorithmic systems are confronted with all sorts of value conflicts in making design choices (cf. C. W. Clegg, 2000). According to Bannister (2017), all possible public values should be considered when designing digital technologies for governmental contexts. That does not mean that all values that are deemed important for a specific public algorithmic system can be realised. After all, values can be in conflict with each other. As a result of conflicting values, designers are confronted with hard choices when formulating the socio-technical specification of algorithmic systems (Dobbe et al., 2021). Alternatives in the specification can represent different values and, at the same time, may be contrary to other values. A hard choice emerges when such alternatives are on a par (Chang, 2017).

The value conflicts and resulting hard design choices bring political struggle into the design process of public algorithmic systems. Following from the multi-actor setting of socio-technical design processes (cf. De Bruijn & Herder, 2009), multiple actors will have, would like to have, or should have a say in making design choices. These actors have different perceptions of the public algorithmic system, emphasise specific elements in the system, and represent different interests, values, and viewpoints in the system and/or in the design process. In addition, the effects of a specific (to be designed) public algorithmic system on values are often obscure or uncertain. For example, the emergence of new algorithmic applications, the strategic behaviour of stakeholders (De Bruijn & Herder, 2009), unclear system boundaries, and an incomplete overview of all effects of design choices (Sclove, 1995) are all degrees of uncertainty that designers have to deal with.

In general, designer have a better information position about these uncertainties compared to other actors in the system, and their official position provides them with more direct influence over decision-making (Bratteteig & Wagner, 2016). This position grants designers with *designerly power* in the political struggle that the design process is. Designerly power relates to the influence that designers have on the constitution of public algorithmic systems due to their role in making design choices. The coordinating role of technical designer in Chapter 4 is an example of designerly power.

Such designerly power can become problematic when designers dominate other actors in the design process. This happens in the technocratic approach towards designing. The political choices are shifted towards experts with an unclear political or democratic mandate (Beck, 1992; Hajer, 1995). The lack of countervailing power or mechanisms to provide democratic legitimacy can result in arbitrary use of designerly power in making design choices. In this case, designers would base their design choices on own will or pleasure, instead of basing the choices on democratic deliberations or political mandates (cf. Grimmelikhuijsen & Meijer, 2022; Hildebrandt, 2011; Zouridis et al., 2020). At the same time, possibilities for citizens to engage in making or questioning design choices are cut off.

Notwithstanding, it is not only the abuse of power by official designers that can result in problematic socio-technical specifications. Debates on design choices by politicians can result in negotiated nonsense (Herder & De Bruijn, 2008) or politicians can refrain from making normative choices (Dobbe et al., 2021, p. 4). As a result, official designers are left in the dark when they need to formulate the socio-technical specification. They either need to scrutinise the negotiated output or use their own interpretation of the normative choice. This provides room for designing possibilities for arbitrary conduct into public algorithmic systems.

6.3.2 Arbitrariness following from deficient coordination

Apart from arbitrary use of designerly power, designers can design possibilities of arbitrary conduct into a public algorithmic system. In that case, the socio-technical specification formulated by designers leaves room for arbitrary conduct to emerge. The cases in Section 6.1 show that the misalignment between different socio-technical components in algorithmic systems creates these possibilities for arbitrariness. Alignment in a socio-technical specification is achieved when the institutional and technical artefacts in the systems to be designed are compatible on different levels of aggregation (Koppenjan & Groenewegen, 2005; Künneke et al., 2021). The need for alignment follows from the interactions between system components and the interconnectedness of design choices regarding each individual system component (C. W. Clegg, 2000). Misalignment of institutional and technical components will render the interactions between the components problematic. For example, objection procedures are often put forward as a means to address (unanticipated) errors and mistakes in socio-technical systems. However, in the Post Office case, the presumption in law that software and ICT technologies are always right and reliable does not align with the actual character of those technologies. The presumption eliminates all possibilities for citizens to find recourse when digital technology is involved. The DUO case shows how objection procedures can be effective in reversing faulty decisions made in a public algorithmic system. At the same time, the emphasis of objection procedures on individual cases does not align with the systemic impact of algorithmic applications.

To establish alignment between system components, the different relevant disciplines for the algorithmic system should collaborate and learn from each other. The coordination of designing different system components requires 1) a shared vocabulary (cf. Krafft et al., 2020) about the interactions between socio-technical components – designers do not necessarily need a full understanding of each other's expertise, 2) involving different perspectives on algorithmic systems to stimulate reflexivity on design output (Elish & Boyd, 2018), and 3) the evaluation of the public algorithmic system as a whole. The latter requires a diagnostic approach in which designers anticipate and react to emerging possibilities for arbitrary conduct in public algorithmic systems in the interactions between components (cf. Dobbe et al., 2021).

However, as identified in Chapter 4, current design practices of public algorithmic systems are characterised by inefficient and ineffective coordination between designers. The lack of coordination between different disciplines can result in misaligned socio-technical specifications. A special situation in this respect is the case where rules and regulations are not appropriately translated to the logical inner workings of an algorithmic system. This can be the result of, for example, vagueness in the specification or the urge to simplify inherently complex situations in formalistic models (Dobbe et al., 2021; Alkhatib, 2021). Thereby, system designers may impose a system's logic that does not align with laws, regulations, and policies related to the processes in which the algorithmic system functions. This relates to the first manifestation of arbitrary conduct – reasoning based on own will or pleasure. These problems may not be new for public administration, and the role of the algorithmic application may not be decisive, but designers need to understand the role of algorithmic applications in creating Kafkaesque situations in public administration. If not, public organisations, representing the designers, are responsible for unfair decision-making in concrete situations by faulty public algorithmic systems.

6.4 Conclusion

Where Chapters 4 and 5 considered the process aspects of designing public algorithmic systems, this chapter focused on the output of interactions between designers: the socio-technical specification of public algorithmic systems. This chapter aimed to better understand harmful algorithmic systems and the role that designers play in the emergence of such systems. More specifically, it examined how and why Kafkaesque situations are caused by particular socio-technical specifications of algorithmic systems formulated in design activities.

This concluding section relates the findings to the design practices identified in Chapters 4 and 5.

When citizens affected by public algorithmic systems are in a Kafkaesque situation, they are cornered in inscrutable and impenetrable public algorithmic systems that make consequential decisions. Kafkaesque situations find their origin in possibilities for arbitrary use of power in the socio-technical specification of the algorithmic system that might materialise when the system is deployed in its context. Through analysing documented cases of harmful public algorithmic systems, this chapter identified manifestations of arbitrary conduct that can be related to algorithmic Kafka. Designers should be aware of the possibilities of arbitrary conduct manifestations (reasoning based on own will or pleasure, no means or space to engage or contest, incomprehensible and unpredictable, and unfairness of decision-making in concrete situations) within the socio-technical specification they formulate. In doing this, designers should mostly focus on such possibilities in the interactions between institutional and technical components.

Possibilities for arbitrary conduct partly originate in the interactions between designers. This chapter shows that manifestations of arbitrary conduct can be introduced in the socio-technical specification of public algorithmic systems if legitimacy and coordination in the design process are deficient. Both the lack of legitimacy and of coordination correspond with the two action situations identified in Part I. Coordination between disciplinary designers is currently lacking in design practices of public algorithmic systems. Similarly, the distance between political and official designers indicates a lack of legitimacy in current design practices.

In sum, when aiming to prevent, mitigate, or correct algorithmic Kafka, interactions between designers of public algorithmic systems have to be reshaped. More specifically, arbitrary conduct should be reduced in both legitimating design choices and coordinating design activities. Democracy and the Rule of Law provide fundamental mechanisms to reduce arbitrary use of power. The next chapter examines how the presuppositions of democracy and the Rule of Law can be ingrained in the socio-technical design practices of public algorithmic systems.

Chapter 7

Socio-technical design practices in a democratic and Rule of Law context

> Normen, protocollen, en modellen waarbinnen de uitvoerders hun werk moeten doen zijn hulpmiddelen. Niet minder maar ook niet meer. Dat geldt (dus) ook voor de toepassing van ICT in publieke dienstverlening. De (interne) voordelen daarvan worden tot nu toe overgewaardeerd en de (externe) effecten voor burgers schromelijk onderschat. Juist burgers in een zwakke positie worden daar als eersten het slachtoffer van.[1]
>
> – *Herman Tjeenk Willink; Groter denken, kleiner doen, p.73*

Chapter 6 elaborated on the possibilities for arbitrary conduct in public algorithmic systems and related the emergence of these possibilities to the design process. As discussed in Section 2.5, arbitrary use of power can be addressed through democracy and the Rule of Law. Democracy provides ways to legitimise the use of power, and the Rule of Law protects citizens against misuse of that power. Both democracy and the Rule of Law form the basis for the organisation of states in liberal democracies. Accordingly, they can also form the basis for design practices within public organisations. Part I identified presuppositions behind current design practices of public algorithmic systems that arguably are contrary to democracy. This chapter studies the implications of grounding design practices for public algorithmic systems on presuppositions of socio-technical designing, democracy, and the Rule of Law. It answers the following question:

How to synthesise socio-technical designing, democracy, and the Rule of Law into a theory of designing public algorithmic systems?

The question is answered through abductive reasoning aimed at synthesis (see Section 3.3.2). First, Section 7.1 juxtaposes socio-technical designing, democracy, and the Rule of Law. The juxtaposition focuses on identifying symbioses, shared challenges, and contradictions between

1 Norms, protocols, and models that structure the work practices of public servants are means to an end. Nothing less but also nothing more. The same goes for using ICT in public services. The (internal) advantages of these tools are currently overrated and the (external) effects for citizens are seriously underrated. Especially citizens who are in vulnerable positions are the first to be victims of this.

the three presuppositions. The findings of Section 7.1 are synthesised with all learnings from previous chapters into a meta-theory. Section 7.2 discusses this meta-theory which comprises the assumptions that underlie the design theory formulated in Chapter 8. Section 7.3 concludes this chapter by presenting the desired interactions between designers that follow from the meta-theory. These desired interactions form the starting point for the design theory.

7.1 Juxtaposing democracy, the Rule of Law, and socio-technical designing

As discussed in Section 3.3.2, we understand juxtaposition as comparing and contrasting dissimilar concepts. The dissimilar concepts we juxtapose in this section are the presuppositions of socio-technical designing, democracy, and the Rule of Law. We provide a short overview of our interpretation of these concepts (as presented in Section 2.3 and 2.5).

We identified five characteristics of socio-technical designing:

- Systemic: consecutively and consistently design system components and their interactions
- Emergent: anticipation and reaction to system consequences in design processes through co-evolution and learning processes
- Contingent: technological and institutional structures determine the design context
- Multi-stakeholder: approach design as a social process of interactions between heterogeneous actors involved or affected by the system to be designed
- Political: value trade-offs make design a political process characterized by power struggle

For the concept of democracy, we formulated our interpretation of the three fundamental tenets of democratic theory in line with our definition of design:

- The public has primacy in making design choices and the public is formed by those individuals (possibly) affected by the consequences of design choices
- A division of labour based on representative democracy: the public frames problems and sets goals which are translated into concrete design choices by representatives
- Self-correction is guaranteed by approaching design as an inquiry, i.e., looking for design alternatives by questioning established orders

Our interpretation of the Rule of Law follows the socio-legal perspective that goes beyond a focus on (formal) rules to reduce arbitrary use of power:

- Establish argumentative practices in which design choices can be contested; legitimacy of design choices depends on the soundness of arguments
- Introduce interventions (not necessarily formal rules) that ensure a balance of power between involved and affected actors

Despite their dissimilarities, the three presuppositions have intersecting characteristics. In general, the presuppositions share a focus on pluralism, inquiry, and co-evolution. Using these intersections as starting point, we elaborated the juxtaposition of the three presuppositions.

This resulted in four symbioses (Section 7.1.1), three shared challenges (Section 7.1.2), and three contradictions (Section 7.1.3) between the presuppositions (see Section 3.3.2 for the method used to arrive at these juxtapositions).

7.1.1 Symbioses

Despite the heterogeneity of the presuppositions of socio-technical designing, democracy, and the Rule of Law, they share fundamental elements. In this section, we identify symbioses between the three presuppositions – i.e., when characteristics of one presupposition are shared and compatible with other presuppositions and, therefore, reinforce the function of the different presuppositions. We identified four symbioses: iterative response, reflexive checks and balances, scaffolding political struggle, and pluralistic argumentation. The four symbioses are discussed as follows. The first three paragraphs of each symbiosis discuss the characteristics of one of the three presuppositions. The fourth paragraph discusses the symbiosis related to a characteristic in socio-technical designing, democracy, and the Rule of Law.

Symbiosis 1: iterative response

Iteration is fundamental in all three concepts. In response to the phenomena of emergence and co-evolution (see Section 2.3.3), iteration is a default practice in socio-technical designing (Bauer & Herder, 2009). To some extent, designers and organisations should submit to the uncertainty inherent to designing socio-technical systems and acknowledge the idiosyncrasy in the design process. Recurrently revisiting a socio-technical specification enables designers to react to new situations, that partly originate from their own design choices. Iteration can be done through structural feedback mechanisms of anticipation and reaction. In this, anticipation is about envisioning or predicting consequences through testing or social experimentation (cf. Van de Poel, 2020). Reaction focuses on monitoring the system, and empowering involved and affected actors to raise issues in public algorithmic systems (cf. Dobbe et al., 2021).

For both the democratic as well as the Rule of Law presuppositions, the practice of iteration follows from their emphasis on contestation. Democratic decision-making is constantly in flux; the public and their representatives are always entitled to challenge decisions, for example, because new situations emerge or when they learn from past situations. This relates to the self-corrective tenet of democracy that stipulates that the public retrospectively considers earlier interventions in society and responds to problems that arise from these interventions (Olsen, 2009). Moreover, it is associated with flexibility and reflexivity within democratic institutions, and the possibility to contest decision-making (Dewey, 1927/2016; Spicer, 2019). The value of self-correction lies in facilitating and promoting continuous learning. Reflecting and reacting on the consequences of earlier decision-making provides the institutional flexibility to respond to uncertainty and complexity in society (Olsen, 1997). Dewey's (1927/2016) interpretation of democracy incorporates iteration by considering democracy as an experimental, curious and reflective approach to finding the right actions to address problems.

This self-corrective element manifests in Rule of Law procedures that institutionalise contestation – a central aim in the Rule of Law (cf. Hildebrandt, 2018). The Rule of Law engenders iterative practices by establishing countervailing powers between involved actors so that design choices can be challenged. Moreover, mechanisms, such as, objection procedures, the right to protest, and the role of an independent judiciary provide citizens with means to find recourse for unfair or unlawful decisions (Waldron, 2011).

Fundamental to all three concepts is the iterative approach to deal with emergent properties of practices and systems. Iteration is a way to deal with changing circumstances and situations, but also to reflect on and address unanticipated or undesirable consequences of own actions. Socio-technical designing, democracy, as well as the Rule of Law emphasise that not all outcomes of complex systems can be controlled, understood, or predicted. Accordingly, they prescribe iteration as a way to anticipate and react to emergence. The first symbiosis of the three presuppositions is found in the need for iterative response to new situations that emerge while designing.

Symbiosis 2: reflexive checks and balances

While a design process needs flexibility to iteratively react to unexpected or undesirable situations that emerge, institutionalising a design process, and decision-making processes in general, is also a way to attain control over the design object – i.e., the public algorithmic system and possible Kafkaesque situations emerging in that system. Socio-technical designing encompasses intended activities to change the form and function of socio-technical systems. As such, the socio-technical design approach can be interpreted as a process to control the function and form of to be designed artefacts. Nevertheless, control over design processes is always limited because of the uncertainty caused by co-evolution of problem and solution, and because of the conflicting interests and power differences associated with multi-actor settings. In order to ensure the desired level of control over the design output, a reflective and responsive attitude at the side of designers is conditional to the feedback mechanisms discussed in the first symbiosis (Dorst, 2019b).

The self-governance tenet in democracy rejects the phenomenon of power exercised by a few over other individuals. The public has primacy in steering society and, consequently, has final control over means of steering. Governments should respect citizens' freedom, including illegitimate restriction of that freedom through government action. This also includes guaranteeing individual and collective interests through deciding the goals of a policy or system (Christiano, 1996). In line with Dewey, the very fact that public algorithmic systems can inflict harm on citizens creates a public – i.e., the association of those that can potentially be harmed by algorithmic systems. Democratic theory provides several options to institutionalise the prevention of power accumulated at a few people over this public (Olsen, 1997). The pragmatic approach to democracy that best aligns with the nature of socio-technical design processes is that of the choice of aims model (Christiano, 1996; Dewey, 1927/2016). Similar to what Sclove (1995) proposes for bringing democracy into design processes, the choice of aims model assigns the responsibility to prescribe goals and constraints for a public algorithmic system to be designed to a democratic body that represents the public. In this case, the democratic body must be empowered to steer the goals and constraints through information provision by experts.

As discussed in Section 2.5.2, the main aim of the Rule of Law is to curb arbitrary use of power. The aim to reduce arbitrary conduct introduces the balance of power as a guiding principle for positioning actors in decision-making, since 'concentration of powers, regardless of the institution concerned, engenders opportunities for arbitrariness' (Mak & Taekema, 2016, p. 29). This balance of power can be approached by establishing countervailing powers, ensuring a separation of power, or organising checks and balances.

Socio-technical designing, democracy, and the Rule of Law structure decision-making processes in such a way that the outcomes are adequate, legitimate, and/or just. A commonality in their approach in structuring processes is the distribution of power. Democracy assigns final decision-making power to the public, socio-technical designing decentralises decision-making power to all involved or affected actors, and the Rule of Law provides a balance of power between decision-making actors. The three presuppositions share the idea of distributing power among all involved and affected actors in design processes. Distribution of power can be done by institutionalising checks and balances. Still, to emphasise the primacy of the public in the distribution of power, we refer to reflexive checks and balances. All actors with decision-making power should be reflexive to the needs and interests of the public.

Symbiosis 3: scaffolding political struggle
As discussed throughout this thesis, design choices are the corollary of value trade-offs and therefore bring political struggle. Our empirical studies show that the normativity and complexity inherent to socio-technical designing may result in inertia, passivity or simplification when designers are confronted with the politics in design processes. Chapters 4 and 5 observed this phenomenon in current design processes. Public organisations fall back to the practices that they know such as traditional engineering processes, or fashionable approaches like agile. Following Koppenjan & Groenewegen (2005), it is important to first delineate the design process before starting to design technological and institutional artefacts for the socio-technical system. This design process should provide procedural structure to the political struggle that is inherent to designing (Clegg, 2000).

Democracy is a way to channel political struggle while preserving self-government of citizens and fair decision-making in which all citizens can participate. Following Dewey, we distinguish two structuring principles for political struggle in the design process: representation, and questioning of design choices. First, Dewey (1927/2016) argues for a representative form of democracy. He acknowledges the value of expertise (e.g., needed for inquiry into design choices) but dismisses elitism. He argues that a public is only formed when it has officials that translate the needs, interests, and problems of the public and act on those. Officials execute solutions or measures within the frame of problems and goals set by the public. Moreover, the public makes trade-offs between alternatives provided by these officials. In other words, experts are important in the division of labour in democratic practices but are constrained by boundaries set by the public. Second, in line with the idea of democracy as inquiry, Dewey (1927/2016) considers democracy as a way of questioning established orders. Accordingly, citizens should be able to question design choices (to be) made.

The Rule of Law, although itself not immune to political struggle (Krygier, 2007), provides ways to address the detriments of political struggle. Political struggle provides opportunities for arbitrariness. For example, Flyvbjerg (1998), and Hajer (1995, 2003) describe how the power of making design choices easily shifts to unchosen or strategically behaving actors when design practices are manifested. This power drift can be negated through Rule of Law practices. These practices can be, for example, the checks and balances discussed before, the protection of individual rights, or the establishment of an independent judicial system (Mak & Taekema, 2016).

In all three presuppositions, political struggle is a given. Value trade-offs are unavoidable when decisions regarding the organisation of society, the state, or systems are to be made. Still, political struggle about these trade-offs can be displaced, unproductive, or harmful. Therefore, socio-technical designing, democracy, and the Rule of Law prescribe procedures to arrive at civil, fair, and resourceful political debates. In a way, the concepts provide scaffolds for political struggle in decision-making that also apply to design process of public algorithmic systems.

Symbiosis 4: pluralistic argumentation

Socio-technical designing, democracy, and the Rule of Law all emphasise the importance of collecting arguments from different perspectives. The multi-actor nature of socio-technical designing can be mobilised to elicit knowledge, expertise, and interests from all involved and affected actors in the system (to be designed). Literature on socio-technical designing emphasises the need to appreciate and cherish the different perspectives, frames, tools, and practices among disciplines relevant to the system. This will advance the quality of the system. At the same time, it is laborious and troublesome, and maybe even unfeasible, to fully synthesise the input of all applicable disciplines into one coherent output (De Bruijn & Herder, 2009). Contradictions, misalignments, and conflicts are likely to arise at some point. A more effective approach might be to be prepared to meaningfully resolve frictions between disciplines.

Democracy emphasises the need for a meaningful public debate. In such a public debate, claims and/or proposed actions are compared to interests in order to improve decisions (Dewey, 1927/2016). Accordingly, bringing democratic politics to design practices is about giving the public the opportunity to question design choices. At the same time, it brings conflicts and diversity of arguments into view. It is important to pay equal respect to all members of the public in such a debate (Christiano, 1996). For Dewey (1927/2016) this means that all citizens should have equal opportunity to influence decision-making through, for example, education, providing access to the same information or knowledge, or sustaining a meaningful and fruitful public debate. However, such empowerment of citizens is easier said than done (Flyvbjerg, 1998).

Within the Rule of Law, argumentation is a means to bring rationality into decision-making and relates to the deliberative basis of designing. It prescribes that argumentative practices should be established in the design process of public algorithmic systems or in seeking recourse for (semi-)automated decisions. This also requires the empowerment of citizens to play a meaningful role in argumentative practices. These argumentative practices are structured by principles of good governance. In the Dutch context, much has been written about the application of these principles in Dutch law (i.e., abbb's[2]) to public algorithmic systems (Meijer et al., 2019; Meuwese et al., 2024; Widlak et al., 2020; Wolswinkel, 2020). Wolswinkel (2020) argues that two principles are especially relevant for algorithmic practices: motivation (*motiveringsbeginsel*) and meticulousness (*zorgvuldigheidsbeginsel*). Following the motivation principle, governmental organisations are expected to provide well-grounded argumentation that supports their decision-making. The meticulousness principle prescribes that all interests of involved or affected actors should be considered and that proportionality is the main aim when making trade-offs. These principles can be helpful in steering argumentation behind design choices for public algorithmic systems.

The three presuppositions consider argumentation as an important means to gather and share information between actors. Argumentation brings rationality into the design process,

2 *Algemene beginselen van behoorlijk bestuur* (general principles of good administration) are fundamental to Dutch administrative law. They structure the behaviour of public servants.

results in the cumulation of knowledge, and is key when actors with different perspectives and interests deliberate to arrive at some sort of consensus on design choices. To ensure the quality of such deliberation, pluralism of actors and perspectives is a condition for arriving at a, as comprehensive as possible, representation of the addressed problem and its potential solutions. Accordingly, we refer to the fourth symbiosis as pluralistic argumentation.

7.1.2 Shared challenges

The notions of socio-technical designing, democracy, and the Rule of Law as discussed in Chapter 2 are mostly theoretical. However, this thesis aims to bring these concepts to practice. This section discusses the challenges of translating the theoretical concepts into practice. We focussed on challenges that are shared by all three presuppositions. Accordingly, these shared challenges are important considerations to address when restructuring design practices of public algorithmic systems.

Shared challenge 1: contingent on context
Socio-technical design practices are not clearly demarcated or readily transferable to other organisational and social contexts. Design practices are *contingent on context* because they are socially shaped by a specific constellation of actors. The experiences, interests, and perceptions shape the subsequent practice and, therefore, are probably unfit for contexts with a different set of actors. Moreover, design practices in an organisation are shaped by the historical evolution of institutions in that specific organisation. The same goes for democratic and Rule of Law practices. These practices should be tuned to the context in which they are situated. Formulating a template practice that works for all organisational and social contexts is unattainable. For example, because different jurisdictions have their own (historically grown) institutions for democracy and the Rule of Law. Although institutions that shape design practices are always context-dependent, fundamental building blocks can be identified that can inform or inspire the practices in each specific organisation (also see Ostrom, 2005).

Shared challenge 2: favourable attitude
The literature on democracy and the socio-legal perspective on Rule of Law share the understanding that only using formal, legalistic, and top-down institutions to achieve desired practices is not enough. The involved and affected actors should exhibit behaviour conducive to the fundamental premises of democracy and the Rule of Law. In other words, these actors should support the institutions that are formalised and act in accordance with the institutions (also see Chapter 5). Such a *favourable attitude* towards democracy and the Rule of Law can be achieved through the education of the public (see Dewey, 1927/2016) and/or leadership in institutional change within public organisations (see Selznick, 1984). Considering socio-technical systems, this aspect of a favourable attitude is emphasised in subdisciplines such as system safety (Leveson, 2012). This discipline refers to the need for a safety culture within organisations – i.e., the need to ensure that people who are critical in safeguarding a system feel safe themselves to raise possible safety issues, without the fear of being retaliated against or other suffering from it, and that active follow-up of raised issues is organised (Dekker, 2012). Such a favourable attitude among involved actors cannot be designed or controlled but emerges from human behaviour in practice.

Shared challenge 3: steering polycentric decision-making

The final shared challenge relates to polycentricity (see Section 2.4.1). *Polycentricity* refers to the phenomenon of decision-making being situated in several different but interrelated arenas. There is no central position of power in the network of action situations that makes all final decisions (Aligică & Tarko, 2012). This distribution of power in polycentric networks can present challenges for institutionalising socio-technical designing, democracy, and the Rule of Law in the design process of public algorithmic systems. As discussed before, socio-technical designing happens in multiple action situations. As such, the design process is uncontrollable (Kroes et al., 2006). Despite decentralised conceptions of democracy, the core concept puts final decision-making power in the hands of the public. However, the only way to bring this to practice, is to create a so-called division of labour in which experts support the public (Dewey, 1927/2016). In the polycentric web of decision-making arenas, power can easily drift to experts who lack a democratic mandate (Hajer, 1995). Although the Rule of Law can be associated with the centralised division of power within bureaucracy, the notion of checks and balances makes polycentricity a strength. It assigns clear boundaries to the power of specific decision-making arenas, and it establishes control mechanisms between these arenas. Again, like in democracy, these practices are not set in stone. In sum, all three concepts are confronted with the complexity that polycentricity brings.

7.1.3 Contradictions

Notwithstanding the symbioses and shared challenges, we are still juxtaposing quite dissimilar concepts. Therefore, some of the characteristics within the presuppositions are conflicted. This section identifies three contradictions between socio-technical designing, democracy, and the Rule of Law. When using the three presuppositions to shape design processes of public algorithmic systems, trade-offs posed by these contradictions need to be addressed.

Contradiction 1: public versus expert

Socio-technical designing and democracy especially contradict in their emphasis on the expert and the public respectively. Socio-technical designing stipulates that experts from relevant disciplines should be involved in the design process (Baxter & Sommerville, 2011; De Bruijn & Herder, 2009). As such the comprehensiveness and quality of the design output can be ensured (Clegg, 2000). Although the literature on socio-technical designing emphasises the political nature of making design choices (Clegg, 2000), it still prescribes expert-led design processes in which these expert designer will involve other involved or affected actors such as citizens. In democracy, the citizen (as part of the public) has primacy in decision-making (either through representation or direct engagement). Following Dewey's (1927/2016) conception of democracy, the public can be represented by officials that design measures and solutions in society within the boundaries set by that same public. Following De Bruijn et al. (2002), a specific chronology can ensure the primacy of the public: negotiation of design choices always precedes expert scrutiny. This negotiation can take place in public debates. But as discussed in the polycentric challenge, since the public often lacks the knowledge or awareness of design process, they are often unable to meaningfully engage in public debates on design boundaries or goals. Another way of ensuring the primacy of the public is by strengthening the position of politicians that represent the public in the design process. Rules and procedures can clearly demarcate the latitude of experts in the design process.

Contradiction 2: flexibility versus (fixed) procedures
The conflict between flexibility and fixed procedures mostly occurs between socio-technical designing (and to a lesser extent democracy) on one side and the Rule of Law on the other side (cf. Meijer & Thaens, 2021). Designing is a creative exercise that is often based on tacit knowledge, and that follows an idiosyncratic course of design activities (Bots, 2007; Dorst, 2019a; Simon, 1969/1996). This asks for flexibility on the part of designers to react to new insights or new situations emerging during the design process. The need for flexibility contradicts the Rule of Law's emphasis on stable institutions and fixed procedures. This emphasis follows from the rationale that government actions should be predictable and that arbitrary conduct should be reduced. Public organisations, in their role as designer, need to find a balance between being flexible to react to idiosyncrasy in designing, and formally institutionalising public control over design processes.

Contradiction 3: majority decision versus protection of minorities
A classic tension in democratic theory is that of majority decision opposed to the protection of minorities. In the case that majority decision is considered the best possible option to reflect the will of the people, the possibility that the majority tyrannises minorities emerges. Traditionally, minorities are protected through guaranteeing individual human rights and a neutral position of government towards ideas about a good life (Cunningham, 2002). The latter two are in the realm of the Rule of Law (Dworkin, 1980). However, there is still heated debate whether constraining majority decisions through the Rule of Law can be reconciled with the primacy of the public in democracy (Waldron, 2016). A pragmatic approach would be to adapt the socio-technical design process to the way that democracy and the Rule of Law (and the balance between the two) are institutionalised in the particular jurisdiction in which the design process is situated.

7.2 Synthesis: Meta-theory

Section 7.1 showed how the concepts of socio-technical designing, democracy, and the Rule of Law can be related to each other. This section synthesises these insights with the findings in Chapter 6 into a meta-theory. Essentially, the meta-theory summarises what has been discussed in previous chapters. Accordingly, it synthesises our perspective on socio-technical designing in the context of public organisations and their democratic and Rule of Law practices.

The concept of meta-theory for design theory was introduced by Love (2000). A meta-theory explicates assumptions that underlie a specific design theory. The meta-theory in this section clarifies the context, fundaments, and interpretations behind the design principles for institutional interventions formulated in our design theory. Love (2000) describes ten abstraction levels on which assumptions can be described, see Table 8.1. Together, these ten levels provide the philosophical underpinnings of the design theory, the perspective on the design process, and the perspective on the design object.

Table 7.1 Framework for meta-theory by Love (2000, pp. 305–306)

Philosophical matters	
Ontology of design	The philosophical study of the ontological basis for design theory and the activity of designing. It is at this level where human values, and the values and fundamental assumptions of research, are included in critiques of theory
Epistemology of design theory and the theories of objects	This is the level that contains those analyses and discussions about the critical study of the nature, grounds, limits and criteria or validity of design knowledge
General design theories	This is the level that is concerned with the details of those general theories which seek to describe the whole activity of designing and its relationship to the objects involved
Design process	
Theories about the internal processes of designers and collaboration	This level includes the descriptions of theories about the reasoning and cognition of individual designers, of negotiated design in collaborative design teams, and of cultural design effects on designers' output
Design process structures	The level that includes the theories about the underlying structure of design process, and the influence of domain, culture, artefact type and other similar attributes and circumstances
Design methods	The level in which theories about and proposals for design methods and techniques are described
Mechanisms of choice	The level of descriptions about the way that choices are made between different objects, processes, or systems, and how solutions are evaluated
Object	
Behaviour of elements	The level at which the behaviour of elements which may be incorporated into objects, processes and systems is described
Description(s) of objects	The level that encompasses simple descriptions of objects, processes and systems
Direct perceptions of reality	This is the level at which we 'sit on chairs', 'watch sunsets', 'hear the sound of a bird'

A design theory is typically elaborated on one of the ten levels. The design principles in our design theory are elaborations of 'design methods'. According to Love (2000), theories on design methods mostly provide knowledge on the methods and techniques that individual designers use. We use a slightly different interpretation of design methods as we prescribe institutional interventions that policymakers can use to organise design practices for public algorithmic systems in their public organisations. The design practices that are expected to emerge after implementation of the institutional interventions can be considered as design methods.

The ten abstraction levels of the meta-theory are interconnected and form a hierarchy. Starting from the top 'ontology of design' level, each level structures the level below. At the same time, starting from the bottom 'direct perceptions of reality' level, each level below demarcates the scope of upper levels. Since the design theory presented in Chapter 8 is situated in the 'design methods' level, the design principles are dependent on both the upper philosophical matters as well as the objects they concern. Therefore, we start by discussing the levels pertaining to philosophical matters (Section 7.2.1) and to objects (Section 7.2.2). Thereafter, Section 7.2.3 will elaborate the assumptions on the levels in which the design principles are situated.

7.2.1 Philosophical matters: socio-technical designing, democracy, and the Rule of Law

The juxtaposition in Section 7.1 forms the theoretical basis of the design theory in Chapter 8. In combination with the theoretical lens presented in Chapter 2 and the epistemology of design theory discussed in Chapter 3, we can synthesise the concepts of socio-technical designing, democracy, and the Rule of Law into the philosophical matters that form the basis of this thesis' design theory.

Ontology of design
Design-as-a-verb comprises sets of practices in which designers perform activities to transform an unsatisfactory situation into a preferred situation. Design practices consist of interactions between designers that are shaped by institutions and attitudes. These institutions and attitudes enable and/or constrain the behaviour of designers, but do not have a deterministic influence on behaviour. The institutions and attitudes exist before a design process starts but can be adapted throughout the process. Presuppositions shared among designers form the basis for the institutions and attitudes. Since the organisation of a design process influences the form and function of the design output, the presuppositions that underlie the design process will be recognisable in the design output.

Epistemology of design theory and the theories of objects
This thesis' design theory is a prescriptive theory. Grounded in descriptive, explanatory, and predictive theories, the design theory formulates design principles for institutional interventions that can reshape design practices. Actors, individuals or organisations, can use these principles to constitute institutional interventions that fit their own context. Because of this context-dependence and the uncertainty in human behaviour, the effects of institutional interventions on shaping design practices can never be fully predicted.

General design theories
The act of designing is a deliberative effort in which knowledge and interests are combined to create artefacts. The way in which this deliberation is approached can be based on different presuppositions. Part I of this research shows the implications of technocratic and businesslike presuppositions on organising design practices. The design theory presented in Chapter 8 is based on the presuppositions of socio-technical designing, democracy, and the Rule of Law. Synthesising these presuppositions results in four characteristics for design practices: inquiry, political steering, checks and balances, and flexibility.

First, design knowledge is the product of inquiry by designers. Through abductive reasoning, designers contextualise their expert knowledge to the specific context of the system to be designed, and the problem they are addressing. This adaptation of expert knowledge requires creativity and learning. As designing is about inquiry and abduction, design knowledge is constantly challenged and scrutinised in debates on design choices.

Second, inherent to these debates on design choices is a political struggle. Especially hard choices (i.e., value trade-offs in design choices) are to be made in a public debate. This follows from the primacy of the public over executive designers in making these hard choices. The primacy of the public can be ensured by establishing political steering on design goals and design constraints. However, as experts in the design process have a knowledge advantage and will scrutinise the outcomes of public debates, they have a strong position of power in the design process. Checks and balances are needed to reduce arbitrary use of designerly power.

Third, checks and balances between different designer positions can play a role in ensuring the quality and legitimacy of design output. Such checks and balances can be created by purposely establishing positions of different types of designers, and by formalising the actions these designers can perform, information available to these designers and control to be assigned to these positions. Moreover, checks and balances can be achieved through formalising feedback or dissent channels for designers and citizens. Apart from the formal structures, checks and balances require a specific attitude on the part of executive designers. They have to be reflective on their own work and be responsive to the input and comments of others at the same time. Moreover, since designers of public algorithmic systems are situated in a public administration context, they are expected to act along the lines of principles of good governance.

Fourth, considering the emergent properties, uncertainties, vagueness, and complexity of socio-technical systems, designing such systems is inherently idiosyncratic and continuous. The outcomes and effects of socio-technical systems cannot be fully predicted. This requires flexibility from designers, and iteration of the system design.

7.2.2 Object: socio-technical systems

The design theory in this thesis concerns design practices for socio-technical systems. More specifically, it deals with public algorithmic systems as object of design. This section outlines our perspective on such systems following the insights obtained in Chapter 6 and the definitions discussed in Chapter 2 (Section 2.2).

Behaviour of elements

The objects within this design theory are socio-technical systems. These systems consist of institutional, technological, and agential components that interact (Orlikowski, 1992). More specifically, the different components are entangled; they co-constitute each other. As the components are entangled, socio-technical systems have emergent properties; the system's outcome is more than the sum of its parts. Accordingly, designers should pay special attention to designing the interactions between components. In addition, the constitution of both the individual components as well as the interactions between components change over time. Consequently, the system's form and function in the future is inherently uncertain. Finally, a socio-technical system depends on the context in which it is situated and is subject to path-dependent choices made in the past.

Description of objects
Designers can obtain an overview of a socio-technical system through a socio-technical specification. Such a specification identifies the institutional and technological artefacts, and the human agents situated in the system. Moreover, it explicitly specifies the interactions between the components. The socio-technical specification also draws the system boundary and, therefore, should describe how the system interacts with elements outside of the system boundary. A socio-technical specification can inform designers working on the individual components. From the perspective of democracy and the Rule of Law, the socio-technical specification functions as a way to inform involved actors about design choices, and as a discussion piece.

Direct perceptions of realities
Citizens, as affected actors, are confronted with the consequences of socio-technical systems. Following Dewey (1927/2016) citizens form the public for designing public algorithmic systems. Accordingly, citizens have primacy in socio-technical designing in a democratic and Rule of Law context. Citizens' freedom should be guaranteed, and citizens should be protected from arbitrary use of power. In the case of public algorithmic systems, citizens can be confronted with Kafkaesque situations. In these situations, citizens perceive the public algorithmic system as incomprehensible, senseless, and rigid. The user of the system is also an affected actor. For those working with the algorithmic system, the system is mostly perceived as an internal work process that is dispersed over several actors responsible for different components of the system. For them, the system is complex because of organisational and technical inflexibility and scale.

7.2.3 Design process: coordinating and legitimising design choices

After elaborating the theoretical underpinnings and the object to which socio-technical designing is directed, we turn to the levels on which the design theory of this thesis resides. The next four levels present the assumptions related to the design process.

Theories about the internal processes of designers and collaboration
Designers are, like all human agents, 'creatures of bounded rationality' (Simon, 1969/1996). Their information on the context in which they design – both the context in which the system to be designed will be implemented, as well as the context in which design practices are situated – is limited. Designers anticipate and react to the design context and the behaviour of other designers. For them, the design process is a learning process in which their knowledge and experience on the artefact or system increases over time. In addition, they gradually gain insight into the perceptions, interests and values that other designers bring to the table. Collaboration between designers should concentrate on the co-evolution of problem and solution.

Design process structure
Designing a socio-technical system is a multi-actor effort that concurrently takes place in multiple but interconnected action situations. These action situations set the stage in which designers interact with each other. The interactions of designers result in a sequence of transformation of the system or artefact (Bots, 2007). This sequence is context-dependent and can

therefore not be fixed beforehand. Consequently, designing is not only a procedural but also a normative and substantive effort.

Design methods

Design practices should reduce arbitrariness in the design process of public algorithmic systems. Chapter 6 concluded that arbitrary conduct can manifest in coordinating and in legitimating design choices. Coordination of design choices should focus on the socio-technical components and their interactions within public algorithmic systems. The components and interactions are represented in a socio-technical specification, which is formulated by dedicated designers, so-called system-level designers. The expertise of the different system-level designers working on the specification should represent the different components in the public algorithmic systems. These system-level designers focus on designing the interactions between components and can consequently instruct component designers to create system components that are compatible with the whole system. The system-level designers also have a role in legitimating design choices. They are involved in a dialectic with representative designers that provide normative direction to design choices based on their democratic mandate. The overview of the system in the socio-technical specification enables the identification of critical design choices and reflects on the implications of design choices. This can inform public debate on the specific public algorithmic system and enable representative designers to engage in shaping the system. Since representative designers have a democratic mandate, they can decide the goals and constraints of the system to be designed. Chapter 8 elaborates on the role of both system-level as well as representative designers in the dialectic on design choices.

Mechanisms of choice

Making design choices is a political process happening on several hierarchical levels. Decision-makers at higher hierarchical levels give a political assignment to designers on lower levels. The other way around, executive designers provide advice on design choices to decision-makers. These executive designers collaboratively formulate design alternatives and identify hard or critical design choices that they will advise on. Both the alternatives as well as the identified hard or critical design choices are value-laden and subject to power dynamics between executive designers. As the executive designers' advice advances higher up the organisational hierarchy, the political debate on design choices becomes more public. At these higher organisational levels, other power dynamics are at play; stakeholders will try to influence both the interpretation of the advice, and the choices that are made according to their own interests (Flyvbjerg, 1998).

7.3 Conclusion

This conclusion discusses the theoretical underpinnings for embedding socio-technical design practices for public algorithmic systems in a democratic and Rule of Law context. The juxtaposition of the three presuppositions resulted in symbioses, shared challenges, and contradictions.

We considered a *symbiosis* between the presupposition to occur when characteristics of one presupposition are shared and compatible with other presuppositions and, therefore, reinforce the function of the different presuppositions. We identified four symbioses: iterative response, reflexive checks and balances, scaffolding political struggle, and pluralistic argumentation. Together, socio-technical designing, democracy, and the Rule of Law can instigate iterative and deliberative design practices in which political struggle is structured by checks and balances between reflective and responsive actors.

The three presuppositions *share challenges* in bringing their predominantly theoretical nature to practice. We identified three shared challenges: contingent on context of system to be designed, favourable attitude, and steering polycentric decision-making. The design practices are situated in a polycentric network of action situations that are dependent on their context. Moreover, designers need to have a favourable attitude towards supporting the institutions that structure the design practices in accordance with the symbioses.

Following their dissimilar nature, the three presuppositions also *contradict* each other. We identified three contradictions: public versus expert, flexibility versus (fixed) procedures, and majority decisions versus protection of minorities. These contradictions present trade-offs that the design theory has to deal with: whether it will put emphasis on public over experts, focus on the protection of minorities over majority decision-making, and whether fixed procedures can be minimised to engender flexibility.

This chapter also presented the meta-theory that synthesises the assumptions underlying the design principles for institutional interventions in design practices that will be presented in Chapter 8. The assumptions give directions on what it means to embed socio-technical design practices of public algorithmic systems in a democratic and Rule of Law context. More specifically, we can derive desired interactions between designers from the meta-theory. Insight into desired interactions is needed to formulate the design principles.

We infer the following four desired interactions between designers of public algorithmic systems. The first two interactions concern the inquiry that is performed in design through deliberation between designers to address coordination problems. Considering public algorithmic systems, there is a need for structural and continuous deliberation and alignment of design choices on a system-level. Within this inquiry two forms of argumentation should form interactions between designers:

1. Pluralism of argumentation. The different socio-technical components in public algorithmic systems should be represented in deliberations about design choices. In order to do this, a level playing field between designers from different disciplines and expertise is required. In other words, a balanced power distribution within design team that works on system-level design choices should be ensured.
2. Substantive argumentation. Stakeholders in a public algorithmic system and its design process should be informed about (critical) design choices and the underlying considerations. To ensure substantive argumentation, interactions between designers should focus on anticipating and reacting to causes of harm inflicted on citizens by algorithmic systems.

Deliberation does not only occur between executive designers. Critical design choices should also be coordinated and deliberated with representative designers. As such the lack of legitimation of design choices can be addressed. This requires a dialectic between political and system-level designers: a dialogue that aims to scrutinise, question, and improve design choices on a system-level. Design outputs of both groups of designers are the medium for this dialectic.

Two interactions are desired in this dialectic:

1. Iteration of design choices. Both political and system-level designers make design choices and should be open for reflections and critique of those choices. Accordingly, representative designers should refine their objectives for a public algorithmic system based on the socio-technical specification provided by system-level designers. Similarly, system-level designers should iterate their design choices based on the objectives determined by representative designers.
2. Primacy of the public. Representative designers, as representatives of the public, have primacy in making critical design choices. This power distribution between representative designers and system-level designers should be inherent to their interactions in the dialectic. Representative designers should be enabled by system-level designers to make the critical choices. In addition, it is up to representative designers to assess whether the design output of system-level designers fulfils the politicians' objectives.

Part III

Create & assess

Chapter 8

Design principles for institutional interventions in design practices of public algorithmic systems

> It is not the business of political philosophy and science to determine what the state in general should or must be, what they may do is to aid in creation of methods such that experimentation may go less blindly, less at the mercy of accident, more intelligently, so that men may learn from their errors and profit by their successes.
>
> – *John Dewey; The public and its problems, p. 83*

Part I and Part II researched the current and desired situation of design processes of public algorithmic systems respectively. This chapter focuses on formulating the design theory that bridges the gap between these two situations. As such, this chapter presents the main deliverable of this research. The design theory comprises design principles for institutional interventions (see Section 2.4.3 and Section 3.2.2). The design principles prescribe mechanisms and rule types that public organisations can use to realise institutional interventions that reshape their current design practices.

The main idea behind the design principles is to create a new action situation in the design process of public algorithmic systems: an action situation in which system-level designers formulate the socio-technical specification. The system-level designers function as an intermediary between the action situations of coordination and legitimation, see Figure 8.1. A system-level designer is a position filled by executive designers. The socio-technical specification, that system-level designers formulate, provides an overview of a public algorithmic system that can be used to coordinate the design of individual system components and to facilitate public debate on design choices. In other words, the system-level designers interact with politicians by equipping them with information for a meaningful public debate through the specification, and the politicians direct the design activities of the system-level designers. At the same time, the system-level designers interact with component designers. Component designers create the individual system components of public algorithmic systems, for example, a policymaker who formulates work instructions or a data scientist who writes the code for an algorithmic application. The position of component designer is also filled by executive designers. System-level designers coordinate and scope the design tasks of component designers, and the component designers provide their insights and requirements to the system-level designers.

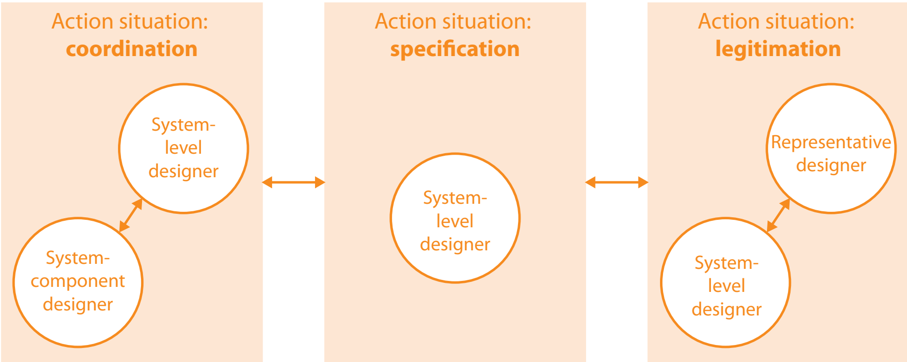


Figure 8.1 Action situations for which the design principles prescribe institutional interventions

Although the three operational action situations are on the same hierarchical level (see Section 2.4.2), power differences between the action situations exist. The representative designers have a democratic mandate and, therefore, set the boundaries and aims that both the system-level as well as the component designers should adhere to. Accordingly, we address the lack of democratic control over system-level bureaucrats identified by Zouridis et al. (2020). The system-level designers have coordination power over component designers; they provide component designers with specific design tasks. At the same time, both the politicians and system-level designers are dependent on their subordinate levels for information about the specifics of the component designs and the socio-technical specification respectively.

This chapter presents four design principles that prescribe institutional interventions to structure the 'specification' – design principles 1 and 2 – and the 'legitimation' – design principles 3 and 4 – action situations. We did not include the coordination action situation in this iteration of the design theory. This action situation is dependent on the socio-technical specification formulated in the specification action situation. Therefore, we first focus on the system-level designers who work on the socio-technical specification in collaboration with representative designers. In subsequent iterations of the design theory, the coordination action situation should also be specified.

Sections 8.1 up to 8.4 each describe one design principle in the design theory. These four sections start with identifying the gap between the current interactions between designers identified in Part I and the desired interactions that can be derived from Part II. Thereafter, the institutional interventions that bridge the gap and the expected effects of the interventions are presented. Finally, the actual design principle with associated rule types is formulated. Section 8.5 concludes this chapter by reflecting on the design principles.

8.1 Design principle 1: position of system-level designer

The first design principle focuses on establishing a group of system-level designers that will populate the 'specification' action situation. The main task of these designers is to formulate the socio-technical specification.

Gap between current and desired interactions

The need for the 'specification' action situation arises from the lack of coordination between executive designers from different disciplines. Following the findings of Chapter 6, this lack of coordination can result in possibilities for arbitrary conduct in public algorithmic systems. Part I of this thesis identified two interactions between designers related to this lack of system-level coordination: the coordinating role of technical designers and the asynchronous work rhythms. Our empirical research showed that designing public algorithmic systems is already approached as deliberating, aligning, and seeking agreement on design choices. Nonetheless, this deliberation happens ad hoc and between disciplinary designers who are situated in organisational silos.

Among the organisational silos, technical designers have obtained a coordinating role in design practices of public algorithmic systems. In combination with the client-supplier relationship between technical and domain designers, the knowledge and expertise of non-technical designers is underrated. Consequently, the design process tends to concentrate on the technical artefacts in public algorithmic systems.

Moreover, the variation in work rhythms between different disciplinary designers impedes a comprehensive overview of public algorithmic systems and a systemic approach towards designing the whole socio-technical system. The ad hoc and project-based approach currently demonstrated in design practices impedes coordination between executive designers with different disciplinary backgrounds (e.g., engineering or legal backgrounds), makes the design process vulnerable to losing knowledge when personnel changes, and reinforces a focus on developing algorithmic applications in a procurement dynamic. The project-based and chronological order of design processes in public organisations does not correspond with the continuous evolution of public algorithmic systems.

Considering the insights from Part II of this research, the current interactions between designers should be transformed into the following desired interactions. Instead of the ad hoc approach, design choices on a system-level (i.e., not only focusing on technological artefacts) should be structurally and continuously deliberated and aligned. In this deliberation, pluralism of argumentation should be ensured by creating a level playing field between designers from different disciplines. This addresses both the coordinating role of technical designers as well as the disregard of non-technical disciplines. The level playing field requires mutual understanding between disciplinary designers on the interactions between system components. After all, system-level designers will not be dealing with the details of each system component but need to understand when a complex web of system components is coherent.

Institutional interventions and expected effects

The institutional interventions prescribed by design principle 1 aim to establish both the 'specification' action situations as well as the position of the system-level designer. The system-level designers are assigned the task of formulating the socio-technical specification. To arrive at such a specification, the system-level designers are a group of disciplinary designers that represent all disciplines associated to the components in the public algorithmic system. The form of the specification should provide system-level designers with a shared vocabulary to discuss the alignment of institutional and technical artefacts (see Section 6.3.2). This does not mean that they should fully understand the ins and outs of each other's expertise. Instead, such a language should transcend their discipline and focus on creating an overview of socio-technical components and their interactions.

We expect that the use of the interventions based on design principle 1 results in a forum for deliberation about the interactions between system components and, thereby, supporting the alignment between system components. More specifically, this means that we expect that the shared vocabulary of a socio-technical specification results in enlarged mutual understanding between system-level designers on how system components interact. This mutual understanding is dependent on pluralism of disciplines among system-level designers, as we expect that pluralism will engender a comprehensive overview of the public algorithmic systems (see symbiosis 4 in Section 7.1.1). Finally, we expect that system-level designers' focus on the socio-technical specification can ensure continuous design. In the case that the forum of system-level designers exists throughout the whole life cycle of the algorithmic system, it will maintain an overview of the system and can identify needed adaptations. As system-level designers are performing the same design activity in this action situation, the problem of variation in work rhythms is also solved. At the same time, designers of the individual system components can hold on to their own work rhythm.

Design principle:
Establish an organisational level responsible for the socio-technical specification of public algorithmic systems in order to ensure mutual understanding about and the design of interactions between system components

This design principle establishes the new position of system-level designers who work on the socio-technical specification of public algorithmic systems. Three rule types elaborate the principle:

- A position rule that establishes the position of system-level designer.
- A boundary rule that determines what official designers should be included in the group of system-level designers and what disciplinary knowledge should be represented in the action situation. The disciplines should reflect the components in the public algorithmic system.
- A choice rule that defines the actions these system-level designers can take. System-level designers can formulate the socio-technical specification and coordinate the alignment between components in the public algorithmic system.

8.2 Design principle 2: the socio-technical specification

Where design principle 1 establishes an action situation in which a socio-technical specification is formulated, the second design principle defines the features of this socio-technical specification.

Gap between current and desired interactions
Part I observed that those actors who are executive designers have a narrow view on public algorithmic systems. They tend to consider algorithmic applications as simple automation. This attitude reduces public algorithmic systems to a technological artefact. Executive designers

pay little attention to how algorithmic applications interact with institutional artefacts and human agents. Consequently, they lack insight into the implications of design choices for the whole public algorithmic systems or into the possible effects of the system beyond its system boundaries. This impedes anticipation and reaction to Kafkaesque situations emerging from public algorithmic systems.

Instead, Part II argues that providing substantive argumentation behind design choices should be encouraged and that implications should be made explicit. Argumentation is standard practice within the Rule of Law, for which the principles of motivation and meticulousness play an important role (see symbiosis 4 in Section 7.1.1). Following from these principles, interactions between designers should focus on anticipating and reacting to causes of harm inflicted on citizens by algorithmic systems.

Institutional interventions and expected effects

Apart from providing a common language that supports understanding between system-level designers, the socio-technical specification should also support designers in forming the argumentation behind design choices and ensure the thoroughness of the inquiry of implications of design choices. Principles for good governance provide grounded guidelines that encourage designers to anticipate implications of design choices and make those explicit (see Section 7.1.1). We focus on two good governance principles, which also form the expected effects of design principle 2: motivation and meticulousness.

First, following from the motivation principle, governmental organisations are expected to provide well-grounded argumentation behind their decision-making. We expect that our interventions will strengthen the motivation behind design choices and make those motivations known. Similar to motivating decisions, the Rule of Law prescribes meticulousness in making decisions that affect citizens. A decision-making process is meticulous when all interests of involved or affected actors are considered and when proportionality is ensured in trade-offs (cf. Wolswinkel, 2020). We expect that checking the socio-technical specification on possibilities for arbitrary conduct will increase the meticulousness of making design choices.

Design principle:

Establish formal conditions to which the socio-technical specification of public algorithmic systems should adhere to, in order to standardise the motivation behind and meticulousness of design choices

The second design principle concerns the output that system-level designers should deliver. More specifically, it sets requirements for the output to ensure information flows and pluralistic argumentation. The requirements are set in the following three rules:

- An information rule that establishes how the socio-technical components and their interactions should be communicated in the socio-technical specification.
- A scope rule that defines a motivated and meticulous socio-technical specification of different design alternatives as the outcome of the 'specification' action situation and details the requirements for motivation and meticulousness.
- A pay-off rule that determines the costs (i.e., consequences) associated with design choices that do not adhere to the criteria for motivation and meticulousness.

8.3 Design principle 3: reflection and learning

The socio-technical specification that the system-level designers formulate is a means to communicate with representative designers. Design principle 3 concerns the 'legitimation' action situation and structures the dialogue between the two groups of designers. In practice, this action situation already exists. The following two design principles are ways to restructure the existing action situation.

Gap between current and desired interactions

The empirical results in Part I show a disconnect between representative and executive designers. The gap between representative and executive designers is widened by miscommunication between and misinterpretation by both parties. The third design principle concentrates on the observation that executive designers share insufficient information about public algorithmic systems with representative designers. Moreover, it addresses the reactive and incidental political debate that was observed among representative designers.

Part II provides insight into desired interactions that support an informed and meaningful public debate on design choices for public algorithmic systems. Both representative and system-level designers should focus on the design output provided by the other group of designers to instigate reflection and learning. This requires both groups of designers to engage in a dialogue based on their design outputs. In other words, representative designers should refine their objectives for a public algorithmic system based on the socio-technical specification provided by system-level designers. Similarly, system-level designers should be reflective of what political objectives (formulated by representative designers) mean for the design choices that they make when formulating the socio-technical specification.

Institutional interventions and expected effects

The institutional interventions instigated by design principle 3 establish a dialectic interaction between representative and system-level designers in making design choices. Therefore, it defines the roles of both groups of designers in that dialectic interaction and demarcates the design outputs that form the medium through which the two types of designers interact. Representative designers are responsible for setting the normative objectives and boundaries of the to-be-designed public algorithmic systems (see symbiosis 2 in Section 7.1.1). Thereby, representative designers should focus on design goals and design constraints. On the other hand, system-level designers should provide representative designers with insight into what kind of public algorithmic systems are possible within these goals and constraints (see the discussion on the contradiction 'public vs. expert' in Section 7.1.3). Therefore, the system-level designers should draft the socio-technical specification and reflect on possible manifestations of arbitrary conduct in the algorithmic system that will emerge from the specification. Through the information provided in these design outputs, the politicians and system-level designers can sustain a dialogue.

The use of these institutional interventions is expected to resolve miscommunication between the two groups of designers. The exchange of design outputs will result in enlarged mutual understanding between representative and system-level designers. Providing the design output to each other, enables designers to clarify, adapt and improve their own design output. At the same time, using the design outputs in response to one another, is expected to create a dialogue that stimulates reflection and learning on both sides. Representative designers are informed by system-level designers about the consequences of their own design goals and design constraints. This enables reflection on the desirability and feasibility of these goals and constraints. System-level designers can use the reaction of representative designers to their socio-technical specification to enlarge their understanding about what the design goals and design constraints actually mean. Finally, a dialogue between politicians and system-level designers brings flexibility to the design process. Both groups of designers are provided with new knowledge on which they can base changes to their own output. As such, the dialogue also counterbalances the formalism that follows from the socio-technical specification prescribed by design principle 2. Flexibility is also established by giving system-level designers a mediating role between component designers and politicians. System-level designers bring all information relevant to politicians together, while component designers can perform their design activities undisturbed and can follow their own work rhythm.

Design principle:
Establish a structured dialectic between representative designers and system-level designers via their design outputs in order to stimulate learning among and reflection by both groups of designers

Design principle 3 prescribes the interaction between the two types of designers. This interaction is mediated through design outputs. Two rules define the outputs that each group of designers will deliver:

- A scope rule that demarcates the design output delivered by system-level designers to representative designers. They should formulate a design advice that informs politicians about different system alternatives, the critical design choices in those alternatives, and the (possible) implications of those design choices.
- A scope rule that demarcates the design output by representative designers. Representative designers formulate design goals and design constraints that demarcate the design space of system-level designers.
- A choice rule that assigns the responsibility for both the representative as well as the system-level designer to communicate their design output to the other designer.

8

8.4 Design principle 4: checks and balances

Where design principle 3 divides the roles between system-level designers and politicians, design principle 4 prescribes institutions that guarantee power balances between these two positions.

Gap between current and desired interactions
The final interaction observed in current design practices concerns representative designers who refrain from determining a normative or political instruction that adequately steers the problem-formulating and -solving activities of executive designers. Chapter 6 argues that the lack of democratic legitimacy in the design process is an important factor in the emergence of manifestations of arbitrary use of power, mainly if executive designers use their designerly power arbitrarily. In addition, the policy instruments developed and implemented by public

organisations do not address the reluctance of representative designers to make hard design choices.

Instead, representative designers should be enabled to make hard design choices. System-level designers should provide representative designers with the right information (see design principles 2 and 3) but also provide a preferred design alternative for the algorithmic system to be designed. On the other hand, decision-making power among the designers should be based on the principle of self-government of citizens. To ensure the primacy of the public, representative designers should assess whether the design output of system-level designers fulfils their design goals and design constraints (see symbiosis 3 in Section 7.1.1).

Institutional interventions and expected effects

The institutional interventions following from design principle 4 should assign control over design choices to the specific positions in the action situations. The control is based on the primacy of the public and creating a level playing field for pluralistic argumentation. First, as design principle 1 and 2 firmly establish the position of system-level designers, their influence on design choices can increase. Accordingly, the possibility of arbitrary use of designerly power also increases. To address the inherent tension between experts and the public in democracy, the institutional interventions in design principle 4 counteract the influence of system-level designers on design choices by making the hierarchy between representative and system-level designers explicit. Representative designers will have the last say in determining whether the algorithmic system design falls within the goals and constraints that they formulated.

In addition, this design principle also determines the checks and balances within the team of system-level designers to prevent the domination of one designer with a particular discipline over other disciplines. System-level designers should strive for consensus on their design alternatives and, if that is not achievable, assign decision-making power to one of the system-level designers. A final means to address power imbalances within design teams is to provide individual designers with the possibility to add their individual opinion about the advice that their team sends to the political level.

We expect that the use of the institutional interventions will result in support for making design choices and increased political steering by representative designers. As mentioned before, system-level designers provide representative designers with information about critical design choices that are (to be) made, and the implications of these design choices. Representative designers are responsible for making these hard design choices. Their means of political steering are strengthened by 1) having the initiative to set design goals and design constraints, and 2) making the final assessment whether the output of system-level designers is compatible with their goals and constraints.

Design principle:

Establish checks and balances between system-level designers, and between representative designers and system-level designers in order to reduce arbitrary use of designerly power

The last design principle focuses on structuring the power relations between designers. The starting point is reducing arbitrary use of designerly power. Regarding the interactions between system-level designers, several aggregation rules ensure a balance of power:

- An aggregation rule that determines that the group of system-level designers should strive for consensus on their design choices, and that assigns one of the system-level designers the responsibility of making the final decision on a preferred design alternative in case of conflicting views.
- An aggregation rule that assigns the right to individual designers to write an individual opinion considering the choices made in the design advice.
- An aggregation rule that determines that design activities of system-level designers cannot be assigned to external parties and that only activities on the component level can be outsourced.

Regarding the interaction between representative and system-level designers, a rule that ensures that representative designers (as representatives of the public) have primacy in determining the form and function of public algorithmic systems:

- A choice rule that governs that politicians will determine whether the socio-technical specification falls within the goals and constraints that they have formulated.
- An aggregation rule that assigns final decision-making power over the socio-technical specification to representative designers.

8.5 Reflection

This chapter contains the formulation of four design principles that policymakers can use as a template for institutional interventions in order to reconfigure current design practices for public algorithmic systems. The interventions create the position of system-level designers who formulate the socio-technical specification of public algorithmic systems. In addition, these system-level designers have the task of coordinating the creation of the individual components of the algorithmic system and the interactions between these components. Furthermore, these designers are engaged in a dialectic with representative designers to guarantee the legitimacy of design choices pertaining to the socio-technical specification. We arrived at four design principles that support public organisations in establishing the position of system-level designers:

DP 1: *Establish an organisational level responsible for the socio-technical specification of public algorithmic systems in order to ensure mutual understanding about and the design of interactions between system components*
DP 2: *Establish formal conditions to which the socio-technical specification of public algorithmic systems should adhere to, in order to standardise the motivation behind and meticulousness of design choices*
DP 3: *Establish a structured dialectic between representative designers and system-level designers via their design outputs in order to stimulate learning among and reflection by both groups of designers*
DP 4: *Establish checks and balances between system-level designers, and between representative designers and system-level designers in order to reduce arbitrary use of designerly power*

Concerning the coherence and consistency between principles, the four principles together structure all variables in the action situations of specification and legitimation. DP1 assigns the design activities that system-level designers are expected to execute. DP2 demarcates the contents of the socio-technical specification that is delivered by system-level designers. DP3 structures the interaction of the new position with representative designers, and DP4 distributes influence on the specification over the different types of designers and assigns primacy in decision-making on design choices to representative designers.

The design principles mostly prescribe formal institutions. But as emphasised in the meta-theory, a favourable attitude of responsiveness and reflectiveness among designers is elemental to socio-technically designing public algorithmic systems in a context of democracy and the Rule of Law. The formal institutions are probably insufficient in instigating such attitudes but should be complemented by informal institutions. These informal institutions are less susceptible to deliberate design. Nonetheless, formal institutions based on the design principles will reshape interactions in such a way that attitudes among designers will also change (see the learning loops in the IAD framework in Section 2.4.2). Apart from the formal institutions based on the design principles, public organisations should provide system-level, component, and representative designers with support in working more responsively and reflectively. This asks for leadership among actors in design process of public algorithmic systems (Selznick, 1984).

Chapter 9

Evaluation of the design theory

> No change in a set of rules is ever sufficient to solve the next set of problems created by new opportunities and constraints that continually arise in an evolving human community.
>
> – *Elinor Ostrom, Understanding Institutional Diversity, p. 214*

The design principles presented in Chapter 8 can be used by public organisations to develop institutional interventions in their design processes of public algorithmic systems. For the purpose of evaluating the principles, we specified our own institutional interventions based on the design principles. This chapter presents the test of these interventions. Through this test, we evaluated the extent to which the design principles result in the expected interactions between designers. This evaluation episode closes the last generate-test cycle of this research. Thereby, we answer the following question:

To what extent do institutional interventions based on the design principles result in the expected interactions between designers?

We performed an explorative evaluation of the design theory through a simulation of a design process. Section 3.3.3 described our evaluation approach, which starts by formulating the goal of the evaluation. The goal of the evaluation is to test whether the institutional interventions in design processes result in the expected interactions between designers. Based on this goal, we selected simulation as the most suitable evaluation strategy. We performed a simulation of a design process in which representative and system-level designers were asked to produce design outputs in reaction to a problem formulation that could be addressed with an algorithmic system. The problem formulation reads as follows:

The national government has asked municipalities to execute the allocation of energy cost allowances. This energy cost allowance aims to support households that are no longer able to pay their energy bills because of increasing energy prices. It is the municipality's task to ensure that all eligible households will receive the allowance. However, the municipality lacks an overview of all eligible households.

This chapter discusses the results of the simulation performed in a Dutch municipality. First, Section 9.2 discusses how we translated the design principles into institutional interventions. It describes the individual evaluation episode and presents the evaluation measurements. Section 9.3 presents the simulation results and evaluates the four design principles individually. Section 9.4 concludes this chapter.

9.1 Instantiations: institutional interventions

To test and evaluate the design theory, we translated the design principles into institutional interventions that can be used in practice. In design science, this translation is called an *instantiation* – i.e., a practical intervention in reality that represents the design theory in such a way that it can be tested (Gregor & Jones, 2007). Instantiations are not one-on-one representations of the theory. In translating the design theory to practice, the principles are interpreted and pragmatically situated in a particular context.

The instantiations used in the simulation are based on the rule types presented in Chapter 8, in the specific context of Dutch municipalities. We created the institutional interventions through two trial simulations with respectively two and three fellow PhD candidates. This section presents the institutional interventions as used in the final simulation. See Appendix B1 for an overview of all changes made to the interventions after the two trials. This section will highlight the most fundamental changes to the interventions.

The following four subsections present the instantiations for each of the four design principles and the measurement variables related to the principles. These measurement variables are the evaluation properties for our design theory, see Section 3.3.3. Together, the measurements relate to the goal of the evaluation: to test whether the institutional interventions in design processes result in the expected interactions between system-level and representative designers. Table 9.1 presents the instantiations and measurements related to each design principle. We elaborate on these in the next four subsections.

9.1.1 Instantiation of design principle 1: participating actors

As discussed in the previous chapter, the main idea in the design theory is to establish the position of system-level designer (i.e., comprises all designers working on the socio-technical specification of a public algorithmic system; moreover, system-level designers are a subset of executive designers). This section elaborates on how we defined that position in the simulation. Since the four design principles also concern the interaction of system-level designers with representative designers, this section also discusses the actors who were considered representative designers in the simulation.

Table 9.1 Instantiations and measurement variables related to each of the four design principles

DP	Design principle	Instantiation	Measurement
1	Establish an organisational level responsible for the socio-technical specification of public algorithmic systems in order to ensure mutual understanding about and the design of interactions between system components	System-level designers are a group of executive designers whose expertise reflect the socio-technical component in a public algorithmic system	• Mutual understanding between system-level designers
2	Establish formal conditions to which the socio-technical specification of public algorithmic systems should adhere to, in order to standardise the motivation behind and meticulousness of design choices	A socio-technical specification comprises a system map, the identification of critical design choices, and a reflection on possible manifestations of arbitrary conduct	• Motivation • Meticulousness
3	Establish a structured dialectic between representative designers and system-level designers via their design outputs to stimulate learning among and reflecting by both groups of designers	Four alternating design sessions in which representative designers formulate a political instruction, and system-level designers draft a design advice	• Mutual understanding between groups of designers • Reflection and learning
4	Establish checks and balances between system-level designers, and between representative designers and system-level designers in order to reduce arbitrary use of designerly power	Instruct system-level designers to arrive at consensus on a preferred design alternative, provide individual system-level designers with the possibility to write an individual opinion on the design advice, and give representative designers the responsibility to assess the design advice	• Enabling making choices • Political steering

System-level designers

We formed a group of system-level designers in line with the position and boundary rule in design principle 1. Accordingly, the involved system-level designers had to represent the different socio-technical components in public algorithmic systems. The aim was to involve five to six public servants who represent similar disciplines as the interviewees in the explanatory empirical study. We were looking for designers working on technical artefacts, domain-related institutional artefacts, and compliance-related artefacts to formulate the socio-technical specification of the public algorithmic system in the simulation (in line with the choice rule under design principle 1). The categories of participants are described broadly since every public organisation has its own distribution of roles and responsibilities among public servants. We presented the following criteria to the participating public organisation and asked them to propose participants for the simulation that met the criteria:

- **Technology;** actors that are engaged in developing and embedding algorithmic applications. We distinguish two functions:
    - Designers or developers of algorithmic applications, for example, a machine learning or regression model – e.g., data scientists and data analysts

  - Those constituting the information architecture of the organisation. These can be technically- or policy-oriented functions that aim to embed the algorithmic applications in the broader/general information architecture of the organisation – e.g., information architects
- **Domain;** an algorithmic system is always related to a specific policy. For example, the designed algorithm can be part of the execution of benefit provision (verstrekking). In that case, the knowledge of policymakers in the area of benefit policies is needed in the design process. These policymakers are closer to public servants who have to work with the algorithmic system compared to the technical designers. Therefore, the design team includes:
  - Policymakers who are responsible for or are knowledgeable about the policy that forms the basis of the algorithmic system. These are policymakers with domain expertise. Preference: policymakers who have already been involved in design process(es)
- **Compliance;** algorithmic systems must comply with laws and regulations. Actors who check compliance have an important function in guaranteeing or protecting boundaries in the design process. We distinguish two functions:
  - Policymakers who work on new initiatives in the field of responsible use of algorithmic systems. These are policymakers who develop new policies regarding data and digitalisation. For example, policymakers who develop the governance of algorithms
  - In addition, more 'traditional' compliance functions that are important in the design process. These are functions that ensure that algorithmic systems comply with current/existing laws and regulations. For example, privacy officers, security officers, legal advisors

This list only comprises actors working within the public organisation. Thereby, it also enforces the aggregation rule prescribed by design principle 4 (i.e., system-level designers cannot be external actors).

Eventually, three public servants participated in the simulation as system-level designers. S1 is an information strategist working at the municipality for 9 years. He was the only participant who was not involved in a design process before. S2 is a policymaker specialised in the social domain (e.g., income-related topics), which is related to the specific problem used in this simulation. S2 works for more than 10 years at the municipality, of which 2.5 years as a policymaker. Finally, S3 is the CIO (chief information officer) of the municipality. He worked for the municipality for 23 years but left the organisation just before session 4 took place.

Initially, two other public servants were supposed to join the simulation: a privacy officer and an application architect. Together, the five public servants would represent the full list presented above. We postponed the session we initially planned because one of the participants was ill. Unfortunately, the privacy officer and application architect were not able to join the newly planned session because of sickness and personal circumstances. We decided to continue the session as the CIO, being the main contact for this explorative evaluation, was leaving the municipality. The privacy officer did fill in the first questionnaire. His answers are included in the research data to enrich our insight into the current situation of design practices in the public organisation. During the simulation sessions, the public servants were often referring to what the privacy officer would likely add to the discussion.

Representative designers

The position of representative designer in the 'legitimation' action situation is held by actors in public organisation that have the democratic mandate to set policy goals, create legislation, and control executive branches of government. Mostly, politicians in a representative body have this democratic mandate. Nonetheless, public organisations designing public algorithmic systems do not always interact directly with a representative or political body. In the Netherlands, an executive agency depends on the ministry it is associated with for political guidance. In that case, the group of politicians in the experiment could be replaced by policymakers of the associated ministry.

The position of representative designers already exists in public organisations and, therefore, it is not established through the design theory. However, for the purpose of the simulation, we had to identify the actors who hold the position of representative designer in the participating public organisation. Following the situation in Dutch municipalities, we consider city council members as representative designers.[1] This is also in line with representative democracy being the starting point of the design theory.

Our aim was to engage three to four politicians, preferably representing a broad range of political parties or ideologies. In the end, three local politicians participated in the simulation. All three politicians were members of political parties at the left side of the political spectrum and occupied their seat in the city council since the most recent elections (March 16, 2022). P1 is a council member for the Green Party (*GroenLinks*), P2 is a council member for the socialist party (*SP*), and P3 is an auxiliary member for the council members of the Green Party.

Measurement variables

Mutual understanding between system-level designers

The main expectation from the institutional interventions following from design principle 1 was to enlarge mutual understanding between system-level designers through collaboratively formulating the socio-technical specification. In this simulation, we considered mutual understanding to be reached when an individual designer comprehends the statements or outputs of another individual designer. We operationalised mutual understanding as the alignment between the intention or provided meaning by one individual or group given to oral statements or written outputs, and the interpretation of those by another individual or group.

In observing mutual understanding, we focused on:

- Instances of misunderstandings between participants in deliberating design choices
- The extent to which the socio-technical specification is used to reach understanding within the deliberation

1 The Alderman can also be considered a representative designer in the case of municipalities. In reality, the Alderman would have a bridging function between city council members and public servants. Nevertheless, we decided not to include an Alderman in the simulation in order to test the dialectic between political and system-level designers. The Alderman would bring a third position to the action situation. Including the position of Alderman or Minister (in the case of a national public organisation) is a possible direction for improving the design theory in future research.

9.1.2 Instantiation of design principle 2: system map

As prescribed by design principle 1, system-level designers are responsible for formulating the socio-technical specification of a public algorithmic system. This section elaborates on the form of the socio-technical specification used in the simulation based on design principle 2.

In the simulation, the socio-technical specification was communicated through a system map (see information rule design principle 2). A system map is a diagram of all socio-technical components in the public algorithmic system, and the interactions between the components (the instruction to the simulation included an explanation of the system map, see Appendix B3.2). We asked system-level designers to depict institutional artefacts as triangles, technological artefacts as squares, human agents as circles, and interactions as arrows. The squares, triangles, circles, and arrows should specify the artefact, human agent, or interactions in words. Participants had the possibility to use a white board and blank sheets of paper to draw the system map.

The scope rule of design principle 2 determines that the outcome of the 'specification' action situation should be a motivated and meticulous socio-technical specification. Participants were instructed to identify critical design choices in the socio-technical specification, write down their arguments behind design choices, and to reflect on their own design using the four manifestations of arbitrary conduct. Critical design choices were defined as: 1) choices that relate to specifying the components of the design alternative, and 2) choices for which a public servant does not have a mandate to make a decision.

Measurement variables

Motivation

Design principle 2 is expected to result in institutional interventions that strengthen the motivation and meticulousness of design choices. We defined motivation in line with the interpretation of the motivation principle (*motiveringsbeginsel)* codified in Dutch administrative law:

- The extent to which the system map and associated argumentation can support design choices
- The extent to which the system map and associated argumentation are internally consistent
- The extent to which the system map and associated argumentation are insightful to actors other than the system-level designers
- The extent to which the motivation is (made) knowable to the politicians
- The extent to which the system map and associated argumentation support system-level designers in formulating the motivation behind design choices

Meticulousness

Meticulousness concerns the mapping of the interests of stakeholders in the system and communicating the trade-offs made between the interests of stakeholders. Similar to motivation, we followed the interpretation of the meticulousness principle (*zorgvuldigheidsbeginsel*) codified in Dutch administrative law:

- The extent to which all socio-technical components relevant to the problem formulation are identified and elaborated in the system map and associated argumentation

- The extent to which both the system-level designers and the politicians consider and examine the consequences of public algorithmic systems for 1) affected citizens, and 2) internal practices in the public organisation
- The extent to which the system map and associated argumentation support system-level designers in guaranteeing meticulousness in arriving at design choices

9.1.3 Instantiation of design principle 3: political instruction and design advice

The third design principle considers the structured dialectic between representative and system-level designers. This dialectic happens via the design outputs of both groups of designers. This section elaborates on the two design outputs – i.e., a political instruction and a design advice – used in this simulation and discusses the instantiation of the structured dialogue.

We asked the representative designers to draft a political instruction. The contents of the instruction followed from the second scope rule in design principle 3. The instruction consisted of a list of design goals – i.e., statements on what the to-be-designed system should achieve – and a list of design constraints – i.e., boundaries to the design space (see the second scope rule of design principle 3). In addition, we asked representative designers to make a prioritisation of both the goals as well as the constraints.

The system-level designers were asked to produce a design advice (see the first scope rule of design principle 3). The design advice elaborated several alternative socio-technical specifications that satisfy the political instruction. To ensure that several of the design alternatives concerned public algorithmic systems, we included in the instructions to system-level designers that at least two of the design alternatives should incorporate an algorithmic application. For each design alternative, the system-level designers had to create a socio-technical specification (in the form of a system map), identify the critical design choices in the specification, and discuss the implications of design choices (see Section 9.2.2). The main goal of the design advice was to inform the representative designers about the types of public algorithmic systems possible within the boundaries that they have formulated.

Design principle 3 focuses on the exchange of the two design outputs between political and system-level designers in a structured dialectic. In this simulation, we interpreted this as an alternation of sessions in which representative and system-level designers separately worked on their own design outputs. Both the representative as well as the system-level designers were informed about each other's output at the start of each session (and were instructed to work with the input). The simulation consisted of four sessions: two sessions with politicians that alternated with two sessions with public servants. The chronology of the dialectic is presented in Figure 9.1.

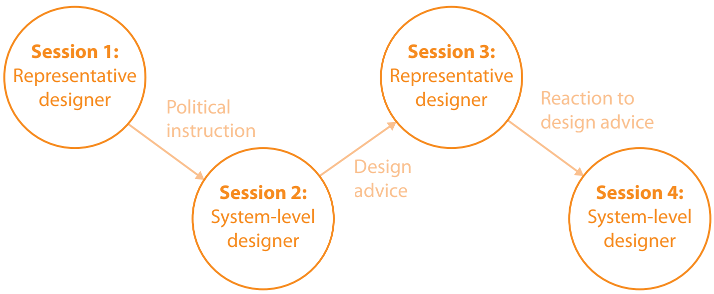


Figure 9.1 The actors in the four sessions in the simulation

9

The first two sessions were mostly aimed at the creation of artefacts. The last two sessions focused on reacting to and reflecting on the design output of the other designers. Appendix B3 presents the slides used by the instructor for instructing the participants in each session.

In the first session, the politicians drafted the political instruction based on the problem formulation. This design activity started with an individual part in which the politicians could identify goals and constraints themselves before collectively formulating goals and constraints. Thereby, we prevented politicians from influencing each other's thought processes. The formulation of goals and constraints was followed by prioritising both lists with goals and constraints. The politicians were only informed about this prioritisation after the identification of goals and constraints. As such, politicians were first stimulated to think about a variety of goals and constraints before making political choices between them.

After concluding the session with the politicians, the first session with system-level designers focused on drafting their design advice. The instructor started the session with a presentation of the problem formulation and the political instruction formulated by the representative designers. The drafting of the advice was divided into five steps. This provided the public servants with an overview of what we expected them to do, because the trial sessions showed that without dividing session 2 into steps, the participants were overwhelmed by the task they were asked to perform (see Appendix B1). Therefore, we decided to split the session up as follows. First, the system-level designers refined the design assignment through discussing the problem formulation and political instruction with the aim of arriving at a shared interpretation of these inputs. Second, the public servants produced preliminary design alternatives – i.e., general descriptions of the idea behind each alternative. Again, the trial sessions showed that participants have the inclination to start with one alternative, leaving no time for specifying other alternatives. In addition, through this second step, we could instruct designers to create at least two algorithmic systems. The third step consisted of elaborating the three alternatives in system maps. Designers were asked to reflect on their own designs by identifying critical design choices and their implications in step 4. We presented the manifestations of arbitrary conduct to the designers. They could use this for their reflection on their own design choices. Finally, participants were asked to formulate their advice (including giving arguments for a preferred alternative, see Section 9.1.4).

In the third session, we presented the advice of the public servants to the politicians. First, we asked the politicians to compare the advice with their own political instruction. Based on this comparison, we asked them to reflect on their own goals and constraints and to adapt them (or their prioritisation) where needed. Finally, we asked the politicians to formulate questions and/or instructions for the public servants in reaction to the design advice.

In the last session, we returned to the public servants. Like the third session, we presented all the material from the preceding sessions to the public servants. We also asked the public servants to both reflect on their own advice as well as react to the politician's output of session 3. As such, this session was the last step in the dialectic prescribed by design principle 3. The time constraint of this PhD project did not allow for adding more sessions. In reality, the dialectic can continue from here with new alternating sessions with politicians and public servants. These sessions would be similar to sessions 3 and 4.

Measurement variables

Mutual understanding between the two groups of designers
Like we expected enlarged mutual understanding among system-level designers in design principle 1, we expected the interventions based on design principle 3 to enlarge mutual understanding between representative and system-level designers. Using the same interpretation of mutual understanding mentioned before, we focused our observations for this design principle on:

- The extent to which the design outputs are understood by participants
- In the case that the design outputs are not readily understandable, the extent to which the interpretations of the outputs by other designers align with the intentions of the makers

In addition, we also considered mutual understanding within the group of representative designers. To evaluate this, we used the measurement variables used for design principle 1.

Reflection and learning
We expected that the confrontation with a reaction (e.g., the design advice) to a group's own work (e.g., the political instruction) will result in reflection on that own work. For example, the representative designers were expected to adapt their political instruction after being informed by the design advice. Such a reflection can result in a reaction provided to the other group and, therefore, to co-evolutionary learning. In that case, the representative designers will also provide instructions to the system-level designers for advancing their design advice. Within this design principle, we measured the extent to which the design outputs of the two designer groups resulted in new insights regarding and/or in refining the designer's own output. Ultimately, this will enhance the quality of design outputs.

We operationalise reflection and learning as:

- The extent to which the output of the other group of designers results in accentuation, adaptation or adjustment of the group's own output
- The extent to which a group of designers provides instructions and/or asks questions to the other group in reaction to their design output

9.1.4 Instantiation of design principle 4: decision-making in the sessions

Design principle 4 concerns the establishment of checks and balances among designers and guarantees the primacy of the public. The instantiation of the principle focused on decision-making regarding design choices. It structured the way in which the representative designers could influence the socio-technical specification, but also how system-level designers ensure a balance of power among themselves in making design choices on the specification.

To provide different system-level designers with equivalent opportunities to influence the socio-technical specification, we mostly focused on the second session of the simulation. System-level designers were instructed to strive for consensus in arriving at a preference for one of their alternatives. This was an instantiation of the first aggregation rule under design principle 4. In line with the second aggregation rule, we planned to ask the public servants to write an individual opinion at the end of session 2 (as part of step 5). In the actual simulation, there was not enough time to carry out this task. The participants in session 2 mostly agreed on the advice they were given, which also reduced the need to write individual opinions. However, the consensus on the preferred alternative might also be explained by the fact that the output of the simulation had no consequences in reality.

The political instruction described under design principle 3 was the main instantiation to ensure steering of design choices by representative designers. The primacy of the public was partly pursued by having the first session of the simulation with representative designers. Moreover, representative designers were asked to assess the design advice by system-level designers, whereas system-level designers were only asked to interpret the political instruction. The assessment by representative designers is an instantiation of the choice rule under design principle 4.

Measurement variables

Enabling making design choices

For this final design principle, we focused on studying the extent to which information provided by the system-level designers enables representative designers to decide which design alternative best fits their political instruction. If the prescribed system map works as expected, it conveys information about the socio-technical specification that enlarges understanding, provides motivation behind design choices, and communicates the meticulousness of design choices (see the instantiation of design principle 2). However, that does not say anything about whether the system map enables participants to make design choices concerning the socio-technical specification.

We focused our observations of designers making choices on:

- The extent to which the system map supports both representative as well as system-level designers in obtaining an overview of the system
- The extent to which the system map supports both representative as well as system-level designers in identifying critical design choices

Political steering

The political instruction aimed to give the politicians a tangible instrument to determine the substantive and normative direction of the design process. Moreover, the politicians could use the political instruction for assessing whether the design choices or socio-technical specification of the system-level designers stayed within that demarcated space. Together, giving direction to the design process and checking the system-level designers' output was considered political steering by representative designers.

We operationalise political steering as:

- On the representative designers' side
    - The extent to which politicians are able to formulate a political instruction in reaction to a problem formulation presented to them
    - The extent to which the political constellation of the city council is reflected in the political instruction
    - The perception of politicians on the extent to which the political instruction provides them with influence on the substantive and normative course of the design process

- On the system-level designers' side
  - The extent to which the political instruction is useful for system-level designers in making design choices
  - The extent to which the political instruction is recognisable in the design choices or the socio-technical specification
  - The extent to which the system-level designers refer to the political instruction when deliberating on design choices and the socio-technical specification
- Interaction between the two sides
  - The extent to which the dialectic interaction ensures the primacy of the public in determining the substantive and normative direction of design choices
  - The extent to which the political instruction supports politicians in assessing the work of the system-level designers

9.2 Evaluation results

This section discusses the outcomes of the simulation and what these outcomes mean for the evaluation of the design principles. Section 9.2.1 discusses the outcomes of the two surveys and the simulation. Section 9.2.2 uses these results to reflect on the four design principles.

9.2.1 Simulation results

This subsection discusses the results from both surveys and from the observations during the simulation. It first provides a concise overview of the answers given by three politicians and four public servants to questions on the current situation of designing in their organisation. Thereafter, the course of the simulation is discussed by pinpointing where participants adapted or diverted from the approach prescribed in the simulation. Finally, this section presents an overview of the answers given to the survey distributed after the simulation. The survey was filled in by the two politicians and the two public servants who participated in sessions 3 and 4. We also asked the other politician (P3) and public servant (S3) to fill in the post-test, but they did not respond to our request.

Current situation perceived by participants

When asked to characterise their interaction with executive designers of algorithmic systems in the current situation, the representative designers mostly pointed out the many formal and informal communication channels available. Nonetheless, they were critical about the quality of information flows between them and public servants. P2 and P3 stated that they have too little insight into the argumentation behind design choices. This impedes them from executing their 'controlling' and 'norm setting' roles. At the same time, they indicated issues at the side of the city council. While P3 considered the city council of this particular municipality as highly interested in and critical of algorithmic systems (across the full political spectrum), P2 stated that her colleagues often lack the necessary knowledge and competencies in order to ask the right questions, elicit relevant information, and participate in meaningful political debates. When asked for needed improvements in the current design process, P2 and P3 referred to the need for more financial resources to increase capacity within the public organisations and more support from public servants to the city council. P1 stated she would like to see a more positive attitude that also considers the affordances of algorithmic systems within the municipality.

System-level designers also referred to efforts related to information provision when characterising their interaction with politicians. Regarding the public servants, the main aim behind providing information is to share their considerations behind design choices with the city council, and, thereby, engage them in designing algorithmic systems. They do not refer to the need for legitimation of design choices. S3 considered using policy instruments as a means to bring consequences for citizens into view and to identify errors in algorithmic models. At the same time, the public servants referred to the distance between the council and the public servants in the design process. But when asked for improvements to the process, the system-level designers stated they want to prevent politicians from mingling in their own expertise. In addition, they stated that (organisational) structures to come to design choices or to make trade-offs can be more effective and efficient.

Course of the simulation

Session 1

The duration of this session was 1½ hours. The city council members started with a lengthy discussion on their views on the provided case. After 30 minutes, they started to write down a few design goals. When they could not think of other goals, they turned to design constraints. In total, they formulated four design goals and four design constraints. Moving to the prioritisation of goals and constraints, the interactions between the council members became more 'political'. Throughout the entire session, but especially at this point, the politicians checked with the instructor whether they were performing the tasks as intended. The politicians explained their own opinions to each other, whereas before they focused on brainstorming ideas together. The prioritisation did not take much time (~10 minutes). The politicians concluded that, without it being their intention, the goals and constraints already were positioned in the right order. There was only some debate whether the design goal on '*menselijke maat*' should precede the design goal related to the purpose of the system (see Appendix B4.1 for the final prioritisation). They decided to change the order and prioritise the '*menselijke maat*' goal. Likewise, writing down the clarification, which concluded the session, did not take much time (~5 min).

Session 2

The three-hour-long second session of system-level designers started with a critical reflection on the formulation and prioritisation of the goals and constraints by politicians. They especially did not understand why the politicians prioritised the '*menselijke maat*' design goal before the goal related to the purpose of the system (i.e., exactly the two goals explicitly deliberated by the politicians). At the same time, they started to work on the design task and formulated three alternatives in step 2: 1) actively provide allowances; 2) passively apply for allowances; and 3) a combination of the first two alternatives. Moving to step 3, the public servants had to make themselves familiar with using the system map. Due to time constraints, they only drafted system maps for alternatives 1 and 2. In step 4, the participants focused on scrutinising their own designs using the manifestations of arbitrary conduct. They concluded that all manifestations were discussed before and were mitigated in their designs. Closing the session in step 5, the system-level designers concluded that alternative three was the preferred system alternative. The design advice is presented in Appendix B4.2.

Session 3

In session 3, the politicians had 1½ hours to reflect on the design advice and their own political instruction. The politicians started by reading the advice individually before starting to discuss components in the system maps that were unclear to them. The discussion continued with pointing out the components they liked in the design alternatives and the components they missed. The city council members closed this part of the session by comparing the design alternatives with their own goals. They concluded that both the elaborated design alternatives (1 and 2) adhered to their goals, but they agreed with the system-level designers that a combination of the two alternatives would be the best option. Subsequently, the council members reflected on their own political instruction. The politicians concluded that their design goals and design constraints were clear and that they did not have to adapt their instruction. Instead, they mostly discussed what they were still missing in the design alternatives, without relating this to the formulation of their own goals and constraints. The third session ended with the council members drafting a response to the public servants. They formulated eight questions for the system-level designers (see Appendix B4.3).

Session 4

The one-hour session that concluded the simulation evolved into a discussion between the instructor and the two remaining system-level designers. This mostly happened because the system-level designers quickly came to the conclusion that the questions posed by the city council members provided no new information on the political instruction and that a reiteration of their design advice was not needed. Therefore, the instructor asked whether the questions posed by the representative designers were similar to questions typically asked by the city council. This resulted in a discussion on the role of the council and the questions that they ask. The system-level designers stated that, generally, politicians ask a lot of questions and that, in practice, public servants try to already answer most of the possible questions before they are even asked.

Reflection by participants

When asked to reflect on the simulation, the participating politicians referred to three affordances of the political instruction: 1) it supports in formulating requirements for algorithmic systems; 2) the structure of the instruction makes their own output clearer for other actors; and, 3) P1 mentioned how the political instruction helped them in assessing the design alternatives. Concerning the system map, the politicians were positive about the insights it provided into the system alternatives. At the same time, they indicated issues related to the explanation of the maps and incompleteness of the maps that need to be resolved (see Section 9.2.2). P1 and P2 had diverging opinions on whether the system map provided them insight into the consideration behind design choices made by designers. When asked for improvements to the design process in the simulation, the politicians mentioned similar issues as in the questionnaire preceding the simulation: they mostly would like to have more information about design choices made by the public servants.

Regarding the political instruction, the public servants stated that they considered it as steering their design activities. At the same time, the response of politicians to their design advice did not affect their ideas about their own design advice. Concerning the system map, S1 and S2 held diverging opinions. Where S1 considered it as only providing limited support in forming argumentation behind design choices, S2 emphasised how the maps provided more insight into the algorithmic systems.

9.2.2 Evaluation of the four design principles

This section evaluates the four design principles based on the simulation. For each design principle, we discuss the related measurement variables presented in Section 9.1. We focus on both the interactions within the two groups of designers and between the groups of designers.

Design principle 1

DP 1: *Establish an organisational level responsible for the socio-technical specification of public algorithmic systems in order to ensure mutual understanding about and the design of interactions between system components*

The evaluation of design principle 1 concentrates on the extent to which bringing designers from different disciplines together to formulate a socio-technical specification increased mutual understanding between system-level designers. When asked to reflect on the simulation, system-level designers S1 and S2 stated that the system map was conducive to communicating about design choices and that drafting the map provided them a new perspective on public algorithmic systems. We also observed how system-level designers used the system map to explain system components (and their implications) associated with their own discipline to one another. Thereby, they tested their own assumptions concerning the input of the other experts. For example, in session 2, the 'technology-oriented' designers (S1 and S3) asked the policymaker (S2) many questions. In response, the policymaker could explain all sorts of details concerning social welfare policy. Especially the policymaker stated how he learned from mapping the system: 'I found it a meaningful and valuable approach to address an issue. The shift towards algorithmically thinking was especially instructive [trans.].' The system map provided S2 more insight into the effects of different policy-related choices on the whole system: 'by mapping the algorithmic system, the consequences of different policy-related choices could clearly be demonstrated [trans.].'

Apart from instigating mutual understanding, working on the socio-technical specification through the system map also helped to resolve misunderstandings. While designing the second alternative, the designers were discussing how allowance requests would be checked. The policymaker seemed to think that checks would only be performed manually. S1 and S3 needed some time to explain that these checks can also be automated and that, considering the large number of expected applications, this would be the only feasible option. They were constantly referring to the specification when explaining the need for automation.

Design principle 2

DP 2: *Establish formal conditions to which the socio-technical specification of public algorithmic systems should adhere to, in order to standardise the motivation behind and meticulousness of design choices*

First, regarding design principle 2, we expected that establishing formal conditions (i.e., for the system map and for reflecting on the map) would strengthen the motivation behind design choices. The system map supported designers in developing argumentation in different ways. While drafting the system maps, the system-level designers started to think in scenarios that would emerge from different design choices. By scrutinising their design choices through scenarios, the designers contemplated the possible consequences of their design choices. Although we observed this 'scenario' approach among the group of system-level designers, S1 and S2 voiced diverging views on the extent to which the system map supported them in formulating the argumentation behind design choices in the final questionnaire. Where S2 recognised this affordance (see the results of design principle 1), S1 stated in the reflection that the design approach in the simulation provided him with limited support in providing argumentation behind design choices.

Although the system-level designers verbally discussed many arguments and considerations behind design choices, they only documented a few of their design choices and related argumentation in their design advice. They mostly wrote down design choices that they discussed at the end of the second session. Moreover, while they selected alternative 3 as their preferred alternative, this was the only alternative not elaborated in a system map. Consequently, the design choices for this alternative could only be implicitly derived from the documentation of the other two alternatives.

Although we only expected to support motivating design choices at the side of system-level designers, we observed that the political instruction supported politicians in making their choices. As one of them stated during session 1: 'the debate is held on a deeper level; the debate is focused more on the system in this way, and not only about whether a system can be discriminating or not [trans.].' It seems that asking politicians to formulate concise design goals and design constraints provides them with a structure to motivate the norms that they set.

Regarding meticulousness of making design choices, we observed how drafting the system map encouraged system-level designers to define the (interactions between) institutional and agential components in their system alternatives (see the system maps in Appendix B4.2). Nonetheless, we also remarked an inclination to elaborate technological components. These components have more detailed descriptions and denominations compared to the institutional components. Moreover, the designers started with the technological components when creating the system maps. Accordingly, we observed a higher number of technological artefacts in the maps. However, this can also be explained by the fact that we asked the public servants to come up with at least two algorithmic systems.

Using the manifestations of arbitrary conduct, the system-level designers were reflecting on the consequences for citizens of their system alternatives. Although the designers concluded that their alternatives did not include any of the manifestations, the discussion instigated by the manifestations was important in making a decision on the preferred alternative. Of interest here is the confidence of the system-level designers in their own design choices. They showed high trust in the objection procedures to filter out any mistakes in the system. They included these procedures mostly because it is a legal requirement for policies and systems that lead to this type of decision-making. However, they did not fully elaborate on what these objection procedures would look like and how they would be affected by automated and algorithmic decision-making. Considering the problems leading to Kafkaesque situations (see Chapter 6), the specification of such objection procedures needs more detail.

Design principle 3

DP 3: *Establish a structured dialectic between representative designers and system-level designers via their design outputs in order to stimulate learning among and reflection by both groups of designers*

Regarding the interactions between representative designers, we observed that the participating politicians obtained a better understanding of the algorithmic systems through the system maps. When reflecting on the simulation, both politicians saw potential in how the system maps provide insight into the workings of the system. They also provided suggestions for improving the information provision through the system map: let public servants give a verbal explanation of the maps, ensure that the maps are complete, and make the links between the political instruction and the system maps explicit.

Notwithstanding, exchange of the political instruction and the broader advice showed that mutual understanding between the two groups of designers was not fully achieved. The outputs, both the political instruction as well as the design advice, seemed to be too brief to provide meaningful insight for the other group. Several times, participants stated that they did not understand the output of the other group, or that they would like more information or explanation about the outputs. Exemplary are the considerations behind design choices. Although the design approach in the simulation supported system-level designers in discussing their considerations, the politicians mentioned their need for more information and explanation on these considerations in the questionnaire after the simulation.

Participants seemed to associate the unclarity in the outputs with a lack of information provision. They also raised their concerns regarding the lack of information provision in the current situation in the survey that preceded the simulation. At the same time, we observed that participants were not always aware of information provided to them in preceding sessions. This raises the question of whether the need for information can ever be satisfied.

Most of the time, the participants were able to arrive at the interpretation intended by the other group through deliberation despite the perceived lack of information. They did this based on past experience and by learning from each other's expertise. For example, P2 shared her knowledge on the topic of energy allowances with the other politicians. She also stated at the end of session 3: 'I find it important that every council member can bring in and use his or her own expertise and knowledge, which is actually happening in this approach [trans.].' As such, the politicians together learned more about algorithmic systems.

While we observed learning among participants, reflection on their own design output was missing from both groups of designers. The system-level designers showed some reflection while elaborating on the earlier-mentioned scenarios. But in response to the manifestations of arbitrary conduct, the system-level designers stated that they were convinced about their own design choices. In session 4, one of the public servants, when asked whether they would like to change elements in their designs, stated: 'we made a conscious choice for alternative 3, so I do not really see a reason to adapt that [trans.].' As such, session 4 did not stimulate reflection among the system-level designers. Similarly, the representative designers did not adapt their goals and constraints after they were confronted with the alternatives that followed from their political instruction.

The lack of reflection on their own output in sessions 3 and 4 can have different explanations. First, both the political as well as the system-level designers stated that their outputs did not need changes after the first draft. However, this would be surprising considering the co-evolutionary nature of design. Second, we observed that groups have stereotypes of one another. Participants made statements about the shortcomings of the other groups in reality. This might be a reason to mostly look for issues in the output of the other group. In session 3 and 4, we mostly observed participants reacting to the output of the others and pointing out what was missing or wrong. Third, the design outputs in the simulation did not have any consequences. As the advice would never be implemented in reality, the participants could afford to not fully think through the consequences of design choices. Similarly, the time span of the simulation was short compared to the design processes in reality. More time to spend on the activities could give participants the space to critically assess their own work.

Design principle 4

DP 4: *Establish checks and balances between system-level designers, and between representative designers and system-level designers in order to reduce arbitrary use of designerly power*

The evaluation of the first three design principles already touched upon the extent to which the design principles enable designers in making design choices. The public servants used the system map to discuss the implications of design choices and based their choices on that discussion. Moreover, the manifestations of arbitrary conduct enabled the system-level designers to make a choice for a preferred design alternative. Politicians' insight into the alternative algorithmic systems was increased, but they stated that this does not enable them to make a decision. Especially, the lack of insight into the considerations of designers was mentioned as an issue in this respect. P2 stated that she did not feel the confidence to make a decision based on the information provided throughout the simulation.

Notwithstanding, the design advice presented to the politicians provided them with a fundamental choice concerning the role of citizens in the alternative algorithmic systems: whether or not to put full responsibility on citizens to request the allowance. Instead of acknowledging this fundamental choice on the role of citizens, the politicians requested more communication channels for citizens in the system alternatives. The system-level designers did not think this was viable because of efficiency reasons. Accordingly, the deliberation was shifting from the role of citizens to the consequences for the public organisation itself.

Moving to the measurement variable 'political steering', we observed that the participating politicians perceived an increase in their influence on the design of algorithmic systems. As mentioned before, the political instruction supported council members in formulating their requirements and clarifying boundaries for the system to be designed. Moreover, the politicians recognised their own goals and constraints in the design alternatives drafted by the system-level designers. Especially session 1 provided the political debate between council members with more structure compared to real practice. At one point P2 stated: 'this is an efficient way of debating, maybe we should do it like this more often [trans.]'. On the other hand, the participating politicians easily reached consensus on the prioritisation of goals and constraints in the simulation. At the end of session 1, they reflected on how the debate would be tougher when council members from other political parties would participate.

The system-level designers actively used the political instruction in both sessions. S1 stated: 'I experienced the goals and constraints as a directional framework.' Apart from using the instruction, the public servants were quite critical on the substance of the design goals. Unaware of the considerations of the politicians to prioritise the goal of '*menselijke maat'* (i.e., guaranteeing the well-being of citizens, and considering harmful algorithmic systems in the past), the system-level designers often expressed that they thought it was strange to prioritise this goal over the goal related to 'effectivity'.

In general, the two groups of designers reflected on how the simulation changed the dynamic between politicians and public servants several times. The simulation instigated a discussion among the participating politicians on the city council's role: whether they only have a 'controlling' function or also a 'norm-setting' function. The public servants mostly expressed their worries that politicians would mingle too much with their tasks and expertise, and that they might slow down the design process. Moreover, the system-level designers did not think that the interaction between politicians and public servants in the simulation aligns with the constitutional positions of these actors. They want to preserve a strict distinction between politicians prescribing the 'what' and public servants determining the 'how'. The system-level designers were also uncomfortable with giving politicians instructions. In reality, they would also not ask questions to politicians. Here, the Alderman can have a mediating function between the council and public servants.

9.3 Conclusion

This chapter evaluated the design principles formulated in Chapter 8 by testing institutional interventions based on those principles in the simulation of a design process in a Dutch municipality. In this simulation, politicians and public servants were asked to work on a fictional design assignment. The simulation was preceded by a questionnaire to examine the current design practices in the participating municipality and followed by a questionnaire in which participants reflected on the simulation. This concluding section reflects on the extent to which the interventions resulted in expected interactions. Section 10.1.4 will present the needed changes to refine the design principles that follow from the evaluation in this chapter.

The first two principles mostly concern the problem of coordination in formulating a socio-technical specification. The simulation showed how establishing the positions of system-level designers who produce that specification enlarged mutual understanding about, motivation of and meticulousness of design choices. The specification encouraged participants to align different system components; and although the considerations behind design choices need more elaboration, the participants stated they would use the specification in real practice. The establishment of a system-level designer position does not conflict with current practices in public organisations, which might make implementation uncomplicated. In contrast, the last two principles challenge the status quo of practices in public organisations.

Participants were reluctant or sceptical about the dialectic between political and system-level designers in order to guarantee the legitimacy of design choices. Our observations showed that the design process in the simulation did engender learning but did not stimulate significant reflection on one's own design output (designers mostly criticised the output of the other group of designers). Moreover, participants lacked information to make design choices, and they questioned whether the political steering they experienced in the simulation could be transferred to reality (the seem to be resistant to changing the existing interaction).

The evaluation presented in this chapter was explorative in nature. This has implications for the validity of its outcomes. First, although we simulated a real design process, the interactions between designers will be different in reality. For example, the design choices in this simulation were not consequential, and participants stated that they would have more time to reflect on design choices in reality. Second, the evaluation was performed at only one public organisation. For a design theory in progress, this is not a problem because the design principles can be refined based on the findings of this exploratory evaluation. But to improve the design theory, the simulation should be scaled up and performed at several public organisations. Third, the simulation performed was susceptible to response, researcher, and other forms of biases (see Section 3.3.3); partly because of the artificial situation that was created. Eventually, the design principles should be tested in several case studies of real design processes. Finally, we did not test the interaction between system-level designers and component designers in the simulation. Especially for ensuring alignment in the socio-technical specification, this interaction is of importance.

The evaluation's validity influences its generalisability. The simulation is used to iterate the design principles to make them more generalisable (see Section 10.1.4). At best, the principles can be generalised to other Dutch municipalities, since they have similar organisational structures. However, the scope of the design theory goes beyond Dutch local governments. Therefore, to enhance the generalisability of the principles, they should be tested in other public organisations. In principle, the simulation can be performed at every public organisation that designs public algorithmic systems with consequences for citizens. Here, we should note that the simulation is based on the Dutch context and that the instantiations are embedded in current institutional practices in Dutch public organisations. To perform the simulation in other (national) contexts, the principles should be translated into instantiations contingent on the simulation's context.

Chapter 10

Conclusions and discussion: curbing algorithmic Kafka

The research presented in this thesis was motivated by the role of design practices in the emergence of harmful public algorithmic systems. Public algorithmic systems can put citizens in Kafkaesque situations, and inflict physical, emotional, mental, or material harm on citizens. Algorithmic Kafka emerges from possibilities for arbitrary conduct in the constitution of algorithmic systems. Following from the fundamental principles underlying democracy and the Rule of Law, governments should protect citizens from any arbitrary use of power by public organisations. Nonetheless, current design practices of public algorithmic systems do not provide this protection and are lacking democratic legitimacy.

Considering this failing protection and lack of democratic legitimacy of the design process of public algorithmic systems, this research aimed to *create design practices that reduce the emergence of Kafkaesque situations in public algorithmic systems by stimulating designers to reflect on and respond to the consequences of their design choices.* This thesis focused on institutions that structure these practices and on institutional interventions that reshape practices in such a way that they adhere to democratic and Rule of Law presuppositions. The research was divided in three parts related to three research questions: diagnosis, appraisal, and create and assess. Section 10.1 presents the conclusion for each of the three research parts and reflects on the design theory. Thereafter, Section 10.2 elaborates both the scientific as well as the societal contributions of this thesis. Section 10.3 discusses the limitations of our research. In Section 10.4, we identify possibilities for future research and list policy recommendations.

10.1 Revisiting the research questions

This section provides answers to the three research questions formulated in Chapter 1. After answering the questions, we will reflect on the main deliverable of this research: the design theory.

10.1.1 Diagnosis

The first research part focused on identifying current design practices for public algorithmic systems in public organisations by conducting qualitative empirical research. In addition, we scrutinised institutional interventions that public organisations are currently implementing to adapt their design practices. Part I answered the following research question:

Research question 1
What presuppositions underlie the design practices for public algorithmic systems that have emerged in public organisations?

This research studied current design practices through conducting interviews with designers and policymakers, observing meetings in which policy instruments were developed, and by analysing documents that described current design processes and the policy instruments. Using the IAD framework by Ostrom (2005), we identified design practices in two action situations: the collaboration between designers with different disciplinary backgrounds and the political steering of design choices. Focusing on the interactions between designers, and the institutions and attitudes that structure these interactions, we identified four impediments to the design practices in addressing emerging and harmful Kafkaesque situations in algorithmic systems. First, the design choices made by designers are disconnected from meaningful public debate. Second, design activities get bogged down in coordinative and deliberative interactions between designers. Third, a procurement-oriented dynamic between designers created an implicit hierarchy between designers that hinders systemic design of algorithmic systems. Fourth, conflicting work rhythms of the different designers involved result in slack in the design process.

In general, the institutions and attitudes that structure the design practices can be aggregated into two presuppositions: a technocratic and a businesslike presupposition. The technocratic presupposition is reflected in the coordinating role of technical experts and the disconnection between public servants and politicians involved in the design process. The businesslike presupposition instigated the integration of strategies and practices from the private sector, with the associated jargon of agile, scrum, and client-supplier relationships, into design practices of public algorithmic systems. These two presuppositions are also recognisable in public algorithmic systems that are currently designed and deployed.

Public organisations acknowledge that their current algorithmic practices are insufficiently addressing algorithmic risk or harms. Therefore, they are developing their own institutional interventions to adapt their practices: policy instruments. This research assessed four of these policy instruments. The instruments fill an institutional void in design processes of public algorithmic systems by establishing information flows between designers and assigning responsibilities within public organisations. At the same time, the instruments are still based on the two presuppositions prevalent in current design practices. Consequently, they fall short in preventing, mitigating, or correcting Kafkaesque situations in public algorithmic systems.

10.1.2 Appraisal

Part II of the research focused on outlining desired design practices. Using abductive reasoning to study literature describing cases of algorithmic Kafka and literature on political philosophy and philosophy of law, we derived the theoretical basis for design practices that are embedded in current democratic and Rule of Law contexts. We answered the following research question:

Research question 2

What design practices that curb algorithmic Kafka are prescribed by the synthesis of the presuppositions of socio-technical designing, democracy, and the Rule of Law?

We performed two abductive analyses to answer this research question. In the first analysis, we focused on sense-making concerning the origins of algorithmic Kafka in designing public algorithmic systems. We used the perspectives of Kafkaesque situations and of arbitrary conduct to enrich our understanding of well-studied cases of harmful public algorithmic systems. The second abductive analysis concentrated on synthesising three contested and disparate concepts. The synthesis of the presuppositions of socio-technical designing, democracy, and the Rule of Law resulted in a meta-theory that forms the basis for the design principles in our design theory.

Based on the analysis of the childcare allowances scandal, the Robodebt scheme, the Post Office scandal, and the DUO case, this thesis defines algorithmic harms as a loss – i.e., of material, physical, emotional, and/or mental nature – suffered by a citizen who is exposed to the output of an algorithmic system. Such harms are inflicted on citizens who find themselves in Kafkaesque situations created by public algorithmic systems. These Kafkaesque situations emerge in public algorithmic systems when possibilities for arbitrary use of power are materialised. This means that reasoning within an algorithmic system is based on own will or pleasure, that affected citizens have no space or means to engage in or contest algorithmic decision-making, that the systems are unpredictable and/or incomprehensible, or that the system makes unfair decisions in concrete situations. Designers can bring possibilities for arbitrary conduct into the socio-technical specification of algorithmic systems in two ways. First, designers can use their own (designerly) power arbitrarily, which reduces the legitimacy of design choices and the specification based on those choices. Second, problems in coordination between designers can result in misaligned technical and institutional artefacts in algorithmic systems. Such misalignment can result in the emergence of arbitrary conduct.

Arbitrary use of power can be reduced by following principles of the Rule of Law – which main aim is to reduce arbitrary conduct – and democracy – providing legitimacy to decision-making. Chapter 7 explored how these two presuppositions from political philosophy and philosophy of law can be aligned with the socio-technical characteristics of design processes of public algorithmic systems. Juxtaposing the three presuppositions resulted in four symbioses, three shared challenges, and three contradictions. The symbioses show that the three concepts together can instigate iterative design practices in which political struggle is structured by checks and balances between reflective and responsive actors. The three presuppositions share challenges that complicate navigating design practices. The design practices are situated in a polycentric network of action situations that are dependent on their context. Moreover, designers need to have a favourable attitude (i.e., act in accordance with democratic and Rule of Law principles) in order to arrive at the design practices that follow from the symbioses. Finally, the contradictions pose trade-offs that the design theory has to deal with: whether it will put emphasis on public over experts or the other way around, whether it will focus on protection of minorities over majority decision-making, and whether the number of fixed procedures can be minimised to engender flexibility. The juxtaposition of the three presuppositions resulted in a meta-theory. This meta-theory forms the basis for design principles that prescribe institutional interventions to arrive at design practices that are opposite to those based on technocratic and businesslike presuppositions. Embedding the socio-technical design process of algorithmic systems in a democratic and Rule of Law context mostly means that inquiry, iteration, questioning of design choices, and explicitly providing room for public debate should be ensured in a design process in which the power distribution among different designers is balanced.

10.1.3 Create and assess

The first part of this research provided insight into current design practices, and the second part identified desired design practices. The final part of the research formulated design principles for institutional interventions that transform current practices into desired practices. Part III answered the following research question:

Research question 3
What institutional interventions engender interactions between designers of public algorithmic systems that align with democratic and Rule of Law principles?

We formulated a design theory through generate-test cycles. Using the insights from the first two parts of the thesis, we elicited the gap between current and desired interactions between designers. Using the meta-theory presented in Chapter 7 as starting point, we formulated design principles that can support public organisations in arriving at these desired interactions. We tested instantiations of the principles – i.e., specific institutional interventions in the form of instructions for a design process – in a simulation of a design process of a public algorithmic system. Based on the test of these institutional interventions in the simulation, we performed an explorative evaluation of the design theory.

The main idea behind the design theory is to establish the position of system-level designers who, collaboratively, formulate the socio-technical specification of a public algorithmic system. Accordingly, the system-level designers can coordinate the creation of the different system components, inform the political debate with insights into systemic design choices, and interact with representative designers. The design theory consists of four design principles that prescribe formal institutional interventions to reshape design practices:

DP 1: *Establish an organisational level responsible for the socio-technical specification of public algorithmic systems in order to ensure mutual understanding about and the design of interactions between system components*
DP 2: *Establish formal conditions to which the socio-technical specification of public algorithmic systems should adhere to, in order to standardise the motivation behind and meticulousness of design choices*
DP 3: *Establish a structured dialectic between representative designers and system-level designers via their design outputs in order to stimulate learning among and reflection by both groups of designers*
DP 4: *Establish checks and balances between system-level designers, and between representative designers and system-level designers in order to reduce arbitrary use of designerly power*

The goals set in the design principles can be achieved by creating institutions for each individual variable in an action situation (as defined in the IAD framework). We created instantiations of the design principles for a Dutch municipality and tested whether and to what extent institutional interventions based on the design principles resulted in the expected interactions. In a simulation of a design process, we first asked politicians to formulate a political instruction – i.e., design goals and design constraints. Thereafter, system-level designers drafted socio-technical specifications for three design alternatives of public algorithmic systems. Two additional sessions followed in which the politicians and system-level designers reflected on their own design output and responded to each other's design output. The evaluation showed that a system-level design team that drafts a socio-technical specification in the form of a system map (i.e., an overview of all technological, institutional, and agential components of the public algorithmic system and the interactions between these components) increases mutual understanding and stimulates learning among system-level and representative designers. As such, drafting a socio-technical specification stimulates coordination between executive designers from different disciplines. At the same time, the system map does not provide enough information for politicians to assess the motivations behind and meticulousness of design choices. Finally, participants felt uncomfortable with the dialectic instituted between politicians and system-level designers, which aimed to increase the legitimacy of design choices. Especially system-level designers are afraid that their expertise will be undermined when politicians are assessing their work. This suggests that the technocratic presupposition is hindering a responsive attitude, which is needed in such a dialectic, among designers.

10.1.4 Design theory

This thesis presents a provisional instance of the design theory, i.e., design principles embedded in a meta-theory. More iterations of the theory are needed and, considering the explorative evaluation, should be tested more extensively. This section reflects on the affordances and shortcomings of the current theory. Building on that reflection, we explore directions for advancing the design principles in the design theory.

To reflect on the design theory, we return to the requirements for institutional interventions formulated in Chapter 5. In current institutional interventions within public organisations, we identified a need for establishing structural information flows among designers and politicians about algorithmic systems, demarcating roles and responsibilities, and raising awareness on algorithmic risks among designers. The need for structural information flows was also observed in the explorative evaluation of the design principles. Within the design theory, the socio-technical specification creates a medium for representative and system-level designers to communicate design choices. At the same time, the simulation shows that the need for information, especially at the side of representative designers, seems to be insatiable. Second, the design theory demarcates the roles and responsibilities of representative and system-level designers. In next iterations, the roles and responsibilities of component designers (i.e., designers that elaborate the design of individual system components) should be added to the design theory. Third, the requirement of raising awareness on algorithmic risks among public servants is satisfied indirectly by stimulating designers to actively search for manifestations of arbitrary conduct in designs of algorithmic systems. Nonetheless, the assessment of these manifestations in socio-technical specifications can be strengthened and made more profound.

We derived four additional requirements from gaps in current institutional interventions in Chapter 5. First, reshaping design practices requires a mix of institutional interventions. Our design theory prescribes such a mix, as it is grounded in the IAD framework and related perspectives on institutions. Second, the interventions should aim to increase democratic legitimacy of design practices by specifying the role of citizens and political debate. At this point, the design theory assumes that this political debate is happening through representative bodies and that these bodies adequately represent citizens. Third, the institutional interventions should not be presented as a panacea. That is why the design theory only anticipates a reduction of arbitrary conduct and does not claim that such conduct can fully be prevented. The final requirement is that the set of institutional interventions should have an unambiguous goal. The aim of the design theory is to embed the design process of public algorithmic systems in a democratic and Rule of Law context to mitigate, prevent, and correct algorithmic Kafka. The meta-theory specifies what this goal entails.

Accordingly, we can assess the extent to which the design theory brings public organisations closer to this goal of institutionalising design practices that curb algorithmic Kafka. We specified this goal to two sub goals for interactions between designers: increase coordination between designers, and guarantee legitimation of design choices. Regarding the conclusion drawn in Section 10.1.3, we can derive that the first sub goal of coordination has been achieved by establishing a system-level design team. Notwithstanding, there is potential to expand the design theory with design principles on the role of component designers.

Considering the legitimation sub goal, the design principles require further refinement. The institutional interventions on reducing arbitrary designerly power mostly focus on defining the role of politicians in the design process. The formulation of design goals and constraints by politicians was well received by both groups. Nonetheless, the public servants were reluctant to extrapolate the dialectic proposed in the design theory to real design practices. The proposed dialectic in the design theory seems to be too big a move away from the current interaction between politicians and public servants. In the simulation, public servants stated that politicians should stay away from determining the means to achieve an end or address a problem. It seems that the public servants considered a response by the politicians on their system maps as interfering with their own responsibilities. Still, it is unlikely that politicians will actually interfere with public servants' tasks, as they stated in the simulation that at the moment their capacity and resources are already too limited to have a meaningful debate on algorithmic systems.

10.2 Contributions

This section discusses both the scientific and the societal contributions of this thesis.

10.2.1 Scientific contributions

Our research contributes to several scientific debates. This section discusses our contribution to the conceptualisation of algorithmic harms, to the conceptualisation of arbitrary conduct in algorithmic systems, to empirical insight into design practices of algorithmic systems, to the substantiation of the role of institutional interventions in governing algorithmic systems, to positioning democracy and Rule of Law in socio-technical systems literature, and to insights on the lack of democratic legitimacy in design processes of public algorithmic systems. We will elaborate on these six scientific contributions.

Shifting the focus to causal and constitutive factors leading to algorithmic harm inflicted upon citizens

In recent years, attention for algorithmic harms caused by algorithmic systems has increased. This thesis contributes to that research by elaborating the citizen perspective on algorithmic harms and by shifting the focus to the role of designers in creating these harms. Dencik et al. (2019) already argued how the lack of a citizen's perspective can dehumanise citizens and produce a golden view on citizens. We emphasised and demarcated the citizen perspective on algorithmic harms by refining the perspective on algorithmic systems as producers of Kafkaesque situations. The Kafka metaphor is an important perspective on the position of individual citizens in anonymous bureaucratic and (socio-)technological machines. We contributed to this perspective by identifying the elements of Kafkaesque situations created by public algorithmic systems and how these elements are linked to possibilities for arbitrary conduct.

In addition, the research contributes to the knowledge on causes of algorithmic harms. As discussed before, causes and consequences of algorithmic harms are often conflated (Hutiri, 2023). This hinders addressing the harms' causes in the design process and the role of designers therein. We argue that manifestations of arbitrary conduct cover a wide range of causes of algorithmic harms. The concept of arbitrary conduct also provides a way to appoint accountability to either individual agents or socio-technical assemblages that use their power arbitrarily.

Conceptualising a socio-technical perspective on arbitrary conduct

This thesis positioned public algorithmic systems as socio-technical systems that have the potential to mediate arbitrary use of power and consequently instigate Kafkaesque situations. Considering manifestations of arbitrary conduct as an emergent property of a socio-technical system allows for a sociomaterial understanding of such conduct (as the sociomaterial understanding argues that there is no social, i.e., here arbitrary conduct, without the material (Orlikowski, 2007)). However, debates in legal philosophy on reducing arbitrary use of power generally disregard the role of technology in the emergence of such conduct (Nouws & Dobbe, 2024). Bringing in the socio-technical perspective on arbitrary conduct contributes to the need for theorising the nature of arbitrariness (see Krygier, 2016; Mak & Taekema, 2016) by explicating the (relational) role of technology in arbitrary conduct. At the same time, a socio-technical perspective on arbitrary conduct in algorithmic systems incorporates recent calls for considering the infrastructural characteristics of automation and algorithmic systems (Widlak & Peeters, 2025). This means that arbitrary conduct cannot always be attributed to one actor or a specific organisation, but might be distributed within a digital ecosystem (Mueller, 2025) or over a supply chain (Cobbe et al., 2023).

This thesis identified how arbitrary conduct can emerge from the interactions between socio-technical components in public algorithmic systems. The focus on interactions between system components clarifies the role of designing and designers in the emergence of arbitrary algorithmic systems. We elaborated the consequences of arbitrary conduct by designers themselves and confirmed the importance of alignment between institutional and technological components (described by, for example, Peeters & Widlak (2018) and Künneke et al. (2021)) in the specific case of public algorithmic systems. Moreover, by introducing arbitrary conduct as an emergent property, we also position this concept as a bridge for policy-oriented and technical-oriented perspectives on public algorithmic systems that are often divided (Vydra &

Klievink, 2019). Using the concept of arbitrary conduct allows for relating algorithmic systems to notions of democracy and the Rule of Law. As such, the concept forms a bridge between socio-technical literature and literature from political philosophy and philosophy of law.

Providing more granular insight into current design practices of public algorithmic systems

This thesis' introduction identified a lack of empirical understanding of current design practices of public algorithmic systems. Accordingly, to complement the already extensive descriptive insights into the use, implementation, execution, and operation of algorithmic systems, we focused our empirical research on design practices. By focusing on the interactions between designers, this research provided insight into presuppositions structuring the design practices. These findings indicate that design practices are still based on technocratic and businesslike presuppositions that have been discarded in scientific literature (e.g., Bostrom & Heinen, 1977).

In addition, this thesis demonstrates the suitability of the IAD framework by Ostrom (2005) for studying design practices. By using the IAD framework, the design process is considered a collective action that occurs in several different but interrelated action situations. The IAD framework enables researchers of design processes to distinguish the polycentric network in which design choices are made and supports in identifying structural factors shaping the arena in which these choices are made. Moreover, the framework provides a basis for understanding changes within design practices through learning but also how deliberate interventions into the structural factors of design practices can be used to arrive at desirable outcomes of design practices. The explorative evaluation of our design principles already showed how participating designers were changing their approaches towards designing public algorithmic systems. Although the applicability of Ostrom's thinking for governing AI has already been stressed (Grimmelikhuijsen & Meijer, 2022), we are the first (as far as we know) to use the IAD framework for studying algorithmic practices.

Positioning institutional interventions as a way to govern the direction of algorithmic practices

The governance of AI and other algorithmic systems generally focuses on the use of algorithmic systems. This thesis broadens the perspective on AI governance to design practices. Here, the research's main contribution is a design theory encompassing a comprehensive set of institutional interventions to reshape design practices. Considering governance from an institutional perspective ensures that both formal and informal institutions are considered. In addition, the institutional perspective broadens the view to underlying presuppositions, which relates to the literature on institutional logics that already discusses the role of (shared) ideas in the formation of institutions (Thornton & Ocasio, 2008). This thesis shows that, in order to change design practice, both formal and informal institutions as well as institutional logics within public organisations need to be altered. Institutional theory literature provides many insights on how institutions change both organically as well as by design (Klijn & Koppenjan, 2006). This has implications for future algorithmic practices. In order to design interventions in these practices, and consequently design AI governance, current practices must be studied and unravelled first.

Another advantage of considering AI governance from a practice perspective is that it moves away from prevalent understandings of human-in-the-loop. Where this type of AI governance presupposes that actors within public organisation need to develop new skills and competences, our design theory starts from ensuring that all needed skills, competences, and disciplines are represented in a system-level design team. Instead of putting all responsibility on one designated actor, this system-level design team should ensure a way of coordinating all relevant input needed about the algorithmic systems. This starts with the socio-technical specification as shared language.

Applying democracy and the Rule of Law to socio-technical designing

This thesis presents the first synthesis of democracy, Rule of Law, and socio-technical designing. Our synthesis shows that, although dissimilar, the concepts share a considerable number of elements which can be used to firmly embed socio-technical design practices into the context of public organisations. Moreover, it is a fruitful perspective on positioning public organisations as designers of algorithmic systems. Literature on democracy and the Rule of Law provide a wealth of knowledge on how to organise government. By combining this with socio-technical theory on designing systems, the well-established democratic and Rule of Law practices can be applied to the new role of public organisations as designers of socio-technical systems. Therefore, the democratising wheel does not need to be reinvented when restructuring current design practices.

As such, this thesis contributes to scientific debates on democratising design processes (of AI) and the position of AI in the Rule of Law. As mentioned before, most studies concerning democratising AI do not relate to concepts from political philosophy (Himmelreich, 2019), resulting in superficial and popular interpretations of democracy (Himmelreich, 2023). Our pragmatic approach enabled us to translate Dewey's philosophical interpretation of democracy to the design process of public algorithmic systems. At the same time, we move away from prescribing ideal types of democratic design processes such as Participatory Design, or Value-Sensitive Design. We changed the perspective to enhancing design practices that are already present in public organisations. Regarding AI and the Rule of Law, this thesis provides ways to use the Rule of Law for formal requirements, procedural constraints, and substantive rights that can guide the design (both as verb as well as noun) of public algorithmic systems. In addition, the position of socio-technical designers turns out to support democratic debates on the form and function of public algorithmic systems.

Shifting the perspective on the lack of democratic legitimacy in design processes of algorithmic systems

Finally, this research confirms hunches in literature regarding the undermining of democracy in current design processes of public algorithmic systems (Mulligan & Bamberger, 2018, 2019; Van Zoonen, 2020). The distance between the political debate and the design process observed in this thesis is especially exemplary for the lack of democratic legitimacy. Van Zoonen (2020) and Oldenhof et al. (2024) explain this lack with the concept of institutional void. However, based on this thesis, we challenge this explanation. Institutions to ensure democracy and Rule of Law values in design processes of algorithmic systems and other artefacts public administration are already in place; but these institutions should also be upheld by the human agents working within these contexts (see for example Krygier, 2009; Selznick, 1999). This means that these human agents should have an attitude that supports institutions that structure reflectivity and responsiveness in design practices.

Accordingly, it might be more suited to explain the lack of democratic legitimacy as a consequence of *institutional neglect*. Public organisations show little initiative to reflect on their own role as socio-technical designer and how their (design) practices need to be transformed in such a way that they still adhere to democratic and Rule of Law values. This might ask for both structural changes in public organisations but also shifts in attitudes and competences among public servants. Similarly, in discussions on contestability of algorithmic systems authors emphasise the need for a reflective and responsive attitude among designers of algorithmic systems (Alfrink et al., 2022; R. Dobbe et al., 2021; Hildebrandt, 2018). This study shows that countering the institutional neglect includes considering the competency of design as part of the professionalisation of public servants.

10.2.2 Societal contributions

Our research confirms that public organisations struggle to protect citizens from harms inflicted by their algorithmic systems. Recently deployed systems put citizens into Kafkaesque situations, which can partly be explained by issues in current design practices in public organisations. Meanwhile, pressure on public organisations to strengthen their grip on public algorithmic systems is intensified – e.g., by the introduction of laws and regulations, such as the AI act. This section presents contributions of our research that can support public organisations in addressing this challenge.

Discerning design practices in public organisations
Section 10.2.1 discussed the insight that public organisations already perform design activities to conceive their algorithmic systems. Nevertheless, while conducting our research in public organisations and communicating our findings, we often had to explicate what activities of public servants and politicians can be understood as design efforts. Being unaware of their own role as designer might hinder public organisations and public servants in acquiring or developing necessary design competencies. The design practices identified in this research can contribute to raising awareness about or reflecting on design activities within public organisations. Increased insight in design practices might open possibilities for adapting design practices, for which this research provides practicable design principles that public organisations can translate to institutional interventions applicable to their own context.

Preventing, mitigating, and correcting Kafkaesque situations
Eliminating Kafkaesque situations should be a priority for public organisations. This responsibility originates from the fundamental principles underlying democracy and the Rule of Law: governments should protect citizens from their own actions since the inherent power imbalance between citizens and governments creates room for arbitrary use of power. Where scientific literature has identified problems related to algorithmic systems, this study also provides public organisations with points of intervention in these algorithmic systems. Public organisations can focus on the alignment of institutional and technological artefacts to reduce possibilities for arbitrary conduct in their systems. Moreover, the perspective of alignment emphasises that designers of both institutional and technical artefacts should collaborate and that the input of both types of designers is equally important. Accordingly, this research contributes to the reduction of the creation of harmful public algorithmic systems and to the protection of citizens against algorithmic Kafka.

Contribution to public debate
Finally, this research makes a contribution to public debates on public algorithmic systems. First, it introduces arbitrary conduct as a meaningful concept to debate issues in algorithmic systems. Arbitrary use of power is a familiar concept to public organisations, and it emphasises that the use of algorithmic systems is a continuation of a longer trend in government action. Moreover, the concept of arbitrary conduct enables actors from different disciplinary backgrounds to deliberate on complex socio-technical systems; it creates a shared language. Second, the instruments discussed and examined in this thesis are still used, elaborated, and implemented by public organisations. Our findings can be used to fuel the debate in which these policy instruments are assessed, scrutinised, and improved.

10.3 Limitations

The limitations of this research can be divided into four broad themes. Two themes relate to the adopted research approach: the emphasis on interviews (performed online) for qualitative empirical research and the generalisability of the findings of the three individual parts of this research. The other two themes relate to the scope of the study: our focus on institutional interventions for design practices and delivering a design theory that is still in development.

Research approach – emphasis on online interviews
Our empirical findings are mostly based on interviews. Ethnographers debate whether practices, such as the design practices in this thesis, can be elicited through interviews. This research followed Seaver (2017) who argues that, when using appropriate techniques, interviews can result in rich insights into practices. Still, observations of designers working on public algorithmic systems would have enabled us to compare statements of designers with reality. However, observations of design practices are time-intensive; time that was not available in a design science project to be finished in a four-year PhD trajectory. Therefore, we have corroborated our findings by performing consecutive exploratory and explanatory empirical studies. Moreover, we organised workshops in which designers of public algorithmic systems could react and reflect on our preliminary findings.

Using interviews as research method also brings possibilities for researcher bias. Interviewers ask questions based on their own knowledge and interests. The answers by interviewees are always a reaction to the interviewer's perceptions of a practice. We have tried to mitigate the researcher bias by asking designers of public algorithmic systems to chronologically discuss a design project in which they were involved (see Section 3.3.1).

In addition, we performed our interviews online. On the one hand, performing interviews online is practical and advantageous. Practitioners seemed to be more willing to participate in an online interview. This might be explained by an online meeting being less of an effort compared to a physical meeting or it being easier to plan in packed agendas. The advantages on the side of the interviewer are the fact that interviewees are not distracted by what you do or write as an interviewer. Moreover, working from my own workspace, I could easily access all relevant documents and protocols for the interview.

Nevertheless, online meetings have a dynamic that is different from physical interviews. The screens between the interviewer and interviewee make it hard for the interviewer to notice and react to face expressions, movements, etc. of interviewees. Moreover, as an interviewer, you also have to pay attention to the online platform – e.g., whether the recording and transcription are running. At the side of the interviewee, the online interaction might influence their willingness to share insights, or their trust in the interviewer, both positively and negatively. The online interaction can bring or mitigate response bias. Interviewees might feel that they cannot say everything, because they are more directly confronted by the interviews being recorded. At the same time, some people might be more comfortable in giving an interview in their own environment, not being distracted by what happens at the office. Accordingly, interviewees might feel freer to answer questions. Altogether, performing interviews online does influence the results of an interview.

Nonetheless, we deliberately choose to conduct interviews to identify design practices and to do them online. In addition, at the time of our empirical study, our society was still dealing with the COVID-19 pandemic, which made physical meetings difficult to plan and uncertain. This was also a reason for doing interviews online. Planning an interview online made the chance of cancelling interviews lower. Moreover, practitioners were skilled in doing online meetings as it was a prevalent communication channel at the moment.

Research approach – generalisability

The generalisability of our findings differs between the three parts of this study. Therefore, we discuss the three parts separately. First, the empirical findings from Part I Diagnosis. The starting point for the empirical study was to examine a diversity of public organisations, to define broad categories of designers who could participate in the study, and to use triangulation to increase the reliability of our results. Accordingly, the empirical results are generalisable to public organisations operating on national, regional, and local level. However, the participating organisations were all situated in the Netherlands. Therefore, our sample is only representative for the Netherlands. For other national contexts, the empirical findings are not directly applicable but can be used as point of departure for further empirical research. For example, our research confirms insights from studies in the USA (Mulligan & Bamberger, 2018, 2019).

Second, the insights resulting from abductive reasoning in Part II Appraisal can be generalised to other national contexts but raise questions about universality. Regarding the cases studied in Chapter 6, our aim was to compare cases of erroneous and harmful public algorithmic systems in different national contexts thoroughly described in scientific literature. We analysed four cases that satisfied our inclusion criteria. Moreover, the childcare allowances scandal, the Robodebt scheme, the Post Office scandal, and the DUO case were widely covered in the media. Still, it can be argued that the included cases are extreme cases and, therefore, are not fully representative. The cases might be extreme in the sense of the consequences for citizens. But considering the technical and institutional artefacts, the used public algorithmic systems are not much distinct from general practice in public organisations. In addition, studying past cases always brings the risk of hindsight bias. Finally, we relied on the analysis of other researchers by performing a secondary analysis of cases. Thereby, our analysis was inherently steered by the interpretations of these cases by those researchers.

Regarding the presuppositions used in Part II, we have used elements in the concepts of democracy and the Rule of Law that are widely accepted. Democracy and the Rule of Law are starting points for the organisation of most modern states, especially liberal democracies. But each jurisdiction has its own interpretation of these contested concepts. Moreover, the level of adherence to principles related to the concepts varies between countries. To address this variation, we used notions in political philosophy and philosophy of law to arrive at more universal understandings of the presuppositions. Nevertheless, the theoretical notions needed to be contextualised. For that purpose, this research used the context of Dutch public organisations. Our intention is that the design principles in the design theory are principles that are, to some extent, universally applicable. The institutional interventions can be used as examples for how to translate the design principles to practice.

Finally, the evaluation results in Part III Design are not generalisable since the test of our design principles was indirect and exploratory. The evaluation of the principles was indirect because we first had to translate the principles to institutional interventions and because we simulated the design process. The circumstances in the simulation reflected reality, but the effects of our interventions will probably be different when implemented in real practice. Our evaluation was exploratory since we only tested our principles in one local public organisation. At best, the evaluation results are generalisable to medium-sized local governments in the Netherlands. Upscaling and replication of the test is needed to say anything about the generalisability of the design theory. For the purpose of our research, the use of a singular simulation was justifiable as it provided insight on the elements into the theory that worked and elements that need elaboration or iteration.

Research scope – focus on institutional interventions for design practices

In our research, we used a specific scope for the analytical lens and the deliverables of our research. Concerning the analytical lens, we focused on design practices for algorithmic systems in public organisations. This focus is justifiable considering the current emphasis on the use of algorithmic systems in research. However, this scope risks claiming that most problems can be solved by organising design processes 'properly', which would disregard the complexity of addressing algorithmic harms inflicted on citizens through design (cf. Dorst, 2019c). Moreover, the distinction between designing and using systems can often not be made in practice (cf. Orlikowski, 1992). Regarding the research deliverable, we aimed to produce a design theory for institutional interventions. As such, we left required and/or possible technical interventions out of the research scope. And although we emphasised the importance of informal institutions, we have not fully elaborated such institutions in our design theory.

Moreover, following from our pragmatic position, this thesis presupposes specific interpretations of democracy (i.e., a representative form based on deliberation and inquiry) and the Rule of Law (i.e., emphasising the socio-legal perspective). As such, we took a normative position in scientific debates on these contested concepts. Refraining from such normative choices would prevent us to formulate a design theory. Nevertheless, the design theory itself is not a normative but a prescriptive theory. We do not argue that institutional interventions based on our design principles are the only way to achieve goals related to improving design processes of public algorithmic systems. Other interpretations of the contested concepts will likely result in variations of our design principles that might also improve coordination among designers and enlarge legitimacy of the design process. Moreover, as mentioned before, we used Dutch democratic and Rule of Law practices to arrive at instantiations of the prescribed design principles. Especially the interaction between politicians and public servants will probably differ in other national contexts. Finally, our design theory is based on assumptions that can be challenged. We did not include citizens in the evaluation and assumed that the voice of citizens is brought in the design process through city council members. Accordingly, we excluded the public debate on algorithmic systems from the simulation of a design process.

Research scope – a design theory still in development
It is important to stress that the outcome of this thesis is a prescriptive design theory. This means that it provides possible ways to tackle issues prevalent in design practices for public algorithmic systems. It does not take a normative stance on how these design practices should be achieved (cf. Gregor, 2006). And as mentioned before, a design theory such as prescribed in this research always needs to be contextualised.

Moreover, the design theory itself is not yet finalised. This research did not aim to deliver a final theory but was meant to be a first iteration that needs expansion and scrutiny by performing new generate-test cycles. Its prescriptive power can be increased in the following ways. First, there are elements we could not test, such as giving system-level designers the chance to formulate their own individual opinions on the design advice, or the interaction between component designers and system-level designers. Second, where the system map showed immediate potential, other elements need to be enhanced – e.g., the interaction between politicians and system-level designers via design outputs. Third, the design theory provides a first step in embedding design practices in democratic constitutional states. To accomplish such embedding fully (if that is possible) is a long-winded work that requires more structural interventions in the organisation of those democratic constitutional states.

10.4 Recommendations

This section presents directions for future research (10.4.1) and policy recommendations (10.4.2) that follow from the research presented in this thesis.

10.4.1 Future research

Iterate and elaborate the design theory
As mentioned before, the design theory in this thesis is not a final product. Instead, it should be expanded, detailed, and scrutinised in further research. Expanding can be done by exploring new design principles that fit the meta-theory or by examining the changes to design principles when other interpretations of democracy the Rule of Law are used. The design theory can be detailed by refining the design principles, for example, by distinguishing between universal parts of the principles and parts that need to be contextualised. Scrutinising the design theory can be done by upscaling and replicating the current evaluation. First, this is about improving the current simulation of design processes. Thereafter, the design theory should also be tested in real practices. There are two possible ways of evaluating the theory in practice: 1) through case studies in which the prescribed principles can be identified and assessed, and 2) by collaborating with public organisations that translate the design principles to institutional interventions and implement these interventions in an actual design project.

Expand empirical research on issues in current uses and designs of public algorithmic systems
Where scientific and popular debate on AI seems to be consumed by presumed existential risks, this research underlines that there is still a lot to learn from current issues and hazards in

public algorithmic systems. To this moment, there are only a few, well-known cases of which the hazardous public algorithmic systems have been examined. Research does not have to reinvent the wheel, we can learn from domains studying safety in software systems (e.g., system safety - see Leveson, 2012; Dobbe, 2022), or from lessons learned from earlier forms of automation and digitalisation (see Bannister, 2023; Lindgren, 2023; Meijer & Löfgren, 2015).

At the end of this research, the emergence of large language models and generative AI accessible to many people changed the perspective on AI. There can be no doubt that public servants and public organisations are using such systems or are experimenting with using them. Probably they bring a new dynamic to practices in public organisations that also need to be studied (Tangi et al., 2026). For example, what is the role of generative AI in arbitrary conduct? What role can institutional design play in governing the design and use of generative AI in the public sector?

Explore the possibilities of design for informal institutions

In the end, this research mostly focused on formal institutions for design practices of public algorithmic systems. We already emphasised the importance of complementary informal institutions. More research is needed on the informal institutions that follow from the meta-theory. Subsequently, the possibilities for encouraging or stimulating the uptake of such informal institutions needs to be studied. Can informal institutions be engendered by training and educating public servants that design algorithmic systems? Or is this about steering the creation of informal institutions in practice through formal institutions that instigate the emergence of specific informal institutions? The answers to these questions are probably also informative about how to shift prevalent presuppositions in public organisations to desired presuppositions.

Transfer insights to other domains

This research focused on public algorithmic systems, but our findings probably also apply to other contexts. The systems are part of a longer trend of automation in public organisations and share characteristics with other digital and socio-technical systems. The insights from this research can, therefore, be applied to other socio-technical systems and other domains. For example, design practices in urban planning (i.e., a field in which public organisation are designers of socio-technical systems in the built environment) can be compared to those of algorithmic systems. This may result in learnings between the different domains.

10.4.2 Policy recommendations

Use the notions of arbitrary conduct and Kafkaesque situations in reflecting on public algorithmic systems

Although mostly a metaphor from literature, using the notion of Kafkaesque situations is particularly suitable for public organisations to explore harmful algorithmic systems from a citizen perspective. It supports in focusing on the consequences of algorithmic systems that should be prevented, mitigated, or corrected. Using the metaphor of Kafkaesque can counterbalance the prevalent organisational and inward perspective in public organisations.

Similarly, arbitrary conduct is a concept useful for public organisations in understanding their algorithmic systems. The concept is relatable to these organisations and captures many problems associated to algorithmic systems. Using the concept will encourage public organisations to shift their focus from specific technical or institutional components in algorithmic systems to the emergent properties of these systems.

Integrate design practices in algorithmic and AI governance

Based on this research, we recommend that public organisations evaluate their current design practices and explicitly include considerations on design approaches in the governance of algorithmic systems. Our design theory can be used for that purpose. Parts of the design theory that can be directly implemented are the system-level design team and the assignment of the task to formulate the socio-technical specification of public algorithmic systems (e.g., through a system map) to that team of designers. Apart from that, public organisations can explore possibilities for politicians to be included in the deliberation and formulation of design goals and design constraints in their current lawmaking and policymaking tasks.

Use the notion of feedback loops to align different work practices

Feedback loops and iteration can be used as a central concept in design practices of public algorithmic systems. This recommendation is related to the importance of informal institutions and attitudes that support reflectivity and responsiveness within public organisations. Moreover, the notion of feedback is coming back in both the presuppositions underlying current practices as well as those related to desired practices: agile and scrum, socio-technical designing, democracy, and the Rule of Law. Having iteration in common, the notion of feedback can form a basis for integrating practices or making practices complementary. Thereby, feedback loops address the problems of distinct time scales and work rhythms between different disciplinary designers and the need for coordination between those designers. Moreover, feedback loops are also an adequate approach to deal with the emergent properties of complex systems.

Position technology – more specifically, socio-technical systems – in relation to laws and regulations

Apart from lawmaking and policymaking organisations, public organisations should consider themselves as socio-technical system-making organisations. Politicians or representatives of citizens have a clear role in the conception of laws and policies (one can debate whether the role is sufficient or suitable, but often it is defined in constitutional laws). Considering that technology also is structuring and regulatory of nature, it is important to define the position of politicians and representatives in the creation of socio-technical systems. Politicians and representatives do not have to be included in decision-making on all details of such systems but, at least, in determining the design goals and design constraints.

Summary

Public organisations increasingly rely on algorithmic systems for the execution of their tasks and the provision of public services. These public algorithmic systems – comprising institutions, human agents, and automated rule-based and/or data-driven software technologies that predict or generate output – can inflict harms on citizens by creating Kafkaesque situations. Exemplary cases of erroneous algorithmic systems resulting in Kafkaesque situations are the Dutch childcare allowances scandal and the Australian Robodebt scheme. Public organisations are increasingly held responsible for addressing algorithmic Kafka. However, these organisations fall short in adequately preventing, mitigating, or correcting systems that create their own reality and corner citizens.

Unlike most research on public algorithmic systems, this monograph studies how causes of Kafkaesque situations originate in the design process of public algorithmic systems. Specifically, we studied the interactions between designers that could be traced back to harmful properties of algorithmic systems, and how these design processes can be changed in order to prevent, mitigate, or correct such Kafkaesque situations. To achieve the latter, this research aims to realise socio-technical design practices for algorithmic systems in the democratic and Rule of Law context of public organisations. This thesis uses design science as research approach to prescribe a design theory for institutional interventions that (re)shape the design practices for public algorithmic systems. We arrive at these institutional interventions following three research parts: diagnosis, appraisal, and create and assess.

Diagnosis

The first part of this thesis identifies current design practices for algorithmic systems in Dutch public organisations by conducting interviews with designers and policymakers, and observations of policymakers working on structuring design processes. This empirical study mainly focuses on the fundamental ideas shared by designers – i.e., presuppositions – that structure these design practices. We examine the implications of these presuppositions for the output of design processes. Moreover, we assess current institutional interventions developed for and implemented in design processes by public organisations. Part I Diagnosis answers research question 1: w*hat presuppositions underlie the design practices for public algorithmic systems that have emerged in public organisations?*

Dutch public organisations turn out to base their design practices, and the institutional interventions that they develop, mainly on two presuppositions: technocracy and business-like. The technocratic presupposition manifests itself in a disconnect between design choices made by public servants and the public debate that takes place in political bodies. Moreover,

technical experts obtain a coordinating role in the deliberation that is needed in projects of algorithmic system development. The businesslike presupposition takes shape in procurement-based client-supplier relationships between policy domain experts and technical developers. In addition, public organisations adopt business-oriented agile and scrum approaches, resulting in conflicts between the work rhythms of different involved designers. The assessment of institutional interventions of public organisations – i.e., a governance framework, an algorithm register, procurement conditions, and instructions for objection procedures – shows that these policy instruments are based on the same presuppositions.

The technocratic and businesslike presuppositions are a deficient basis for design practices of public algorithmic systems. First of all, the presuppositions also manifest in the outcomes of the design processes. Public algorithmic systems have technocratic and businesslike characteristics, and do not necessarily reflect values associated with public administration, such as democracy and the Rule of Law. Moreover, current design processes do not provide legitimacy to their outcomes, do not create overviews of the designed systems, lack means for integration of system components, and are characterised by slack.

Appraisal

The second part of this thesis elicits a meta-theory for desired design practices through abductive reasoning. First, points for intervention in the design processes are explored. This is done by examining well-studied cases of harmful public algorithmic systems through the lens of Kafka and arbitrary use of power. Thereafter, we formulate the meta-theory which comprises all theoretical foundations that form the basis for the design theory formulated in the third part of the thesis. We do this by juxtaposing insights from socio-technical designing with two elementary presuppositions in public administration: democracy and the Rule of Law. These presuppositions follow from the nature of public algorithmic systems – i.e., socio-technical – and the need to protect from arbitrary conduct – democracy and the Rule of Law. Part II answers research question 2: *What design practices that curb algorithmic Kafka are prescribed by the synthesis of the presuppositions of socio-technical designing, democracy, and the Rule of Law?*

We analysed scientific literature describing and analysing cases of algorithmic Kafka. We study how Kafkaesque situations emerged from the Dutch childcare allowances scandal, the Australian Robodebt scheme, and other cases in the United Kingdom, and the Netherlands. Our analysis indicates that Kafkaesque situations find their origin in possibilities for arbitrary use of power. These possibilities for arbitrary conduct are partly designed into the system by misalignment between institutional and technological components in the systems' socio-technical specification. Such misalignment indicates problems in coordination between designers of the different system components. In addition, arbitrary use of designerly power undermines the legitimacy of the socio-technical specification.

Protecting citizens from arbitrary use of power is the main purpose of both democracy and the Rule of Law. Both premises provide practices for reducing possibilities for arbitrary conduct. This thesis shows how these practices can be synthesised with premises of socio-technical designing. The synthesis is the result of abductive reasoning. First, we juxtapose the three premises and show the symbioses, shared challenges, and contradictions between them. Building on the juxtaposition, we synthesise the three premises in a meta-theory that comprises the assumptions behind the design theory formulated in the next research part. The assumptions give directions on what it means to embed socio-technical design practices of public algorithmic systems in a democratic and Rule of Law context. By using socio-technical designing, democracy, and the Rule of Law as presuppositions, the meta-theory is opposite to the technocratic and businesslike presuppositions that currently underlie design practices.

Create and Assess

In the final part, this monograph prescribes design principles that public organisations can use to create institutional interventions. First, the design principles are generated using the meta-theory formulated in Part II. Thereafter, institutional interventions based on the principles are evaluated in a simulation of a design process. Part III Create and Assess answers research question 3: *What institutional interventions engender interactions between designers of public algorithmic systems that align with democratic and Rule of Law principles?*

The design principles prescribe institutional interventions that are expected to transform current design practices into desired design practices. The main idea behind the design principles is to create the position of system-level designers, occupied by public servants, who formulate the socio-technical specification of public algorithmic systems. The socio-technical specification identifies technological, institutional, and agential system components, and establishes the interaction between these components. In addition, these system-level designers have the task of coordinating the creation of the individual components of the algorithmic system. But more importantly, these designers are engaged in a dialectic with representative designers, i.e., democratically mandated representatives, to guarantee the legitimacy of design choices pertaining to the socio-technical specification. We arrived at four design principles that support public organisations in establishing the position of system-level designers:

Design principle 1: *Establish an organisational level responsible for the socio-technical specification of public algorithmic systems in order to ensure mutual understanding about and the design of interactions between system components*

Design principle 2: *Establish formal conditions to which the socio-technical specification of public algorithmic systems should adhere to, in order to standardise the motivation behind and meticulousness of design choices*

Design principle 3: *Establish a structured dialectic between representative designers and system-level designers via their design outputs in order to stimulate learning among and reflection by both groups of designers*

Design principle 4: *Establish checks and balances between system-level designers, and between representative designers and system-level designers in order to reduce arbitrary use of designerly power*

We evaluated these design principles by simulating a design process of a public algorithmic system. The simulated design process is structured by institutional interventions that follow from the design principles. Our results show that the use of a socio-technical specification in the form of a system map provides a new perspective on public algorithmic systems for all involved actors – i.e., politicians, designers of technical artefacts, and designers of institutional artefacts. Moreover, the system-level design team can resolve several coordination problems. At the same time, participants in the simulation doubt whether the design practices in the simulation are feasible within their public organisation.

In general, this thesis demonstrates how the socio-technical design process of public algorithmic systems can be embedded in democratic and Rule of Law practices. The prescribed institutional interventions support system-level designers in both coordinating and legitimising design choices on the level of the socio-technical system. System-level designers that draft a socio-technical specification can address misalignment through system mapping and reduce arbitrary use of designerly power by strengthening the dialectic between public servants and politicians. The design practices that emerge by establishing the position of system-level designer will curb the emergence of Kafkaesque situations in public algorithmic systems.

Samenvatting

Publieke organisaties leunen zwaar op algoritmische systemen in de uitvoering van taken en het leveren van diensten. Deze publieke algoritmische systemen – bestaande uit actoren, instituties en geautomatiseerde, op regel-gebaseerde en/of data-gedreven, software die output voorspelt of genereert – creëren Kafkaëske situaties en kunnen daardoor schade toebrengen aan burgers. Twee voorbeeldcases van algoritmische systemen die dat soort Kafkaëske situaties voortbrachten zijn de Nederlandse Toeslagenaffaire en het Australische Robodebt schandaal. In toenemende mate worden publieke organisaties verantwoordelijk gehouden voor het adresseren van algoritmische Kafka. Desondanks schieten publieke organisaties tekort in het adequaat voorkomen, mitigeren of corrigeren van systemen die hun eigen werkelijkheid creëren en burgers in het nauw drijven.

Anders dan het merendeel van wetenschappelijk onderzoek naar publieke algoritmische systemen, onderzoekt deze monografie hoe Kafkaëske situaties ontstaan in de ontwerpprocessen van deze systemen. Meer specifiek, hebben we interacties tussen ontwerpers bestudeerd die herleidbaar zijn naar schadelijke elementen in algoritmische systemen en hoe het ontwerpproces kan worden veranderd om Kafkaëske situaties te voorkomen, mitigeren of corrigeren. Om dat laatste te bereiken beoogt dit onderzoek om socio-technische ontwerppraktijken voor algoritmische systemen te realiseren in de context van de democratische rechtsstaat. Dit proefschrift beschrijft een ontwerptheorie voor institutionele interventies die ontwerppraktijken van publieke algoritmische systemen kunnen (her)vormen. De ontwerptheorie volgt uit een *design science* onderzoeksaanpak. Het onderzoek is verdeeld in drie onderdelen: Diagnose, Beschouwing, en Creëren en Beoordelen.

Diagnose

Het eerste deel van deze thesis identificeert ontwerppraktijken voor algoritmische systemen binnen Nederlandse publieke organisaties op basis van interviews met ontwerpers en beleidsmakers en door middel van observaties van beleidsmakers die het ontwerpproces vormgeven. De empirische studie die daaraan ten grondslag ligt, focust op de fundamentele ideeën die ontwerpers delen – ofwel vooronderstellingen – die de ontwerppraktijken ordenen. Daarnaast beoordelen we institutionele interventies die publieke organisaties op dit moment ontwikkelen en implementeren in ontwerpprocessen. Deel I Diagnose beantwoordt onderzoeksvraag 1: *Welke vooronderstellingen liggen ten grondslag aan de ontwerppraktijken van publieke algoritmische systemen die zich voordoen in publieke organisaties?*

Nederlandse publieke organisaties blijken hun ontwerppraktijken, en ook hun institutionele interventies, hoofdzakelijk te baseren op twee vooronderstellingen: technocratie en verzakelijking. De technocratische vooronderstelling manifesteert zich in een ontkoppeling tussen ontwerpkeuzes die worden gemaakt door ambtenaren en het publieke debat dat plaatsvindt in politieke organen. Daarnaast hebben technische experts een coördinerende rol verkregen in de deliberatie die onontbeerlijk is in ontwerpprocessen van publieke algoritmische systemen. De verzakelijking vooronderstelling komt tot uiting in de klant-leverancier relatie die bestaat tussen domeinexperts en technische ontwikkelaars. Bovendien passen publieke organisaties procesaanpakken uit het bedrijfsleven toe, zoals agile en scrum, die zorgen voor conflicterende werkritmes van verschillende betrokken ontwerpers. De beoordeling van institutionele interventies die publieke organisaties inzetten – een governance framework, een algoritmeregister, aanbestedingsbepalingen en een handreiking voor bezwaarbehandelaars van algoritmische besluitvorming – laat zien dat ook deze beleidsinstrumenten op dezelfde vooronderstellingen zijn gebaseerd.

De technocratische en verzakelijking vooronderstellingen vormen een gebrekkige basis voor ontwerppraktijken van publieke algoritmische systemen. In de eerste plaats werken deze vooronderstellingen door in de uitkomsten van de ontwerpprocessen. Publieke algoritmische systemen hebben karaktertrekken van technocratie en verzakelijking, maar ontberen een basis in waarden die worden geassocieerd met openbaar bestuur zoals de democratische rechtsstaat. Daarnaast legitimeren huidige ontwerpprocessen ontwerpkeuzes onvoldoende, bieden ze onvoldoende middelen om systeemcomponenten te integreren en worden ze gekenmerkt door inertie.

Beschouwing

Het tweede deel van deze thesis formuleert, door middel van abductief redeneren, een metatheorie aangaande gewenste ontwerppraktijken. Eerst stellen we vast op welke aspecten van het huidige ontwerpproces van publieke algoritmische systemen geïntervenieerd kan of moet worden. Daarvoor beschouwen we reeds bestudeerde cases van schadelijke publieke algoritmische systemen vanuit het perspectief van Kafka en arbitraire machtsuitoefening. Daarna komen we tot de metatheorie die de theoretische basis voor de ontwerptheorie in Deel III van deze thesis. Dit doen we door de premisses van socio-technisch ontwerpen en de democratische rechtsstaat tegenover elkaar te zetten. Deze drie premisses representeren de aard van publieke algoritmische systemen – d.w.z. socio-technisch – en het vereiste om burgers te beschermen tegen arbitraire machtsuitoefening – d.w.z. democratisch rechtsstaat. Deel II beantwoordt onderzoeksvraag 2: *Welke ontwerppraktijken die algoritmische Kafka bedwingen volgen uit de synthese van premisses van socio-technisch ontwerpen en de democratische rechtsstaat?*

We hebben wetenschappelijke literatuur waarin cases van algoritmische Kafka zijn beschreven en geanalyseerd. De Kafkaëske situaties die zijn gecreëerd in de Nederlandse Toeslagenaffaire, het Australische Robodebt schandaal en andere cases in het Verenigd Koninkrijk en Nederland werden veroorzaakt door mogelijkheden voor arbitraire machtsuitoefening die waren ingebouwd in de socio-technische specificatie van de onderliggende algoritmische systemen. Deze mogelijkheden tot willekeur worden deels verankerd in de systemen doordat institutionele en technologische componenten in de specificatie van algoritmische systemen niet op elkaar afgestemd waren. Problemen in die afstemming wijzen op coördinatieproblemen tussen ontwerpers van de verschillende systeemcomponenten. Daarnaast wordt de legitimiteit van publieke algoritmische systemen ondermijnt door arbitraire machtsuitoefening van ontwerpers zelf.

Het beschermen van burgers tegen arbitraire machtsuitoefening is het hoofddoel van de democratische rechtsstaat. De twee onderdelen, democratie en rechtsstaat, in deze vorm van staatsinrichting bieden handelswijzen om mogelijkheden tot willekeur te beteugelen. Dit proefschrift toont aan dat deze handelswijzen samengevoegd kunnen worden met de fundamentele kenmerken van socio-technisch ontwerpen. Ten eerste hebben we de drie premisses tegenover elkaar gezet en identificeren we mogelijkheden voor symbiose, gedeelde uitdagingen en tegenstellingen tussen de premisses. Daaruit volgend formuleren we een metatheorie die bestaat uit de assumpties die ten grondslag liggen aan de ontwerptheorie die wordt ontwikkeld in het derde onderzoeksdeel. Deze aannames geven richting aan het inbedden van socio-technische ontwerppraktijken in de context van de democratische rechtsstaat. Door socio-technisch ontwerpen, democratie en de rechtsstaat te gebruiken als premisses, bieden ontwerppraktijken gebaseerd op de metatheorie een alternatief voor de huidige ontwerppraktijken die zijn gebaseerd op technocratische en verzakelijkingsvooronderstellingen.

Creëren en Beoordelen

Het laatste thesisdeel schrijft ontwerpprincipes voor die publieke organisaties kunnen gebruiken om institutionele interventies op te baseren. We formuleren de ontwerpprincipes op basis van de metatheorie uit Deel II. Vervolgens worden op deze principes gebaseerde institutionele interventies geëvalueerd in een simulatie van een ontwerpproces. Deel III Creëren en Beoordelen beantwoordt onderzoeksvraag 3: *Welke institutionele interventies brengen interacties tussen ontwerpers van publieke algoritmische systemen voort die voldoen aan de principes van de democratische rechtsstaat?*

De ontwerpprincipes schrijven institutionele interventies voor waarvan wordt verwacht dat ze huidige ontwerppraktijken transformeren naar gewenste ontwerppraktijken. In de basis creëren de ontwerpprincipes de positie van systeemniveau-ontwerpers die de socio-technische specificatie van publieke algoritmische systemen formuleren. Deze systeemniveau-ontwerpers hebben daarnaast als taak om het ontwerpen van individuele componenten van algoritmische systemen te coördineren. In het bijzonder zijn deze systeemniveau-ontwerpers onderdeel van een dialectiek met vertegenwoordigende ontwerpers, d.w.z. democratisch gemandateerde volksvertegenwoordigers, om de legitimiteit van ontwerpkeuzes in de socio-technische systeemspecificatie te waarborgen. We zijn tot vier ontwerpprincipes gekomen die publieke organisaties helpen om de positie van systeemniveau-ontwerpers in te stellen:

Ontwerpprincipe 1: *Stel een organisatieonderdeel in dat verantwoordelijk is voor de socio-technische specificatie van publieke algoritmische systemen om gedeeld begrip over en het ontwerpen van interacties tussen systeemcomponenten te garanderen*
Ontwerpprincipe 2: *Stel formele condities op waaraan de socio-technische specificatie van publieke algoritmische systemen moet voldoen om de motivering achter en zorgvuldigheid van ontwerpkeuzes te standaardiseren*
Ontwerpprincipe 3: *Stel een gestructureerde dialectiek tussen vertegenwoordigende en systeemniveau-ontwerpers in via hun ontwerpoutput om leren en reflectie bij beide groepen van ontwerpers te stimuleren*

Ontwerpprincipe 4: *Stel* checks and balances *tussen systeemniveau ontwerpers onderling en tussen vertegenwoordigende en systemniveau-ontwerpers in om arbitraire machtsuitoefening door ontwerpers te verminderen*

We hebben deze vier ontwerpprincipes geëvalueerd door een ontwerpproces van een publiek algoritmisch systeem te simuleren. Het gesimuleerde ontwerpproces is gebaseerd op institutionele interventies die we hebben gebaseerd op de ontwerpprincipes. De simulatieresultaten tonen dat het gebruik van een socio-technische specificatie in de vorm van een *system map* alle betrokken ontwerpers – dus vertegenwoordigende ontwerpers, ontwerpers van de technische artefacten en van de institutionele artefacten – een nieuw perspectief op publieke algoritmische systemen brengt. Daarnaast stelt het de systeemniveau-ontwerpers in staat om verschillende coördinatieproblemen die voorkomen in het ontwerpproces aan te pakken. Tegelijkertijd betwijfelden simulatiedeelnemers of het implementeren van de voorgeschreven ontwerppraktijken haalbaar is in hun publieke organisatie.

Deze thesis voorziet in algemene voorschriften om het socio-technische ontwerpproces van publieke algoritmische systemen in de praktijk van de democratische rechtsstaat in te bedden. De voorgestelde institutionele interventies ondersteunen ontwerpers om ontwerpkeuzes op systeemniveau te coördineren en te legitimeren. Systeemniveau-ontwerpers kunnen problemen in afstemming binnen de socio-technische specificatie aanpakken met een *system map* en arbitraire machtsuitoefening door ontwerpers verminderen door de dialectiek tussen ambtenaren en politici te versterken. De ontwerppraktijken die ontstaan door de positie van systeemniveau-ontwerper in te stellen zal de ontwikkeling van Kafkaëske situaties in publieke algoritmische systemen verminderen.

Acknowledgements

I would like to express my gratitude to all the people that supported me in the process of writing this dissertation.

Ten eerste, Roel. Het begon allemaal bij jouw uitnodiging om aan dit avontuur, dat me zeer dierbaar is geworden, te beginnen. Je hebt me de afgelopen jaren ongelofelijk veel vrijheid gegeven om zowel mijn onderzoek als mijn promotietraject invulling te geven. Dat is een belangrijke voorwaarde om plezier en interesse in onderzoek te houden. Ik gun het elke promovendus.

Marijn, je enorme kennis van de literatuur heeft me enorm geholpen in het positioneren van mijn onderzoek. Daarnaast is je feedback scherp, constructief en gedetailleerd. Ik ben je ook dankbaar voor de mogelijkheid die je me hebt geboden om me verder te bekwamen in het geven van (academisch) onderwijs.

Officially, Roel and Marijn were my supervisors. But in practice, I would argue that Wiebke's role was as important in finding my way as PhD candidate. Wiebke, in spite of Covid, we became close office mates. Thank you for making me reflect on my own thoughts and ideas, and for helping me to develop my research direction. I must admit that I miss our conversations since you left TPM. Wiebke, you gave me the confidence to follow my own path in academia and showed me that a good academic should always question academic practices.

I also want to thank all other people that I shared an office with. Antragama, thank you very much for bringing cheerfulness and optimism to the office. Maurus, it is so great to see your development from master student to PhD candidate. I am very much looking forward to the insights your research will bring. I know the quality will be high. Cilia, je hebt ons kantoor in een gezellige plek weten te veranderen. Wie had gedacht dat dat mogelijk is in het TBM-gebouw.

Floris en Antonia, we begonnen ongeveer tegelijkertijd als promovendus op onderwerpen die op het eerste gezicht weinig van elkaar verschilden. We hebben er alle drie onze eigen draai aan gegeven, maar het was erg fijn om met jullie over het onderwerp te sparren. Kars, ook jou wil ik bedanken voor de gesprekken over het ontwerpen van AI-systemen door de overheid. Je hebt me kennis laten maken met andere perspectieven op ontwerpen.

Jacqui, thank you for bringing liveliness and good discussions on societal matters to our research group. And for always pointing out the absurdity in the AI field. Eva, dankjewel voor je hulp bij de evaluatie van mijn ontwerptheorie en de analyse van de resultaten. Ik hoop dat we snel echt iets samen kunnen schrijven. Íñigo, our PhD trajectory followed a very similar path. I owe much to the discussions we had over the years. And thank you for making me laugh with your dry humour.

Of course, there are many more colleagues in the ICT section that I want to thank; especially for all the feedback you gave me at ICT colloquia, your help in testing my research approaches, and the gezelligheid at conferences. Maar met name wil ik het secretariaat van de ICT-sectie bedanken. Laura, Ellen en Fanny, dankjulliewel voor alle hulp en ondersteuning.

There are many people that have provided invaluable input to my research. Arjan bedankt voor het delen van je inzichten uit de praktijk opgedaan bij de Kafkabrigade. Je papers maar vooral onze gesprekken zijn erg belangrijk geweest voor de afbakening van mijn onderzoek. Rainer and Lukas, thank you for your feedback and conversations during my short research stay at UCL. Neelke and Ben, thank you for your reflections on my research proposal at my Go/NoGo meeting.

Ik wil ook alle deelnemers aan mijn onderzoek bedanken. Het betekent erg veel voor onderzoekers wanneer mensen die werkzaam zijn in de praktijk hun tijd vrijmaken om, bijvoorbeeld, mee te doen aan interviews. Met name het kernteam van het consortium heeft veel van hun tijd voor mij vrijgemaakt. Omwille van anonimiteit zal ik jullie hier niet noemen, maar ik wil jullie bedanken voor alle hulp.

Elsbeth, het is alweer even geleden dat we aan ons beleidsvoorstel werkten bij de Nationale DenkTank, maar het is wel waar dit onderzoek is ontstaan. Ik heb vaak teruggedacht aan hoe we in dat hok op de OMHP samen hebben geschreven aan het voorstel. Zonder jou was dit onderzoek er zeker niet geweest.

Caroline, zo leuk dat je mijn paranimf bent! Het doet me goed dat we sinds de Korvezee nog steeds samen eten en biertjes drinken. Heel erg fijn dat ik dit bijzondere moment met jou kan delen.

Juultje, natuurlijk ben jij mijn paranimf! Er zijn weinig mensen die me zo goed kennen en begrijpen zoals jij. Ik ben enorm trots op jou.

Mama en papa, dankjewel voor alles wat jullie van kleins af aan voor mij en Julie hebben gedaan. Ook tijdens mijn promotie bleken jullie van onschatbare waarde.

Don, mijn PhD werd veel leuker toen ik jou leerde kennen. Jij helpt me om mijn onderzoek weg te leggen (ook heel belangrijk) en te genieten van al het andere dat we ondernemen. Heel erg bedankt ook voor je hulp bij de opmaak van dit boek. Dat had ik nooit alleen gekund. Maar bovenal wil ik je bedanken voor je onvoorwaardelijke steun.

About the author

Sem Nouws was born on 19 October 1994 in Breda and grew up in Klein-Zundert. From 2007 to 2013, he attended Mencia Sandrode in Zundert and Mencia de Mendoza in Breda for his pre-university education.

Sem completed both his bachelor and master at the faculty of Technology, Policy and Management (TPM) at Delft University of Technology. There he was trained as complex systems engineer with a specialisation in the build environment and spatial development. As treasurer for SHS Delft, a foundation that transforms vacant office buildings into student housing, he brought his engineering schooling to practice. Besides his activities in Delft, Sem obtained a Bachelor Degree in Law from Erasmus University Rotterdam.

Before writing his master thesis, Sem participated in the Nationale DenkTank. This yearly, four-month project brings together 20 young academics to analyse and address a societal problem. The theme of 2019: the digital society. Together with Elsbeth van den Hazel, Sem wrote a policy advice on collective protection of personal data. It was this policy advice that brought Sem into contact with his (to be) co-promotor Roel Dobbe.

Sem started his PhD research at the ICT section of the Engineering Systems and Services department of the TPM faculty in November 2020. He combined his interests in socio-technical theory, philosophy of law and democratic theory in his research by focusing on democratising the design process of public algorithmic systems.

Sem continues his research at the TPM faculty. For his postdoctoral research, he studies institutional prerequisites for using AI applications in mass deliberation.

List of publications

Book chapter

Nouws, S. & Dobbe, R. (2024). The Rule of Law for Artificial Intelligence in Public Administration: A System Safety Perspective. In K. Prifti, E. Demir, J. Krämer, K. Heine & E. Stamhuis (Eds.), *Digital Governance* (Vol. 39, pp. 183-208). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-639-0_9

Conference paper

Nouws, S., Janssen, M. & Dobbe, R. (2022). Dismantling Digital Cages: Examining Design Practices for Public Algorithmic Systems. In M. Janssen, C. Csáki, I. Lindgren, E. Loukis, U. Melin, G. Viale Pereira, M.P. Rodríguez Bolívar & E. Tambouris (Eds.), *Electronic Government* (Vol. 13391, pp. 307-322). Springer International Publishing. https://doi.org/10.1007/978-3-031-15086-9_20
Best Paper Award

Workshops

Nouws, S., Martinez De Rituerto De Troya, Í., Dobbe, R. & Janssen, M. (2023). Diagnosing and Addressing Emergent Harms in the Design Process of Public AI and Algorithmic Systems. In D. Duenas Cid, N. Sabatina, L Hagen & H. Liao (Eds.), *dg.o '23: Proceedings of the 24th Annual International Conference on Digital Government Research* (pp.679-681). Association for Computing Machinery. https://doi.org/10.1145/3598469.3598557

Nouws, S., Janssen, M. & Tambouris, E. (2025). Institutional design for AI-enabled deliberation. *Proceedings of Ongoing Research, Workshops, and Posters of the International Conference EGOV-CeDEM-ePart conference 2025.*

References

Adler, P. S., & Borys, B. (1996). Two Types of Bureaucracy: Enabling and Coercive. *Administrative Science Quarterly*, *41*(1), 61. https://doi.org/10.2307/2393986

Agre, P. E. (1994). Surveillance and capture: Two models of privacy. *The Information Society*, *10*(2), 101–127. https://doi.org/10.1080/01972243.1994.9960162

AI HLEG. (2019a). *A Definition of AI: Main Capabilities and Disciplines*.

AI HLEG. (2019b). *Ethics Guidelines for Trustworthy AI*.

Aizenberg, E., & Van Den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, *7*(2), 205395172094956. https://doi.org/10.1177/2053951720949566

Aken, J. E. V. (2004). Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. *Journal of Management Studies*, *41*(2), 219–246. https://doi.org/10.1111/j.1467-6486.2004.00430.x

Alfrink, K., Keller, I., Kortuem, G., & Doorn, N. (2022). Contestable AI by Design: Towards a Framework. *Minds and Machines*, *33*(4), 613–639. https://doi.org/10.1007/s11023-022-09611-z

Algorithm Audit. (2024a). *Vooringenomenheid voorkomen: Aanbevelingen voor risicoprofilering in het Controle Uitwonendenbeurs-proces: Een kwantitatieve en kwalitatieve analyse*.

Algorithm Audit. (2024b). *Vooringenomenheid voorkomen: Addendum*.

Aligică, P. D., & Boettke, P. J. (2010). *Challenging institutional analysis and development: The Bloomington School* (Repr). Routledge.

Aligică, P. D., & Tarko, V. (2012). Polycentricity: From Polanyi to Ostrom, and Beyond. *Governance*, *25*(2), 237–262. https://doi.org/10.1111/j.1468-0491.2011.01550.x

Alkhatib, A. (2021). To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3411764.3445740

Alkhatib, A., & Bernstein, M. (2019). Street-Level Algorithms: A Theory at the Gaps Between Policy and Decisions. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300760

Allhutter, D., Cech, F., Fischer, F., Grill, G., & Mager, A. (2020). Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. *Frontiers in Big Data*, *3*, 5. https://doi.org/10.3389/fdata.2020.00005

Andrejevic, M. (2020). Data Civics: A Response to the “Ethical Turn.” *Television & New Media*, *21*(6), 562–567. https://doi.org/10.1177/1527476420919693

Archer, M. (1995). *Realist Social Theory: The Morphogenetic Approach* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511557675

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (1998). *Critical Realism*. Routledge. https://doi.org/10.4324/9781315008592

Augustine, G., Lodge, J. S., & Radic, M. (2024). From Paralysis to Publicization: How Victims Confront Organizational Harm. *Academy of Management Proceedings*, *2024*(1), 15622. https://doi.org/10.5465/AMPROC.2024.315bp

Autoriteit Persoonsgegevens. (2024). *DUO: Gebruik van geautomatiseerde risicoclassificering op basis van een risicoprofiel bij Controleproces Uitwonende Beurs (CUB)*.

Balayn, A. M. A. (2023). *On developers’ practices for hazard diagnosis in machine learning systems* [Delft University of Technology]. https://doi.org/10.4233/UUID:EA94239F-5E95-4705-9DEB-32196D74DAAA

Bannister, F. (2017). In Defence of Bureaucracy: Governance and Public Values in a Digital Age. In A. A. Paulin, L. G. Anthopoulos, & C. G. Reddick (Eds.), *Beyond Bureaucracy* (Vol. 25, pp. 27–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-54142-6_3

Bannister, F. (2023). Beyond the box: Reflections on the need for more blue sky thinking in research. *Government Information Quarterly*, *40*(3), 101831. https://doi.org/10.1016/j.giq.2023.101831

Barad, K. (2003). Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. *Signs: Journal of Women in Culture and Society*, *28*(3), 801–831. https://doi.org/10.1086/345321

Barocas, S., & Selbst, A. D. (2016). *Big Data’s Disparate Impact*. https://doi.org/10.15779/Z38BG31

Bauer, J. M., & Herder, P. M. (2009). Designing Socio-Technical Systems. In A. Meijers (Ed.), *Philosophy of Technology and Engineering Sciences* (pp. 601–630). North-Holland. https://doi.org/10.1016/B978-0-444-51667-1.50026-4

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, *23*(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

Bayamlıoğlu, E., & Leenes, R. (2018). The ‘rule of law’ implications of data-driven decision-making: A techno-regulatory perspective. *Law, Innovation and Technology*, *10*(2), 295–313. https://doi.org/10.1080/17579961.2018.1527475

Beck, U. (1992). *Risk society: Towards a new modernity*. Sage Publications.

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bhaskar, R. (1979). *The Possibility of Naturalism*. Routledge. https://doi.org/10.4324/9781315756332

Bhaskar, R. (1998). Philosophy and Scientific Realism. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical Realism: Essential Readings*.

Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. *MIS Quarterly*, *1*(3), 17. https://doi.org/10.2307/248710

Bots, P. W. G. (2007). Design in socio-technical system development: Three angles in a common framework. *Journal of Design Research*, *5*(3), 382. https://doi.org/10.1504/JDR.2007.014883

Bouwmeester, M. (2023). System failure in the digital welfare state: Exploring parliamentary and judicial control in the Dutch childcare benefits scandal. *Recht Der Werkelijkheid*, *44*(2), 13–37. https://doi.org/10.5553/RdW/138064242023044002003

Bovens, M., & Zouridis, S. (2002). From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. *Public Administration Review*, *62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 97–117). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_5

Braithwaite, V. (2020). Beyond the bubble that is Robodebt: How governments that lose integrity threaten democracy. *Australian Journal of Social Issues*, *55*(3), 242–259. https://doi.org/10.1002/ajs4.122

Bratteteig, T., & Wagner, I. (2016). Unpacking the Notion of Participation in Participatory Design. *Computer Supported Cooperative Work (CSCW)*, *25*(6), 425–475. https://doi.org/10.1007/s10606-016-9259-4

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brey, P. (2018). The strategic role of technology in a good society. *Technology in Society*, *52*, 39–45. https://doi.org/10.1016/j.techsoc.2017.02.002

Brownsword, R. (2016). Technological management and the Rule of Law. *Law, Innovation and Technology*, *8*(1), 100–140. https://doi.org/10.1080/17579961.2016.1161891

Buitelaar, E., Lagendijk, A., & Jacobs, W. (2007). A Theory of Institutional Change: Illustrated by Dutch City-Provinces and Dutch Land Policy. *Environment and Planning A: Economy and Space*, *39*(4), 891–908. https://doi.org/10.1068/a38191

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. https://doi.org/10.1177/2053951715622512

Carlsson, S. A., Henningsson, S., Hrastinski, S., & Keller, C. (2011). Socio-technical IS design science research: Developing design theory for IS integration management. *Information Systems and E-Business Management*, *9*(1), 109–131. https://doi.org/10.1007/s10257-010-0140-6

Carney, T. (2019). Robo-debt illegality: The seven veils of failed guarantees of the rule of law? *Alternative Law Journal*, *44*(1), 4–10. https://doi.org/10.1177/1037969X18815913

Cash, P. J. (2018). Developing theory-driven design research. *Design Studies*, *56*, 84–119. https://doi.org/10.1016/j.destud.2018.03.002

Centeno, M. A. (1993). The new Leviathan: The dynamics and limits of technocracy. *Theory and Society*, *22*(3), 307–335. https://doi.org/10.1007/BF00993531

Chang, R. (2017). Hard Choices. *Journal of the American Philosophical Association*, *3*(1), 1–21. https://doi.org/10.1017/apa.2017.7

Checkland, P. (1981). *Systems thinking, systems practice*. J. Wiley.

Checkland, P. (1999). *Soft systems methodology: A 30-year retrospective* (Repr.). Wiley.

Christiano, Thomas. (1996). *The rule of the many: Fundamental issues in democratic theory*. Westview Press.

Christie, J. (2020). The Post Office Horizon IT scandal and the presumption of the dependability of computer evidence. *Digital Evidence and Electronic Signature Law Review*, 49–70. https://doi.org/10.14296/deeslr.v17i0.5226

Clarke, R., Michael, K., & Abbas, R. (2024). Robodebt: A Socio-Technical Case Study of Public Sector Information Systems Failure. *Australasian Journal of Information Systems*, *28*. https://doi.org/10.3127/ajis.v28.4681

Clegg, C. W. (2000). Sociotechnical principles for system design. *Applied Ergonomics*, *31*(5), 463–477. https://doi.org/10.1016/S0003-6870(00)00009-0

Clegg, S., E Cunha, M. P., Munro, I., Rego, A., & De Sousa, M. O. (2016). Kafkaesque power and bureaucracy. *Journal of Political Power*, *9*(2), 157–181. https://doi.org/10.1080/2158379X.2016.1191161

Cobbe, J., Veale, M., & Singh, J. (2023). Understanding accountability in algorithmic supply chains. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1186–1197. https://doi.org/10.1145/3593013.3594073

Collier, D., Daniel Hidalgo, F., & Olivia Maciuceanu, A. (2006). Essentially contested concepts: Debates and applications. *Journal of Political Ideologies*, *11*(3), 211–246. https://doi.org/10.1080/13569310600923782

Cunningham, F. (2002). *Theories of democracy: A critical introduction*. Routledge.

De Bruijn, H., & Herder, P. M. (2009). System and Actor Perspectives on Sociotechnical Systems. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *39*(5), 981–992. https://doi.org/10.1109/TSMCA.2009.2025452

De Bruijn, H., Ten Heuvelhof, E., & In 'T Veld, R. (2010). *Process Management*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13941-3

De Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, *39*(2), 101666. https://doi.org/10.1016/j.giq.2021.101666

De Laat, P. B. (2019). The disciplinary power of predictive algorithms: A Foucauldian perspective. *Ethics and Information Technology*, *21*(4), 319–329. https://doi.org/10.1007/s10676-019-09509-y

Dekker, S. (2012). *Just culture: Balancing safety and accountability* (2. ed). Ashgate.

Dencik, L., Redden, J., Hintz, A., & Warne, H. (2019). The 'golden view': Data-driven governance in the scoring society. *Internet Policy Review*, *8*(2). https://doi.org/10.14763/2019.2.1413

Dewey, John. (with Rogers, P. M. L.). (2016). *Public and its problems—An essay in political inquiry*. Ohio University Press; WorldCat.org. (Original work published 1927)

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

Dobbe, R., Dean, S., Gilbert, T., & Kohli, N. (2018). *A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics* (No. arXiv:1807.00553). arXiv. http://arxiv.org/abs/1807.00553

Dobbe, R. I. J. (2022). System Safety and Artificial Intelligence. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., p. C67.S1-C67.S18). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.67

Dobbe, R., Krendl Gilbert, T., & Mintz, Y. (2021). Hard choices in artificial intelligence. *Artificial Intelligence*, *300*, 103555. https://doi.org/10.1016/j.artint.2021.103555

Dorst, K. (2008). Design research: A revolution-waiting-to-happen. *Design Studies*, *29*(1), 4–11. https://doi.org/10.1016/j.destud.2007.12.001

Dorst, K. (2019a). Co-evolution and emergence in design. *Design Studies*, *65*, 60–77. https://doi.org/10.1016/j.destud.2019.10.005

Dorst, K. (2019b). Design beyond Design. *She Ji: The Journal of Design, Economics, and Innovation*, *5*(2), 117–127. https://doi.org/10.1016/j.sheji.2019.05.001

Dorst, K. (2019c). What Design Can't Do. *She Ji: The Journal of Design, Economics, and Innovation*, *5*(4), 357–359. https://doi.org/10.1016/j.sheji.2019.11.004

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2005). New Public Management Is Dead—Long Live Digital-Era Governance. *Journal of Public Administration Research and Theory*, *16*(3), 467–494. https://doi.org/10.1093/jopart/mui057

Dworkin, R. M. (1980). Political Judges and the Rule of Law. In *Proceedings of the British Academy* (Vol. 64, pp. 259–287).

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, *85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

Endacott, C. G., & Leonardi, P. M. (2024). Artificial intelligence as a mechanism of algorithmic isomorphism. In I. Constantiou, M. P. Joshi, & M. Stelmaszak (Eds.), *Research Handbook on Artificial Intelligence and Decision Making in Organizations* (pp. 342–358). Edward Elgar Publishing. https://doi.org/10.4337/9781803926216.00029

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First Edition). St. Martin's Press.

Farjoun, M. (2010). Beyond Dualism: Stability and Change As a Duality. *Academy of Management Review*, *35*(2), 202–225. https://doi.org/10.5465/amr.35.2.zok202

Feenberg, A. (1999). *Questioning technology*. Routledge.

Fenger, M., & Simonse, R. (2024). The implosion of the Dutch surveillance welfare state. *Social Policy & Administration*, *58*(2), 264–276. https://doi.org/10.1111/spol.12998

Fest, I., Wieringa, M., & Wagner, B. (2022). Paper vs. practice: How legal and ethical frameworks influence public sector data professionals in the Netherlands. *Patterns*, *3*(10), 100604. https://doi.org/10.1016/j.patter.2022.100604

Fleetwood, S. (2005). Ontology in Organization and Management Studies: A Critical Realist Perspective. *Organization*, *12*(2), 197–222. https://doi.org/10.1177/1350508405051188

Flyvbjerg, Bent. (1998). *Rationality and power: Democracy in practice* (S. L. Sampson, Trans.; English [ed.], rev.updated version.). University of Chicago Press.

Fountain, J. E. (2001). *Building the virtual state: Information technology and institutional change*. Brookings Institution Press.

Gallie, W. B. (1956). IX.—Essentially Contested Concepts. *Proceedings of the Aristotelian Society*, *56*(1), 167–198. https://doi.org/10.1093/aristotelian/56.1.167

Galloway, K. (2017). Big Data: A case study of disruption and government power. *Alternative Law Journal*, *42*(2), 89–95. https://doi.org/10.1177/1037969X17710612

Georgiadou, E., Rahanu, H., Siakas, K., Siakas, E., Siakas, D., Ross, M., & Staples, G. (2024). A Multidimensional Perspective of IT Systems Failures and Their Human Consequences – The Case of the UK Post-office IT Horizon System Failure and the Most Widespread Miscarriage of Justice. In M. Yilmaz, P. Clarke, A. Riel, R. Messnarz, C. Greiner, & T. Peisl (Eds.), *Systems, Software and Services Process Improvement* (Vol. 2180, pp. 325–344). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71142-8_23

Giddens, A. (1976). *New rules of sociological method: A positive critique of interpretative sociologies* (2., rev. ed., repr). Stanford Univ. Press.

Giddens, A. (1979). *Central Problems in Social Theory: Action, structure and contradiction in social analysis* (1st ed.). Bloomsbury Publishing Plc. https://doi.org/10.1007/978-1-349-16161-4

Giddens, A. (1986). *The Constitution of Society: Outline of the theory of structuration* (First paperback edition). University of California Press.

Giest, S. N., & Klievink, B. (2024). More than a digital system: How AI is changing the role of bureaucrats in different organizational contexts. *Public Management Review*, *26*(2), 379–398. https://doi.org/10.1080/14719037.2022.2095001

Gratton, G., Guiso, L., Michelacci, C., & Morelli, M. (2021). From Weber to Kafka: Political Instability and the Overproduction of Laws. *American Economic Review*, *111*(9), 2964–3003. https://doi.org/10.1257/aer.20190672

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, *45*, 105681. https://doi.org/10.1016/j.clsr.2022.105681

Green, B., & Chen, Y. (2019). Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 90–99. https://doi.org/10.1145/3287560.3287563

Greenstein, S. (2022). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, *30*(3), 291–323. https://doi.org/10.1007/s10506-021-09294-4

Gregor. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611. https://doi.org/10.2307/25148742

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Gregor, S., & Jones, D. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, *8*(5), 312–335. https://doi.org/10.17705/1jais.00129

Gregor, S., Kruse, L., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. *Journal of the Association for Information Systems*, *21*, 1622–1652. https://doi.org/10.17705/1jais.00649

Grill, G., Fischer, F., & Cech, F. (2023). Bias as Boundary Object: Unpacking The Politics Of An Austerity Algorithm Using Bias Frameworks. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1838–1849. https://doi.org/10.1145/3593013.3594120

Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of Algorithmic Decision-Making: Six Threats and the Need for a Calibrated Institutional Response. *Perspectives on Public Management and Governance*, *5*(3), 232–242. https://doi.org/10.1093/ppmgov/gvac008

Groenewegen, J. (2011). The Bloomington School and American Institutionalism. *The Good Society*, *20*(1), 15–36. https://doi.org/10.1353/gso.2011.0005

Growns, B., Kukucka, J., Moorhead, R., & Helm, R. K. (2024). The Post Office Scandal in the United Kingdom: Mental health and social experiences of wrongly convicted and wrongly accused individuals. *Legal and Criminological Psychology*, *29*(1), 17–31. https://doi.org/10.1111/lcrp.12247

Gürses, S., & Van Hoboken, J. (2018). Privacy after the Agile Turn. In E. Selinger, J. Polonetsky, & O. Tene (Eds.), *The Cambridge Handbook of Consumer Privacy* (1st ed., pp. 579–601). Cambridge University Press. https://doi.org/10.1017/9781316831960.032

Hajer, M. (1995). Politics on the move: The democratic control of the design of sustainable technologies. *Knowledge and Policy*, *8*(4), 26–39. https://doi.org/10.1007/BF02832228

Hajer, M. (2003). Policy without polity? Policy analysis and the institutional void. *Policy Sciences*, *36*(2), 175–195. https://doi.org/10.1023/A:1024834510939

Heikkila, T., & Andersson, K. (2018). Policy design and the added-value of the institutional analysis development framework. *Policy & Politics*, *46*(2), 309–324. https://doi.org/10.1332/030557318X15230060131727

Hendriks, F. (2022). Key values for democratic governance innovation: Two traditions and a synthesis. *Public Administration*, *100*(4), 803–820. https://doi.org/10.1111/padm.12738

Herder, P. M., Bouwmans, I., Dijkema, G. P. J., Stikkelman, R. M., & Weijnen, M. P. C. (2008). Designing infrastructures using a complex systems perspective. *Journal of Design Research*, *7*(1), 17. https://doi.org/10.1504/JDR.2008.018775

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, *19*(2), 87–92.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. JSTOR. https://doi.org/10.2307/25148625

Hildebrandt, M. (2011). Legal Protection by Design: Objections and Refutations. *Legisprudence*, *5*(2), 223–248. https://doi.org/10.5235/175214611797885693

Hildebrandt, M. (2016). Law *as* Information in the Era of Data-Driven Agency. *The Modern Law Review*, *79*(1), 1–30. https://doi.org/10.1111/1468-2230.12165

Hildebrandt, M. (2018). Algorithmic regulation and the rule of law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170355. https://doi.org/10.1098/rsta.2017.0355

Hildebrandt, M. (2019). Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law*, *20*(1), 83–121. https://doi.org/10.1515/til-2019-0004

Himmelreich, J. (2019). Ethics of technology needs more political philosophy. *Communications of the ACM*, *63*(1), 33–35. https://doi.org/10.1145/3339905

Himmelreich, J. (2023). Against "Democratizing AI." *AI & SOCIETY*, *38*(4), 1333–1346. https://doi.org/10.1007/s00146-021-01357-z

Hodgson, G. M. (2006). What Are Institutions? *Journal of Economic Issues*, *40*(1), 1–25. https://doi.org/10.1080/00213624.2006.11506879

Hodson, R., Martin, A. W., Lopez, S. H., & Roscigno, V. J. (2013). Rules don't apply: Kafka's insights on bureaucracy. *Organization*, *20*(2), 256–278. https://doi.org/10.1177/1350508412443581

Hoekstra, M., Chideock, C., & Van Veenstra, A. F. (2021). *Quick scan AI in de publieke dienstverlening II*. TNO.

Huber, C., & Munro, I. (2014). "Moral Distance" in Organizations: An Inquiry into Ethical Violence in the Works of Kafka. *Journal of Business Ethics*, *124*(2), 259–269. https://doi.org/10.1007/s10551-013-1865-1

Hultin, L., & Mähring, M. (2014). Visualizing institutional logics in sociomaterial practices. *Information and Organization*, *24*(3), 129–155. https://doi.org/10.1016/j.infoandorg.2014.05.002

Hutiri, W., Papakyriakopoulos, O., & Xiang, A. (2024). Not My Voice! A Taxonomy of Ethical and Safety Harms of Speech Generators. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 359–376. https://doi.org/10.1145/3630106.3658911

Hutiri, W. T. (2023). *Design Patterns for Detecting and Mitigating Bias in Edge AI* [Delft University of Technology]. https://doi.org/10.4233/UUID:C74CC90A-E55D-489C-B3BB-4C8C4A6DD7E6

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, *33*(3), 371–377. https://doi.org/10.1016/j.giq.2016.08.011

Janssen, M., Van Der Voort, H., & Van Veenstra, A. F. (2015). Failure of large transformation projects from the viewpoint of complex adaptive systems: Management principles for dealing with project dynamics. *Information Systems Frontiers*, *17*(1), 15–29. https://doi.org/10.1007/s10796-014-9511-8

Klijn, E. H., & Koppenjan, J. F. M. (2006). Institutional design: Changing institutional features of networks. *Public Management Review*, *8*(1), 141–160. https://doi.org/10.1080/14719030500518915

Kolkman, D. (2020). The usefulness of algorithmic models in policy making. *Government Information Quarterly*, *37*(3), 101488. https://doi.org/10.1016/j.giq.2020.101488

Kolko, J. (2010). Abductive Thinking and Sensemaking: The Drivers of Design Synthesis. *Design Issues*, *26*(1), 15–28. https://doi.org/10.1162/desi.2010.26.1.15

Konaté, S., & Pali, B. (2023). "You have to talk with us, not about us": Exploring the harms of wrongful accusation on those affected in the case of the Dutch 'childcare-benefit scandal.' *Journal of Victimology*, *16*, 139–164. https://doi.org/10.12827/RVJV.16.06

König, P. D. (2020). Dissecting the Algorithmic Leviathan: On the Socio-Political Anatomy of Algorithmic Governance. *Philosophy & Technology*, *33*(3), 467–485. https://doi.org/10.1007/s13347-019-00363-w

König, P. D., & Wenzelburger, G. (2020). Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics. *Government Information Quarterly*, *37*(3), 101489. https://doi.org/10.1016/j.giq.2020.101489

König, P. D., & Wenzelburger, G. (2021). The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it. *Technology in Society*, *67*, 101688. https://doi.org/10.1016/j.techsoc.2021.101688

Koppenjan, J., & Groenewegen, J. (2005). Institutional design for complex technological systems. *International Journal of Technology, Policy and Management*, *5*(3), 240–257. https://doi.org/10.1504/IJTPM.2005.008406

Koskela, L., Paavola, S., & Kroll, E. (2018). The Role of Abduction in Production of New Ideas in Design. In P. E. Vermaas & S. Vial (Eds.), *Advancements in the Philosophy of Design* (pp. 153–183). Springer International Publishing. https://doi.org/10.1007/978-3-319-73302-9_8

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020). Defining AI in Policy versus Practice. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 72–78. https://doi.org/10.1145/3375627.3375835

Kroes, P., Franssen, M., Poel, I. van de, & Ottens, M. (2006). Treating socio-technical systems as engineering systems: Some conceptual problems. *Systems Research and Behavioral Science*, *23*(6), 803–814. https://doi.org/10.1002/sres.703

Kroll, J. A. (2018). The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180084. https://doi.org/10.1098/rsta.2018.0084

Krygier, M. (2009). The Rule of Law: Legality, Teleology, Sociology. In G. Palombella & N. Walker (Eds.), *Relocating the Rule of Law*. Hart Publishing. https://doi.org/10.5040/9781472564634

Krygier, M. (2014). Rule of Law (and Rechtsstaat). In J. R. Silkenat, J. E. Hickey, & P. D. Barenboim (Eds.), *The Legal Doctrines of the Rule of Law and the Legal State (Rechtsstaat)* (Vol. 38, pp. 45–59). Springer International Publishing. https://doi.org/10.1007/978-3-319-05585-5_4

Krygier, M. (2016). The Rule of Law: Pasts, Presents, and Two Possible Futures. *Annual Review of Law and Social Science*, *12*(1), 199–229. https://doi.org/10.1146/annurev-lawsocsci-102612-134103

Krygier, M. E. J. (2007). The Rule of Law: An Abuser's Guide. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.952576

Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: Anatomy of a research project. *European Journal of Information Systems*, *17*(5), 489–504. https://doi.org/10.1057/ejis.2008.40

Künneke, R., Ménard, C., & Groenewegen, J. (2021). *Network Infrastructures: Technology meets Institutions* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108962292

Leonardi, P. M. (2012). Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Different? Do We Need Them? In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.), *Materiality and Organizing* (1st ed., pp. 24–48). Oxford University PressOxford. https://doi.org/10.1093/acprof:oso/9780199664054.003.0002

Leonardi, P. M. (2013). Theoretical foundations for the study of sociomateriality. *Information and Organization*, *23*(2), 59–76. https://doi.org/10.1016/j.infoandorg.2013.02.002

Leveson, N. G. (2012). *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press. https://doi.org/10.7551/mitpress/8179.001.0001

Lindgren, I. (2023). Ironies of Public Service Automation – Bainbridge Revisited. *Proceedings of the 24th Annual International Conference on Digital Government Research*, 395–404. https://doi.org/10.1145/3598469.3598514

Liu, H.-Y. (2018). The power structure of artificial intelligence. *Law, Innovation and Technology*, *10*(2), 197–229. https://doi.org/10.1080/17579961.2018.1527480

Lorenz, L. (2023). The hybrid work of public sector data scientists. *Journal of Professions and Organization*, *10*(3), 226–242. https://doi.org/10.1093/jpo/joad017

Lorenz, L., van Erp, J., & Meijer, A. (2022). Machine-learning algorithms in regulatory practice: Nine organisational challenges for regulatory agencies. *Technology and Regulation,* 2022, 1-11. https://doi.org/10.26116/techreg.2022.001

Love, T. (2000). Philosophy of design: A meta-theoretical structure for design theory. *Design Studies*, *21*(3), 293–313. https://doi.org/10.1016/S0142-694X(99)00012-5

Mackieson, P., Shlonsky, A., & Connolly, M. (2019). Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis. *Qualitative Social Work*, *18*(6), 965–980. https://doi.org/10.1177/1473325018786996

Mak, E., & Taekema, S. (2016). The European Union's Rule of Law Agenda: Identifying Its Core and Contextualizing Its Application. *Hague Journal on the Rule of Law*, *8*(1), 25–50. https://doi.org/10.1007/s40803-016-0022-1

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Marshall, P. (2022). Scandal at the Post Office: The Intersection of law, ethics and politics. *Digital Evidence and Electronic Signature Law Review*, 12–28. https://doi.org/10.14296/deeslr.v19i0.5395

Maxwell, J. A. (2012). *A realist approach for qualitative research*. Sage.

Mazzucato, M. & Collington, R. (2023). *The big con: how the consulting industry weakens our businesses, infantilizes our governments and warps our economies.* Allen Lane.

McGinnis, M. D. (2011). Networks of Adjacent Action Situations in Polycentric Governance: McGinnis: Adjacent Action Situations. *Policy Studies Journal*, *39*(1), 51–78. https://doi.org/10.1111/j.1541-0072.2010.00396.x

Meijer, A. (2025). Design science in public administration: Producing both situational interventions and generic knowledge. *Perspectives on Public Management and Governance*, *8*(1), 1–11. https://doi.org/10.1093/ppmgov/gvae012

Meijer, A., & Löfgren, K. (2015). The Neglect of Technology in Theories of Policy Change: *International Journal of Public Administration in the Digital Age*, *2*(1), 75–88. https://doi.org/10.4018/ijpada.2015010105

Meijer, A., Schäfer, M. T., & Branderhorst, M. (2019). Principes voor goed lokaal bestuur in de digitale samenleving: Een aanzet tot een normatief kader. *Bestuurswetenschappen*, *73*(4), 8–23. https://doi.org/10.5553/Bw/016571942019073004003

Meijer, A. & Thaens, M. (2021). The Dark Side of Public Innovation. *Public Performance & Management Review*, 44(1), 136-154. https://doi.org/10.1080/15309576.2020.17829

Meuser, M., & Nagel, U. (2009). The Expert Interview and Changes in Knowledge Production. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (pp. 17–42). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276_2

Meuwese, A. C. M., Parie, J., & Voogt, A. (2024). Hoe "algoprudentie" kan bijdragen aan een verantwoorde inzet van machine learning-algoritmes. *Nederlands Juristenblad*, *99*(10), 660–668.

Milchram, C., Märker, C., Schlör, H., Künneke, R., & Van De Kaa, G. (2019). Understanding the role of values in institutional change: The case of the energy transition. *Energy, Sustainability and Society*, *9*(1), 46. https://doi.org/10.1186/s13705-019-0235-y

Mingers, J., Mutch, A., & Willcocks, L. (2013). Critical Realism in Information Systems Research. *MIS Quarterly*, *37*(3), 795–802. https://doi.org/10.25300/MISQ/2013/37:3.3

Ministery of the Interior and Kingdom Relations. (2023). *Geactualiseerde Werkagenda Waardegedreven Digitaliseren 2024*.

Mitchell, M. (2021). *Why AI is Harder Than We Think* (No. arXiv:2104.12871). arXiv. http://arxiv.org/abs/2104.12871

Moe, T. M. (2006). Political Control and the Power of the Agent. *The Journal of Law, Economics, and Organization*, *22*(1), 1–29. https://doi.org/10.1093/jleo/ewj011

Moon, K., & Blackman, D. (2014). A Guide to Understanding Social Science Research for Natural Scientists. *Conservation Biology*, *28*(5), 1167–1177. https://doi.org/10.1111/cobi.12326

Mueller, M. L. (2025). It's just distributed computing: Rethinking AI governance. *Telecommunications Policy*, 102917. https://doi.org/10.1016/j.telpol.2025.102917

Mulligan, D. K., & Bamberger, K. A. (2018). Saving Governance-By-Design. *California Law Review*, *106*(3), 697–784. https://doi.org/10.15779/Z38QN5ZB5H

Mulligan, D. K., & Bamberger, K. A. (2019). *Procurement as Policy: Administrative Process for Machine Learning*. https://doi.org/10.15779/Z38RN30793

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, *16*(4), 317–342. https://doi.org/10.1111/j.1365-2575.2006.00221.x

Munro, I., & Huber, C. (2012). Kafka's mythology: Organization, bureaucracy and the limits of sensemaking. *Human Relations*, *65*(4), 523–543. https://doi.org/10.1177/0018726711430558

Mutch, A. (2013). Sociomateriality—Taking the wrong turning? *Information and Organization*, *23*(1), 28–40. https://doi.org/10.1016/j.infoandorg.2013.02.001

Newman, J., Mintrom, M., & O'Neill, D. (2022). Digital technologies, artificial intelligence, and bureaucratic transformation. *Futures*, *136*, 102886. https://doi.org/10.1016/j.futures.2021.102886

Nissenbaum, H. (2019). Contextual Integrity Up and Down the Data Food Chain. *Theoretical Inquiries in Law*, *20*(1), 221–256. https://doi.org/10.1515/til-2019-0008

Nonet, P., & Selznick, P. (1978). *Law & society in transition: Toward responsive law*. Transaction Publishers.

Noorman, M., & Swierstra, T. (2023). Democratizing AI from a Sociotechnical Perspective. *Minds and Machines*, *33*(4), 563–586. https://doi.org/10.1007/s11023-023-09651-z

Nouws, S., & Dobbe, R. (2024). The Rule of Law for Artificial Intelligence in Public Administration: A System Safety Perspective. In K. Prifti, E. Demir, J. Krämer, K. Heine, & E. Stamhuis (Eds.), *Digital Governance* (Vol. 39, pp. 183–208). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-639-0_9

Nouws, S., Janssen, M., & Dobbe, R. (2022). Dismantling Digital Cages: Examining Design Practices for Public Algorithmic Systems. In M. Janssen, C. Csáki, I. Lindgren, E. Loukis, U. Melin, G. Viale Pereira, M. P. Rodríguez Bolívar, & E. Tambouris (Eds.), *Electronic Government* (Vol. 13391, pp. 307–322). Springer International Publishing. https://doi.org/10.1007/978-3-031-15086-9_20

OECD. (2019). *Recommendation of the Council on Artificial Intelligence* (No. OECD/LEGAL/0449).

Oldenhof, L., Kersing, M., & Van Zoonen, L. (2024). Sphere transgressions in the Dutch digital welfare state: Causing harm to citizens when legal rules, ethical norms and quality procedures are lacking. *Information, Communication & Society*, 1–17. https://doi.org/10.1080/1369118X.2024.2358163

Olsen, J. P. (1997). Institutional Design in Democratic Contexts. *Journal of Political Philosophy*, *5*(3), 203–229. https://doi.org/10.1111/1467-9760.00032

Olsen, J. P. (2009). Change and continuity: An institutional approach to institutions of democratic government. *European Political Science Review*, *1*(1), 3–32. https://doi.org/10.1017/S1755773909000022

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, *3*(3), 398–427. https://doi.org/10.1287/orsc.3.3.398

Orlikowski, W. J. (2000). Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science*, *11*(4), 404–428. https://doi.org/10.1287/orsc.11.4.404.14600

Orlikowski, W. J. (2007). Sociomaterial Practices: Exploring Technology at Work. *Organization Studies*, *28*(9), 1435–1448. https://doi.org/10.1177/0170840607081138

Ossewaarde, R. (2023). Kafkaësk algoritmisch bestuur: Het toeslagenschandaal en de rol van irrationaliteit in de gemechaniseerde fraudejacht. *Tijdschrift over Cultuur & Criminaliteit*, *13*(1), 69–76. https://doi.org/10.5553/TCC/221195072023013001005

Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton University Press. https://doi.org/10.2307/j.ctt7s7wm

Ostrom, E. (2007). A diagnostic approach for going beyond panaceas. *Proceedings of the National Academy of Sciences*, *104*(39), 15181–15187. https://doi.org/10.1073/pnas.0702288104

Ostrom, V. (1973). *The intellectual crisis in American public administration*. University of Alabama Press.

Ozkaramanli, D., Karahanoğlu, A., & Verbeek, P.-P. (2022). Reflecting on Design Methods and Democratic Technology Development: The Case of Dutch Covid-19 Digital Contact-Tracing Application. *She Ji: The Journal of Design, Economics, and Innovation*, *8*(2), 244–269. https://doi.org/10.1016/j.sheji.2022.04.002

Passchier, R. (2020). Digitalisering en de (dis)balans binnen de trias politica. *Ars Aequi*, *69*(10), 916–927.

Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity*, *25*(4), 507–522. https://doi.org/10.3233/IP-200253

Peeters, R., & Schuilenburg, M. (2018). Machine justice: Governing security through the bureaucracy of algorithms. *Information Polity*, *23*(3), 267–280. https://doi.org/10.3233/IP-180074

Peeters, R., & Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture – The case of the Dutch civil registry's master data management system. *Government Information Quarterly*, *35*(2), 175–183. https://doi.org/10.1016/j.giq.2018.02.003

Peeters, R., & Widlak, A. C. (2023). Administrative exclusion in the infrastructure-level bureaucracy: The case of the Dutch daycare benefit scandal. *Public Administration Review*, *83*(4), 863–877. https://doi.org/10.1111/puar.13615

Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012). Design Science Research Evaluation. In K. Peffers, M. Rothenberger, & B. Kuechler (Eds.), *Design Science Research in Information Systems. Advances in Theory and Practice* (Vol. 7286, pp. 398–410). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29863-9_29

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big Data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration*, *35*(1), 24–44. https://doi.org/10.1177/0952076718780537

Pesch, U. (2021). Imaginaries of innovation: Turning technology development into a public issue. *Science and Public Policy*, *48*(2), 257–264. https://doi.org/10.1093/scipol/scab017

Pesch, U., & Vermaas, P. E. (2020). The Wickedness of Rittel and Webber's Dilemmas. *Administration & Society*, *52*(6), 960–979. https://doi.org/10.1177/0095399720934010

Peters, B. G. (2019). *Institutional theory in political science: The new institutionalism* (4th ed). Edward Elgar publishing.

Peters, B. G. (2020). Designing institutions for designing policy. *Policy & Politics*, *48*(1), 131–147. https://doi.org/10.1332/030557319X15579230420090

Pfadenhauer, M. (2009). At Eye Level: The Expert Interview — a Talk between Expert and Quasi-expert. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (pp. 81–97). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276_4

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A Taxonomy of Evaluation Methods for Information Systems Artifacts. *Journal of Management Information Systems*, *32*(3), 229–267. https://doi.org/10.1080/07421222.2015.1099390

PWC. (2024). *Onderzoek misbruik uitwonendenbeurs*.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

Raz, J. (1979). The Rule of Law and its Virtue. In J. Raz (Ed.), *The authority of law: Essays on law and morality* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198253457.003.0011

Redden, J., Dencik, L., & Warne, H. (2020). Datafied child welfare services: Unpacking politics, economics and power. *Policy Studies*, *41*(5), 507–526. https://doi.org/10.1080/01442872.2020.1724928

Reutter, L. (2022). Constraining context: Situating datafication in public administration. *New Media & Society*, *24*(4), 903–921. https://doi.org/10.1177/14614448221079029

Reymen, I. M. M. J., Hammer, D. K., Kroes, P. A., Van Aken, J. E., Dorst, C. H., Bax, M. F. T., & Basten, T. (2006). A domain-independent descriptive design model and its application to structured reflection on design processes. *Research in Engineering Design*, *16*(4), 147–173. https://doi.org/10.1007/s00163-006-0011-9

Rinta-Kahila, T., Someh, I., Gillespie, N., Indulska, M., & Gregor, S. (2022). Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems*, *31*(3), 313–338. https://doi.org/10.1080/0960085X.2021.1960905

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. https://doi.org/10.1007/BF01405730

Romme, A. G. L., & Meijer, A. (2020). Applying design science in public policy and administration research. *Policy & Politics*, *48*(1), 149–165. https://doi.org/10.1332/030557319X15613699981234

Ropohl, G. (1999). Philosophy of Socio-Technical Systems. *Society for Philosophy and Technology Quarterly Electronic Journal*, *4*(3), 186–194. https://doi.org/10.5840/techne19994311

Sætra, H. S., Borgebund, H., & Coeckelbergh, M. (2022). Avoid diluting democracy by algorithms. *Nature Machine Intelligence*, *4*(10), 804–806. https://doi.org/10.1038/s42256-022-00537-w

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed). SAGE.

Santoni De Sio, F., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philosophy & Technology*, *34*(4), 1057–1084. https://doi.org/10.1007/s13347-021-00450-x

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly*, *43*(3), 695–719. https://doi.org/10.25300/misq/2019/13747

Schwartz-Shea, P., & Yanow, D. (2012). *Interpretive research design: Concepts and processes*. Routledge.

Sclove, R. E. (1995). *Democracy and technology*. The Guilford Press.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., & Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. *NIPS*, 2494–2502.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 205395171773810. https://doi.org/10.1177/2053951717738104

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M. & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*(1), 37. https://doi.org/10.2307/23043488

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598

Selten, F., & Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly*, *41*(1), 101885. https://doi.org/10.1016/j.giq.2023.101885

Selznick, P. (1984). *Leadership in administration: A sociological interpretation*. University of California press.

Selznick, P. (1999). Legal Cultures and the Rule of Law. In M. Krygier & A. Czarnota (Eds.), *The Rule of Law after Communism: Problems and Prospects in East-Central Europe* (1st ed.). Routledge. https://doi.org/10.4324/9781315085319

Shelby, R., Rismani, S., Henne, K., Moon, Aj., Rostamzadeh, N., Nicholas, P., Yilla-Akbari, N., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 723–741. https://doi.org/10.1145/3600211.3604673

Siffels, L., Van Den Berg, D., Schäfer, M. T., & Muis, I. (2022). Public Values and Technological Change: Mapping how Municipalities Grapple with Data Ethics. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *New Perspectives in Critical Data Studies* (pp. 243–265). Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0_11

Simon, H. A. (1996). *The sciences of the artificial* (3. ed.). MIT Press. (Original work published 1969)

Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2022). Participation Is not a Design Fix for Machine Learning. *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–6. https://doi.org/10.1145/3551624.3555285

Smuha, N. A. (2021). Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea. *Philosophy & Technology*, *34*(S1), 91–104. https://doi.org/10.1007/s13347-020-00403-w

Spicer, M. W. (2019). What Do We Mean by Democracy? Reflections on an Essentially Contested Concept and Its Relationship to Politics and Public Administration. *Administration & Society*, *51*(5), 724–748. https://doi.org/10.1177/0095399718786881

Steinitz, C. (1995). Design is a Verb; Design is a Noun. *Landscape Journal*, *14*(2), 188–200. https://doi.org/10.3368/lj.14.2.188

Straub, V. J., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. *Government Information Quarterly*, *40*(4), 101881. https://doi.org/10.1016/j.giq.2023.101881

Suchman, L. (2006). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808418

Suchman, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society*, *10*(2), 20539517231206794. https://doi.org/10.1177/20539517231206794

Susser, D. (2016). Information Privacy and Social Self-Authorship. *Techné: Research in Philosophy and Technology*, *20*(3), 216–239. https://doi.org/10.5840/techne201671548

Taekema, S. (2021). Commitment to the Rule of Law: From a Political to an Organizational Ideal. In G. V. Vilaça & M. Varaki (Eds.), *Ethical Leadership in International Organizations* (1st ed., pp. 83–100). Cambridge University Press. https://doi.org/10.1017/9781108641715.003

Tangi, L., Rodriguez Müller, A. P., & Combetto, M. (2026). A Silent Partner: The Shadow Presence of Generative Artificial Intelligence in Public Administrations. In S. Hofmann, L. Danneels, R. Dobbe, A.-S. Novak, P. Parycek, G. Schwabe, V. Spitzer, & J. Ubacht (Eds.), *Electronic Participation* (Vol. 15978, pp. 69–86). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-02515-9_5

Thornton, P. H., & Ocasio, W. (2008). Institutional Logics. In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin, *The SAGE Handbook of Organizational Institutionalism* (pp. 99–128). SAGE Publications Ltd. https://doi.org/10.4135/9781849200387.n4

Timmermans, S., & Tavory, I. (2012). Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis. *Sociological Theory*, *30*(3), 167–186. https://doi.org/10.1177/0735275112457914

Tjeenk Willink, H. (2018). *Groter denken, kleiner doen. Een oproep* (16th ed.). Prometheus.

Torraco, R. J. (2016). Writing Integrative Literature Reviews: Using the Past and Present to Explore the Future. *Human Resource Development Review*, *15*(4), 404–428. https://doi.org/10.1177/1534484316671606

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, *53*, 190–202. https://doi.org/10.1016/j.apergo.2015.09.013

Van Aken, J. E. (2005a). Management Research as a Design Science: Articulating the Research Products of Mode 2 Knowledge Production in Management. *British Journal of Management*, *16*(1), 19–36. https://doi.org/10.1111/j.1467-8551.2005.00437.x

Van Aken, J. E. (2005b). Valid knowledge for the professional design of large and complex design processes. *Design Studies*, *26*(4), 379–404. https://doi.org/10.1016/j.destud.2004.11.004

Van Aken, J. E. (2007). Design Science and Organization Development Interventions: Aligning Business and Humanistic Values. *The Journal of Applied Behavioral Science*, *43*(1), 67–88. https://doi.org/10.1177/0021886306297761

Van Buuren, A., Lewis, J. M., Guy Peters, B., & Voorberg, W. (2020). Improving public policy and administration: Exploring the potential of design. *Policy & Politics*, *48*(1), 3–19. https://doi.org/10.1332/030557319X15579230420063

Van De Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines*, *30*(3), 385–409. https://doi.org/10.1007/s11023-020-09537-4

Van Der Voort, H. G., Klievink, A. J., Arnaboldi, M., & Meijer, A. J. (2019). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly*, *36*(1), 27–38. https://doi.org/10.1016/j.giq.2018.10.011

Van Krimpen, F., De Bruijn, H., & Arnaboldi, M. (2023). Machine learning algorithms and public decision-making. In T. Rana & L. Parker, *The Routledge Handbook of Public Sector Accounting* (1st ed., pp. 124–138). Routledge. https://doi.org/10.4324/9781003295945-12

Van Veenstra, A. F., Djafari, S., Grommé, F., Kotterink, B., & Baartmans, R. (2019). *Quick scan AI in de publieke dienstverlening*. TNO.

Van Zoonen, L. (2020). Data governance and citizen participation in the digital welfare state. *Data & Policy*, *2*, e10. https://doi.org/10.1017/dap.2020.10

Veale, M., & Brass, I. (2019). Administration by Algorithm?: Public Management Meets Public Sector Machine Learning. In M. Veale & I. Brass, *Algorithmic Regulation* (pp. 121–149). Oxford University Press. https://doi.org/10.1093/oso/9780198838494.003.0006

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Vermaas, P. E., & Houkes, W. (2006). Technical functions: A drawbridge between the intentional and structural natures of technical artefacts. *Studies in History and Philosophy of Science Part A*, *37*(1), 5–18. https://doi.org/10.1016/j.shpsa.2005.12.002

Vydra, S., & Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly*, *36*(4), 101383. https://doi.org/10.1016/j.giq.2019.05.010

Wagner, B. (2019). Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems. *Policy & Internet*, *11*(1), 104–122. https://doi.org/10.1002/poi3.198

Waldron, J. (2002). Is the Rule of Law an Essentially Contested Concept (Florida)? *Law and Philosophy*, *21*, 137–164.

Waldron, J. (2011). The rule of law and the importance of procedure. *Nomos*, *50*, 3–31. JSTOR.

Waldron, J. (2016). *Political political theory: Essays on institutions*. Harvard University Press.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, *3*(1), 36–59. https://doi.org/10.1287/isre.3.1.36

Warner, M. (2007). Kafka, Weber and organization theory. *Human Relations*, *60*(7), 1019–1038. https://doi.org/10.1177/0018726707081156

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., … Gabriel, I. (2022). Taxonomy of Risks posed by Language Models. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 214–229. https://doi.org/10.1145/3531146.3533088

Whelan, A. (2020). "Ask for More Time": Big Data Chronopolitics in the Australian Welfare Bureaucracy. *Critical Sociology*, *46*(6), 867–880. https://doi.org/10.1177/0896920519866004

Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco – Or was it something worse? *Australian Journal of Public Administration*, *80*(2), 340–360. https://doi.org/10.1111/1467-8500.12479

Whittaker, M. (2021). The steep cost of capture. *Interactions*, *28*(6), 50–55. https://doi.org/10.1145/3488666

Widder, D. G., & Nafus, D. (2023). Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society*, *10*(1), 20539517231177620. https://doi.org/10.1177/20539517231177620

Widlak, A. C., & Peeters, R. (2025). A theory of the infrastructure-level bureaucracy: Understanding the consequences of data-exchange for procedural justice, organizational decision-making, and data itself. *Government Information Quarterly*, *42*(2), 102021. https://doi.org/10.1016/j.giq.2025.102021

Widlak, A., Van Eck, M., & Peeters, R. (2020). Towards principles of good digital administration. In M. Schuilenburg & R. Peeters (Eds.), *The Algorithmic Society: Technology, Power, and Knowledge* (1st ed., pp. 67–84). Routledge. https://doi.org/10.4324/9780429261404

Wieringa, M. (2023). "Hey SyRI, tell me about algorithmic accountability": Lessons from a landmark case. *Data & Policy*, *5*, e2. https://doi.org/10.1017/dap.2022.39

Williamson, O. E. (1998). Transaction Cost Economics: How It Works; Where It is Headed. *De Economist*, *146*(1), 23–58. https://doi.org/10.1023/A:1003263908567

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136.

Wolswinkel, C. J. (2020). AR meets AI: Een bestuursrechtelijk perspectief op een nieuwe generatie besluitvorming. *Computerrecht*, *2020*(1), 22–29.

Yang, Y. (2022). Beyond Weber and Kafka: Conceptualizing a Critical Realist Model of Bureaucracy. *Administration & Society*, *54*(3), 500–521. https://doi.org/10.1177/00953997211024953

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, *12*(4), 505–523. https://doi.org/10.1111/rego.12158

Yeung, K. (2023). The New Public Analytics as an Emerging Paradigm in Public Sector Administration. *Tilburg Law Review*, *27*(2), 1–32. https://doi.org/10.5334/tilr.303

Young, M. M., Himmelreich, J., Bullock, J. B., & Kim, K.-C. (2021). Artificial Intelligence and Administrative Evil. *Perspectives on Public Management and Governance*, *4*(3), 244–258. https://doi.org/10.1093/ppmgov/gvab006

Zouridis, S., Van Eck, M., & Bovens, M. (2020). Automated Discretion. In T. Evans & P. Hupe (Eds.), *Discretion and the Quest for Controlled Freedom* (pp. 313–329). Springer International Publishing. https://doi.org/10.1007/978-3-030-19566-3_20

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, *38*(3), 101577. https://doi.org/10.1016/j.giq.2021.101577

Appendix A Diagnosis

A1 Observation guide consortium

Observatieformulier bijeenkomsten consortium

Algemeen

Datum:
Aanwezig:
Soort overleg:
Observator:

Observaties

1. Welke problemen identificeert het consortium?

2. Welke invulling krijgen de instrumenten?
   - Rollen, verantwoordelijkheden en afhankelijkheden
   - Inhoud

3. Welke problemen lossen de instrumenten op?

4. Tegen welke problemen lopen betrokkenen aan bij uitwerking/implementatie van instrumenten?

5. Wat is het perspectief van de verschillende organisaties op de problemen en instrumenten?

6. Welke socio-technische uitdagingen spelen er?

7. Overig:

A2 Topic guide interviews consortium

Topic guide interviews consortium

Introductie

- Waarom interview
  Toekomstverkenning, maar ook mijn eigen onderzoek → daar nadruk op ontwerpproces → maar synergie
- Doel interview
  Huidige situatie in [publieke organisatie] in kaart brengen → voornamelijk ontwikkeling (maar ook beetje gebruik) algoritmes → daarna kritische reflectie op de instrumenten die consortium in wil zetten om praktijk te verbeteren
- Opbouw interview
  Ik begin wat algemeen over ontwerpproces
  Daarna instrumenten
- Definities
  AI systeem → socio-technisch, machine, automatiseren, ondersteunen, versterken/verbeteren → vergelijkbaar met algoritme in consortium
- Consent

Algemene vragen

- Welke functie bekleed je bij de [publieke organisatie]?
- Hoe kijkt [publieke organisatie] aan tegen:
  AI systemen?
  Het gebruik van AI systemen?
  Het ontwerpen (ontwerpen breed) van AI systemen?

Inhoudelijke vragen - ontwerpproces

- In geval dat ze goede inkijk hebben op ontwerpproces:
  Kunnen jullie positief/negatief voorbeeld van AI systeem?
  Zouden jullie het ontwerpproces van dat AI systeem kunnen schetsen
- Als ze er minder bovenop zitten:
  Kunnen jullie een globaal beeld schetsen van hoe een ontwerpproces eruit ziet? Misschien hebben jullie een voorbeeld van een ontwikkeld/gebruikt AI systeem waarvan jullie de ontwikkeling van hebben meegekregen?
- Letten op:
  Actoren: Welke functies zijn er allemaal betrokken? (let op! Breed begrip ontwerper); Hoe vindt de afstemming met de politieke arena plaats? Welke stakeholders worden betrokken?
  Acties: Wat zijn de belangrijkste keuzes in het ontwerpproces? Welke elementen in het AI systeem vallen te onderscheiden?
  Posities: Wie maakt welke keuzes?
  Informatie: Hoe wordt informatie verzameld? Welke bronnen zijn er beschikbaar? Hoe wordt met onzekerheid omgegaan?
  Controle: Wat is de verantwoordelijkheidsstructuur?

Uitkomst: Het AI systeem
Kosten en baten: Hoe wordt AI systeem beoordeeld? Wanneer wordt het goed bevonden? Wanneer wordt er gekozen voor het gebruik van AI/algoritme?
Instituties
Fysieke condities
Attributen van ontwerparena: Welke cultuur in de organisatie? Hoeveel ervaring in de organisatie? Private partijen betrekken? Welke houding organisatie?

- Assumpties testen
  Fragmentatie
  Iteratie
  Politiek
  Privatisering
  Machtsverhoudingen
  Scope
  Kritisch
  Technocratisch/socio-technisch
- Wat moet er verbeterd worden aan ontwerpproces?
- Wat ligt die verbeteringen in de weg?

Inhoudelijke vragen - consortium (let op! Kritische reflectie. Laat me vooral ook weten hoe je er tegenaan kijkt en wat er tot nu toe (niet) is bereikt)

Algoritmeregister

- Reflectie op algoritmeregister?
- Welke plek neemt het register in in de levenscyclus van een systeem?
  Wanneer wordt het opgesteld?
  Wanneer moet het aangepast worden?
- Welk(e) functie/doel heeft het register?
  Kan het daarnaast ook nog andere doelen vervullen?
  Voor welke stakeholders bedoeld? Burger, controlerende actoren, interne organisatie? Allemaal?
  Rol in democratische rechtsstaat? Transparantie? Informatievoorziening?
  In wat voor ketens wordt register gebruikt? In andere woorden, waar valt het middel binnen andere procedures?
- Welke ruimte hebben de individuele publieke organisaties bij het invullen van het register?
  Bijvoorbeeld definitie algoritme?
  Menselijke tussenkomst?
  Impact?
- Interacties → gaat ook over organisatorische inbedding
  Op welk aspect van AI of algoritmetoepassingen is het register gericht?
  In hoeverre houdt het rekening met interacties? Bijvoorbeeld met ambtenaren, procedures, beleid, wetgeving? (nu vooral juridisch gericht) Of met andere registers?
  Hoe kan het register ingebed/geïmplementeerd worden in een organisatie? Wie is daar verantwoordelijk voor? Wat voor begeleiding is daar nodig? Coordinatie tussen verantwoordelijken/informatiehouders

Hoe te verankeren/betekenisvol te maken?
- Sturen van ontwikkeling?

Inkoopvoorwaarden
- Gebaseerd op document [publieke organisatie]
- Reflectie op inkoopvoorwaarden?
- Levenscyclus: welke fases worden allemaal gedekt door de voorwaarden?
  Er zit wat monitoring in? →maar hoe zit dat in de governance?
- Inkoopvoorwaarden zijn onderdeel van contract:
  Hoe worden andere samenwerkingen afgedekt? (denk bijv. aan uitvoering) → gaat ook over wisselwerking → gemeente lijkt verantwoordelijk voor alle interacties? Ook gerelateerd aan risicomanagement
  Welke vrijheid hebben publieke organisaties om de voorwaarden aan te passen? Definities? (denk aan nauwkeurig en correct functioneert; volgens een gemotiveerde aanpak ontwikkeld) Hoe moeten keuzes en ontwerpproces overlegt worden?
  Discretionaire ruimte van publieke organisatie?

Algoritmekader
- Reflectie op algoritmekader?
- Wat was het uitgangspunt van het kader? Verbetering huidige praktijk of verandering?
- Welke beperkingen/mogelijkheden geeft het maken van dit kader voor zo'n heterogene groep van organisaties?
- Doet het recht aan polycentrisch karakter ontwerpproces?

Bezwaarprocedure
- Reflectie op bezwaarprocedure?
- Institutionele inbedding
  Doel? → handreiking of meer nodig?
- Reikwijdte → alleen algoritme? → zou uitbreiding wenselijk/mogelijk zijn?
  Alleen bezwaarbehandelaar?

Einde interview
- Zijn er nog dingen die je kwijt wilt? Die we nog niet hebben behandelt maar wel belangrijk zijn?

A3 Topic guide interviews design teams

Interviewguide ontwerpteams

Introductie

- Mezelf
- Mijn onderzoek
- Doel interview: diagnose van het ontwerpproces van publieke algoritmische systemen
- Algoritmische systemen breed gedefinieerd; naast het technische aspect gaat dat ook om de mensen die het systeem gebruiken en alle wet- en regelgeving en beleid dat ervoor zorgt dat het systeem op een bepaalde manier gebruikt wordt
- Consent form
- Opbouw interview
  Eerst het ontwerpproces zoals bedoeld; vervolgens aan de hand van cases kijken hoe het in werkelijkheid gaat
  Herhaling vragen
- Heeft u nog vragen?

<concreet, voorbeelden, open vragen>

<u>Inleidende vraag</u>

1. Kunt u wat meer vertellen over uw functie bij de [publieke organisatie]?

<u>Beoogde uitkomst</u>

1. Wat is uw rol in de totstandkoming van publieke algoritmische systemen?
2. Wat is daarbij uw doel? Waar streeft u naar als [functie/rol]?

<u>Beoogde actie situatie</u>

1. Welke middelen heeft u tot uw beschikking om dat doel voor elkaar te krijgen?
2. Welke acties/activiteiten kunt u ondernemen om dat doel te bereiken?

<u>Beoogde interactie</u>

1. Wat is uw positie ten opzichte van andere betrokken actoren?
2. Tot waar reikt uw invloed op het algoritmische systeem als [functie/rol]?
3. In hoeverre zijn andere actoren afhankelijk van u als [functie/rol]?
4. Welke informatie heeft u tot uw beschikking en wat moeten andere actoren bij u aanleveren?
5. Op welke manier wordt u betrokken in de totstandkoming van publieke algoritmische systemen?

Gerealiseerde uitkomst

Ik zou u nu willen vragen of u een systeem kunt benoemen waarbij u bij de ontwikkeling van een algoritmisch systeem betrokken bent geweest. Belangrijk daarbij is dat het systeem op een later moment onvoorziene consequenties bleek te hebben. Aan de hand van dat voorbeeld wil ik bekijken hoe het ontwerpproces er in de werkelijkheid uitziet. Onvoorziene consequenties zijn in het geval van algoritmische systemen vaak uitsluiting en ondoorzichtigheid. Met uitsluiting bedoel ik dan bijvoorbeeld de situatie waarin een burger niet de publieke dienst krijgt waar hij of zij recht op heeft. Dit volgt dat vaak uit de disciplinerende werking van algoritmische systemen.

1. Kent u een situatie/systeem waarbij uiteindelijk bleek dat mensen werden uitgesloten of dat eigenlijk onduidelijk was wat het systeem precies deed?
2. Wat was het algemene doel in de totstandkoming van dit systeem?
3. Wat was uw rol en uw doel in de totstandkoming?

Gerealiseerde actie situatie

1. Welke functies waren er verder betrokken bij de ontwikkeling van dit systeem?
2. Kunt u een chronologisch overzicht van het totstandkomingsproces schetsen?
3. Welke keuzes zijn er gemaakt die terug te leiden zijn tot de onvoorziene consequenties?
4. Kunt u de dynamiek tussen de verschillende actoren schetsen?

Ideale situatie

1. Wat moet er volgens u veranderen aan het ontwerpproces? Waarom?
2. Hoe komt het dat het ontwerpproces er nog niet zo uitziet? Wat ligt die verbeteringen in de weg?

Afsluitende vraag

1. Gezien hetgeen we het afgelopen uur hebben besproken, is er dan nog iets dat we niet behandeld hebben maar volgens u wel belangrijk is?

Afsluiting

Ik wil u hartelijk bedanken voor uw tijd en de inzichtelijke antwoorden. De komende tijd zal ik nog verschillende collega's van u bij de [publieke organisatie] interviewen. Hetzelfde doe ik ook bij verschillende andere publieke organisaties. De resultaten van mijn onderzoek worden uiteindelijk gedeeld met de [publieke organisatie].

Dan rest nog het volgende: heeft u nog vragen?

A4 Code book interviews design teams

Process codes

Table 3.1 Process codes

Adapting	When designers change design goals, constraints, specifications etc to suit changing circumstances or conditions
Advising	A designer provides useful information or suggests/recommends an act or course of action to another actor or the design team
Aligning	Designers interact to bring their (possibly diverging) perspectives or interests in line or agreement (afstemmen)
Assessing	When a designer forms an opinion, judges, or decides about the value, size, quality, or importance of something in the design process
Brainstorming	When designers come together to generate new ideas or new knowledge
Coding	When a designer creates or edits computer code
Collecting	When a designer brings together data (or other forms of information) from different places
Consulting	A designer asks for or gets information (e.g., about needs, interests, facts) or advise from an involved or affected actor before making a design choice
Coordinating	When a designer organises or plans all the design activities and designers in a particular design process so that it results in a common action, movement, or condition
Convincing	A designer or design teams brings (as by argument) an involved or affected actor to belief, consent, or course of action; also persuading or lobbying
Deciding	When a designer chooses between several possibilities or fixes the course of outcome
Deliberating	To talk together and exchange problems, ideas, progress; either with the intention to reach a decision or to explore issues
Delivering	A designer provides an artefact to another actor in the design process
Describing	A designer provides more detail on how something is done or of what something is like (beschrijven; in kaart/beeld brengen)
Developing	When a designer brings something into existence through intellectual or physical effort
Endorsing	When a designer/decision-maker formally or officially approves or supports a design activity or design choice (goedkeuren; toestemming geven)
Engaging	A designer or design team interests involved or affected actors in something and keeps them thinking about the design process
Escalating	When a design choice or other sort of decision is passed on to a superior (political) decision-making level
Examining	When a designer examines the implications of a public algorithmic system
Explaining	A designer or experts makes something (e.g., the working of an algorithmic application) clear or easy to understand by describing or giving information about it
Exploring	When a designers explores the options for using an algorithmic application
Helping	A designers assists or supports another designer in their tasks or activities
Implementing	When an algorithmic systems is put into operation

Informing	A designer tells or presents another involved or affected actor(s) about the design process or design choices
Integrating/ embedding	When an algorithmic application is integrated or embedded in a broader ICT system or the organisation
Involving	When a designer or design teams includes someone in a design activity
Joining	When a designer actively gets involved in a particular design activity
Learning	When a designer or design team gains knowledge, understanding, or skills through study, instruction, or experience
Measuring	When a designer estimates, appraises, or discovers the (exact) size, value or amount of something
Outsourcing	When the public organisation pays or employs another organisation to do (part of) the design activities or purchases a product from another organisation (uitbesteden)
Policy-making	When a designer drafts the policy in which an algorithmic application is used
Preparing	When a designer makes a dataset ready for using it in an algorithmic model
Prioritising	When a designer lists, arranges or rates, for example, activities or goals in order of importance
Questioning	When a designers (critically) examines, doubts, or disputes steps, goals or choices in the design process
Recording	When a designers keeps information (about design activities or design choices) by writing it down and storing it so that people can refer to it later
Scoping	When a designer determines the problem or solution space
Steering	When a designer sets and holds to a course in order to make actors or activities in a particular way
Substantiating	A designer gives reasons for a decision, idea, belief, or opinion (onderbouwen)
Testing	When a designer checks (through analysis or diagnosis) whether an artefact or process is effective, efficient, safe, works correctly, etc.
Training	When a designer creates or improves an algorithmic model by supplying it with data
Translating	When a designers transfers an idea from a specific discipline, field, form, or appearance to one that is compatible with their own discipline, field, form
Undertaking	When a designer takes responsibility for and begin doing a design activity in response to a question or assignment of another designer
Waiting	When a designer stays in one place until other designers finish design tasks that are contingent to the tasks of the designer

Descriptive codes

Table A3.2 Descriptive codes

Awareness	References to (the need for) understanding of algorithm-related subjects or situation and the realisation of (the existence of) algorithm-related consequences among individual or groups of actors/designers
Challenge-driven	References to design processes that originated in a problem or challenge
Collaboration	References different actors or designers working together in the design process
Complaints	References to users that articulate that something is wrong or not satisfactory in the public algorithmic system
Compliance	References to actors that ensure conformity of algorithmic systems or processes in fulfilling official requirements
Contribution	References to what a designer thinks they contribute to the whole design process
Demand-driven	References to design processes that originated from a question, need, or demand
Dependency	References to situations in which a designer needs, for example an action or information, from another individual or organisation before they can continue their own activities
Design choices	References to choices between alternatives for the algorithmic system made during or before the design process
Design output	References to the artefacts that are delivered by designers
Documentation	References to the role of documentation in which steps in the design process or design choices are recorded
Ethics	References to the role of ethics as a discipline in the design process or to ethical aspects of algorithmic systems
Experience	References to actors that have gained practical knowledge or skills because of recurrently performing particular (design) activities
Expertise	References to the skills of designers in a particular area
External party	References to actors that play a role in the design process, but are not employed or part of the public organisation
Goals	References to the ends or aims that a design team or public organisation wants to achieve by implementing a public algorithmic system
Incentive	References to external event or occurrence that resulted in a change in design practices
Influence	References to extent of influence that an individual designer has over actions of other designers or over the whole design process
Informatiegestuurd werken	References to a broader policy or goal within a public organisation to use data-driven, digitalisation, and automation approaches in public policy and services
Involvement	References to actors being engaged in or committed to the design process
Iteration	Reference to repetition of design activities to improve an artefact – for example, in response to feedback from users
Knowledge	References to actors that are knowledgeable on a specific topic or actors that lack knowledge on a specific topic
Language	References to the discipline-related language that actors use, but that might bring problems in communicating with other disciplines
Legal	References to written, statutory and formal institutions, such as laws and regulations

Life cycle	References to (transfers between) stages, phases, or steps in the design process or during the existence of the public algorithmic system
Maturity	References to degree of competence of a public organisation regarding design processes or deploying algorithmic systems
Motivation	References to the aims of an individual designer in a design process
Organisation size	References to the size, scale, magnitude, or proportion of a public organisation
Organisational culture	References to the culture within the public organisation
Organisational structure	References to how the public organisation is structured
Perspective	References to how other designers consider or regard a particular topic
Policy	References to policy(ies) within the public organisation
Policy instruments	Instruments that public organisations use to ensure responsible development and use of public algorithmic systems
Politics	References to bodies that engage in political decision-making
Pressure	References to other actors that try to make the designer(s) to do something
Publicity	References to media outlets reporting about a specific algorithmic system or about consequences of algorithmic systems in general
Quality assurance	References to the quality of an algorithmic system or one of its components
Question/problem formulation	References to efforts by designers to elicit or elucidate the question or problem in a particular department or organisation
Resources	References to the availability of means that a public organisation can use; mostly related to financial means or staff capacity
Responsibility	References to the duty of a particular designer or user to take care of an activity
Risks	References to possible losses or dangerous situations resulting from deploying a public algorithmic system
Role	References to the positions and functions of a particular actor in the design process
Safeguards	References to precautionary measures to deal with risks in the design process
Sensitive	References to a public algorithmic system that has a delicate position because of, for example, controversy
Situational	References to the context dependency of the design process
Sprint/agile	References to a dominant approach or form of organisation in software development using terms such as sprint, agile, scrum, dailies
Staff changes	References to designers leaving or entering the design team during the design process
Supply-driven	References to design processes that originated from stand-alone experiments or innovations in the organisation – e.g., in an ICT team or data science lab
System on the radar	References to how actors within a public organisation keep up to date about all algorithmic systems that are being developed or used in that organisation
Systemic level	References to the dependencies of algorithmic applications on other components or the role of such application in the organisation
Technical artefact	References to technologies and/or materials used as tools in the design process
Technological limits	References to the boundaries of current technological knowledge or developments

Time	References to the notions of duration or pace in design activities, design practices, and design processes
Trade-off	References to balancing of factors all of which are not attainable at the same time
Understanding	References to achieved understanding (of the algorithmic application, or other subjects) among other actors
Use of a system	References to how the system is or was used in practice

Value codes

Table A3.3 Value codes

Attitudes	Beliefs
algorithms as execution instead of integral to policy	also bias and discrimination in manual processing
another thing to fill in	an algorithm is a means to an end
chance of success	compliance (privacy) goes over the top
choosing the way of least resistance	contribution means that your expertise is considered
concerns	data science brings added value in the sense that it gives us more insights
discrimination is related to questions of proportionality	data scientist needs domain knowledge
enthousiasm	dependence on willingness or interest of individual
explore the possibilities of algorithms	design choices are based on popular opinion, framing and window dressing
external algorithms are not our responsibility	design choices are not always well thought-out
guarded/cautious	designing algo is a multidisciplinary effort
I lack the knowledge	developed too fast
involvement/drive	educating public servants increases awareness
lack of awareness	inflexibility
lack of urgency	inspectors mostly base their decisions on experience
privacy officers ask difficult questions	integrate data science and BI
professionals are working on the technology	interest in topic makes collaboration and understanding easier
resistance	involving legal officers from the beginning makes the process more smooth
start new projects but not evaluating them	involving the right people can suddenly result in performance
support	it is a difficult message when design choices have been made and you question those choices
unknowledgeable others	it is most important to be able to explain your choices
We cannot fail	management is not capable or lacks knowledge to tackle the problems
we need to adapt as organisation and that needs time	management strategically makes certain design choices
we should be more like businesses/private sector	not our responsibility
we use the word algorithm too easy	organisations have an interest in blowing up/exagurating their use of algorithmic systems
	politics should only be able to change a certain percentage of a policy
	public servants act based on good intentions

	reality is always more complex than rational idea of process
	seeing importance as driver for processes
	the important questions are not asked or not answered
	too much fuzz
	unclear affordances of data-driven applications
	we cannot finish project

A5 Preliminary results shared in workshops

Slides presented to the workshop participants

Algemene indruk

Kenschets

- Dynamisch
- Zoektocht naar waarborgen en samenwerking
- Projectmatig werk dat vaak buiten dagelijkse werkzaamheden van betrokkenen valt

TU Delft 5-6-2023

Inrichting ontwerpproces

Er bestaat een **discrepantie** tussen het **structureren** van het ontwerpproces en het **iteratieve karakter** van dat ontwerpproces

Groot aanbod van **instrumenten** maar hun **waarborgfunctie** blijft onduidelijk

Ontwerpprocessen hebben een **lange tijdsduur** waarbinnen veel **personeelswijzigingen** plaats vinden

TU Delft 5-6-2023

Organisatievraagstukken

Ontwerpteams zijn **afhankelijk** van een groot aantal actoren in de organisatie

Silo's in de organisatie zorgen voor **worstelingen** in het ontwerpproces

Correcties op **systeemniveau** vinden niet plaats of gaan traag

TU Delft 5-6-2023

Individuele actoren

Politieke vertaling is vaak niet aanwezig, tenzij er sprake is van politieke druk

Weerstand, gevoel van onbekwaamheid, angst staat **samenwerking** in de weg

Persoonlijk contact en samenwerking zijn cruciaal voor een soepel verloop van het ontwerpproces

TU Delft 5-6-2023

Technologie

Publieke organisaties hebben een **smalle focus op technologie**

Het **scrumritme** staat op **gespannen voet** met het ritme van andere organisatieonderdelen

Een **duidelijke vraagstelling, doelstelling of afbakening** van projecten is niet aanwezig of wordt uitgevoerd door technische afdelingen

TU Delft

5-6-2023

Appendix B Create and assess

B1 Changes after iteration with colleagues

We conducted two trials of the design process simulation with five colleagues. Two colleagues participated in the first trial. These colleagues were two PhD candidates in the field of algorithmic fairness and institutionalisation of AI. The second trial was attended by three fellow PhD candidates. These colleagues were not necessarily involved in research on designing AI in public organisations but were all member of the ICT section at the TPM faculty of Delft University of Technology.

The two trials resulted in a final version of the protocol of the simulation. This appendix discusses the changes made to the protocol based on input from the trial sessions. The trial sessions provided insight into what elements in the simulation needed more explanation or a clearer instruction. In addition, we made changes to the translation of design principles to institutional interventions.

Based on the trial sessions, we:

In general

1. … provided participants with an indication of the time they were expected to spend on steps in the simulation. The time indication provides participants with a general idea of what a task or step entails.
2. … defined the system boundaries of the algorithmic system more clearly. In the trial, participant tended to discuss the underlying policy. However, this policy was considered fixed in the simulation (as it was formulated by national government, the municipality had to focus on execution of the policy). To make sure that simulation participants would focus on the execution of the policy, we explicitly instructed participants to regard the underlying policy as fixed.

Representative designers

3. … changed the template formulation of design goals and design constraints. The formulation of these goals and constraints confused participants of the two trial sessions. More specifically, they did not see the difference between goals and constraints. Therefore, we clarified the description of design goals and design constraints.
4. … provided representative designers with the opportunity to think about design goals

and design constraints before they would start the deliberation about the goals and constraints. This provided all participants to meaningfully take part in the deliberation and prevented that one or more participants would dominate the discussion.

5. … divided the formulation of and prioritisation of design goals and design constraints. In the trials, participants were instructed to identify goals and constraints and prioritise them in one go. However, this resulted in the prioritisation being of less importance to the participants. Therefore, in the actual simulation, we first asked representative designers to formulate the goals and constraints. Only when they were done with the formulation, we asked them to prioritise the goals and constraints.

System-level designers

6. … divided session 2 in five steps. Especially in the first session, the design exercise was too broad. Participants felt overwhelmed by the assignment they had to complete in session 2.
7. … asked system-level designers to first formulate three high-over design alternatives before elaborating the alternatives in socio-technical specifications. The trial sessions showed that thinking of three system alternatives was not intuitive. Accordingly, participants started with one alternative and specified that one alternative. This leaves little time to think of other design alternatives but also limits creativity in coming up with distinct design alternatives. Another option would have been to ask system-level designers to draft one system map, and ask them to identify or highlight alternatives for specific system components. This happened in the first trial session.
8. … changed the identification of possibilities for arbitrary conduct in the socio-technical specification. In the first version of the simulation, this identification was based on hazard analysis as described in system safety (see Leveson, 2012). This analysis turned out to be too difficult and too elaborate to conduct in the time frame of the simulation. Therefore, we simplified the identification. In the final simulation, manifestations of arbitrary conduct were presented to the system-level designers. Based on this list, the designers could identify possibilities of arbitrary conduct in their own socio-technical specifications. Consequently, the designers did not perform a full assessment of their own specifications but conducted a guided scrutinization of their own design choices.

B2 Questionnaires before simulation

B2.1 Questionnaire before simulation representative designers

Vragenlijst quasi-experiment "ontwerpprocessen van publieke algoritmische systemen"

Beste deelnemer,
Het doel van deze vragenlijst is om de huidige situatie betreffende het ontwerpen van publieke algoritmische systemen in uw publieke organisatie in kaart te brengen. Na het quasi-experiment wordt u opnieuw gevraagd een korte vragenlijst in te vullen. Onderstaande vragenlijst bestaat uit 13 open vragen. Geef bij elk antwoord een korte toelichting. Invullen neemt ongeveer 20 minuten in beslag.

Hierbij wil ik u alvast hartelijk bedanken voor uw deelname. Ik waardeer uw tijd en antwoorden.

Een deel van de vragen bevatten termen die niet vanzelf spreken of door verschillende organisaties anders worden gedefinieerd. In deze vragenlijst worden de volgende definities gebruikt:

In een ***publiek algoritmisch systeem*** wordt een algoritmisch model gebruikt om processen en/of besluitvorming in het openbaar bestuur te ondersteunen, te verbeteren of te automatiseren. Dit algoritmisch model bestaat uit software dat voorspellingen doet, classificaties maakt of uitvoer genereert op basis van vooropgestelde regels of door patronen in grote datasets te ontdekken. Het algoritmisch model heeft op zichzelf geen functie. Het model krijgt pas een functie als de uitvoer wordt gebruikt door een ambtenaar en wanneer het model onderdeel is van beleid en/of wet- en regelgeving. Dit betekent dat bijvoorbeeld ook de werkinstructies voor de gebruikers van het algoritmisch model onderdeel vormen van het systeem.

Met ***ontwerpen*** bedoelen we in deze vragenlijst alle activiteiten die worden ondernomen om tot een gewenst algoritmisch systeem te komen. Het gaat hier dus niet alleen om de ontwikkeling van de software. Ook het stellen van kaders, het opstellen van werkinstructies en het adviseren over juridische aspecten van een algoritmisch model zijn voorbeelden van activiteiten die onder ontwerpen vallen.

Wanneer er gevraagd wordt naar iets wat u niet meemaakt of herkent, wil ik u vragen dat aan te geven.

Inleidende vragen

1. Hoe lang bent u raadslid?

Antwoord:

2. Hoe vaak is een publiek algoritmisch systeem onderwerp van gesprek geweest in uw werk als raadslid en op welke manier?

Antwoord:

3. Op wat voor manier houdt u zich bezig met publieke algoritmische systemen die worden gebruikt in uw gemeente?

Antwoord:

Vragen huidige situatie

Sturing op het ontwerpen van publieke algoritmische systemen

4. Op welke manier geeft u als raadslid sturing aan het ontwerpen van publieke algoritmische systemen binnen de gemeente?

Antwoord:

5. Hoe zou u uw invloed op het ontwerpen van publieke algoritmische systemen karakteriseren? Via welke weg heeft u die invloed?

Antwoord:

Onderling begrip

6. Welke informatie krijgt u van ambtenaren over publieke algoritmische systemen?

Antwoord:

7. In hoeverre maakt de informatie die aan de raad wordt verstrekt over publieke algoritmische systemen inzichtelijk hoe die systemen werken?

Antwoord:

Motivering achter ontwerpkeuzes

8. In welke mate heeft u inzicht in de argumentatie achter keuzes die ambtenaren maken in het ontwerpen van publieke algoritmische systemen?

Antwoord:

Zorgvuldigheid bij het komen tot ontwerpkeuzes

9. In welke mate heeft u inzicht in de gevolgen voor burgers die publieke algoritmische systemen (kunnen) hebben?

Antwoord:

In staat zijn om keuzes te maken

10. In hoeverre bent u in staat om uw besluitvormende en controlerende taken als raadslid uit te voeren als het gaat over publieke algoritmische systemen?

Antwoord:

Reflectie

11. Hoe verloopt de opvolging met betrekking tot publieke algoritmische systemen nadat daar in de raad over gedebatteerd is? Is er sprake van terugkoppeling? Is er sprake van politiek debat?

Antwoord:

Vragen over gewenste situatie

12. Wat zou volgens u moeten veranderen aan de rol van de raad in het ontwerpen van publieke algoritmische systemen?

Antwoord:

13. Wat zou volgens u moeten veranderen aan de rol van de ambtenarij in het ontwerpen van publieke algoritmische systemen?

Antwoord:

Dit is het einde van de vragenlijst

B2.2 Questionnaire before simulation system-level designers

Vragenlijst quasi-experiment "ontwerpprocessen van publieke algoritmische systemen"

Beste deelnemer,
Het doel van deze vragenlijst is om de huidige situatie betreffende het ontwerpen van publieke algoritmische systemen in uw publieke organisatie in kaart te brengen. Na het quasi-experiment wordt u opnieuw gevraagd een korte vragenlijst in te vullen. Onderstaande vragenlijst bestaat uit 13 open vragen. Geef bij elk antwoord een korte toelichting. Invullen neemt ongeveer 20 minuten in beslag.

Hierbij wil ik u alvast hartelijk bedanken voor uw deelname. Ik waardeer uw tijd en antwoorden.

Een deel van de vragen bevatten termen die niet vanzelf spreken of door verschillende organisaties anders worden gedefinieerd. In deze vragenlijst worden de volgende definities gebruikt:

In een ***publiek algoritmisch systeem*** wordt een algoritmisch model gebruikt om processen en/of besluitvorming in het openbaar bestuur te ondersteunen, te verbeteren of te automatiseren. Dit algoritmisch model bestaat uit software dat voorspellingen doet, classificaties maakt of uitvoer genereert op basis van vooropgestelde regels of door patronen in grote datasets te ontdekken. Het algoritmisch model heeft op zichzelf geen functie. Het model krijgt pas een functie als de uitvoer wordt gebruikt door een ambtenaar en wanneer het model onderdeel is van beleid en/of wet- en regelgeving. Dit betekent dat bijvoorbeeld ook de werkinstructies voor de gebruikers van het algoritmisch model onderdeel vormen van het systeem.

Met ***ontwerpen*** bedoelen we in deze vragenlijst alle activiteiten die worden ondernomen om tot een gewenst algoritmisch systeem te komen. Het gaat hier dus niet alleen om de ontwikkeling van de software. Ook het stellen van kaders, het opstellen van werkinstructies en het adviseren over juridische aspecten van een algoritmisch model zijn voorbeelden van activiteiten die onder ontwerpen vallen.

Ook wanneer er gevraagd wordt naar iets wat u niet meemaakt of herkent, wil ik u vragen dat aan te geven.

Inleidende vragen

1. Hoe lang bent u werkzaam als ambtenaar met een focus op algoritmes of digitalisering/ICT in het algemeen?

Antwoord:

2. Hoe vaak heeft u een bijdrage geleverd aan het ontwerpen van een publiek algoritmisch systeem?

Antwoord:

3. Wat is uw rol bij de totstandkoming of aanpassing van publiek algoritmisch systemen?

Antwoord:

Vragen huidige situatie

Sturing op het ontwerpen van publieke algoritmische systemen

4. Op welke manier worden de keuzes die u maakt tijdens het ontwerpen van publieke algoritmische systemen gestuurd door het politieke debat in raad?

Antwoord:

5. Hoe zou u uw invloed op het ontwerpen van publieke algoritmische systemen karakteriseren? Via welke weg heeft u die invloed?

Antwoord:

Onderling begrip

6. Welke informatie over publieke algoritmische systemen wordt met de raad gedeeld?

Antwoord:

7. In welke mate kunt u de raad meenemen in het ontwerpen van publieke algoritmische systemen?

Antwoord:

Motivering achter ontwerpkeuzes

8. Op welke manier komt de argumentatie achter keuzes die worden gemaakt tijdens het ontwerpen van publieke algoritmische systemen tot stand?

Antwoord:

Zorgvuldigheid bij het komen tot ontwerpkeuzes

9. Op welke manier worden de (mogelijke) gevolgen van publieke algoritmische systemen voor burgers in kaart gebracht?

Antwoord:

In staat zijn om keuzes te maken

10. In hoeverre bent u in staat om politieke wensen en grenzen in de raad te vertalen naar publieke algoritmische systemen tijdens het ontwerpen?

Antwoorden:

Reflectie

11. Welke processen om fouten in publieke algoritmische systemen op te sporen worden gebruikt in de gemeente?

Antwoord:

Vragen over verbeteringen

12. Wat zou volgens u moeten veranderen aan de rol van de ambtenarij in het ontwerpen van

publieke algoritmische systemen?

Antwoord:

13. Wat zou volgens u moeten veranderen aan de rol van de raad in het ontwerpen van publieke algoritmische systemen?

Antwoord:

Dit is het einde van de vragenlijst

B3 Simulation: slides and other materials

B3.1 Session 1

In session 1, the instructor followed the following structure to guide the representative designers through the simulation:

1. Explain the experiment to politicians
2. Present the problem formulation to politicians
3. Explain what is meant by 'design goal' and 'design constraint'
4. Ask politicians to list goals and restrictions on their own
5. Ask politicians to deliberate on a shared list of goals and constraints
6. Ask politicians to prioritise both the goals as well as the constraints
7. Ask politicians to write a clarification of the lists of goals and constraints to the system-level designers

The following presentation slides were used for the instructions:

Quasi-experiment
ontwerpproces van publieke
algoritmische systemen

Slide 1

Instructies

- In deze sessie gaan jullie het normatief kader van een systeem vaststellen op basis van een probleemformulering
- Een groep ontwerpers zal daarna dat gebruiken om een systeem ontwerpen
- Bedenk daarbij dat het probleem zich leent voor een publiek algoritmische systeem behoort tot de mogelijkheden
- Ik geef instructies maar ook niet te veel
- Blijf binnen de grenzen van wat de gemeente mag en kan doen

Slide 2

Probleemformulering

Probleem

Het Rijk heeft gemeentes gevraagd om de verstrekking van een energietoeslag uit te voeren. Deze energietoeslag is gericht op huishoudens die hun energierekening niet meer kunnen betalen door gestegen prijzen. Het is de taak van de gemeente om te zorgen dat de alle daarvoor in aanmerking komende huishoudens de toeslag ontvangen. De gemeente heeft alleen geen goed beeld van de huishoudens die aanspraak kunnen maken op de toeslag.

Middelen

- Financiële middelen om ervoor te zorgen dat de huishoudens in beeld komen
- Financiële middelen om het toeslagenprogramma uit te voeren
- Data over inkomens, huishoudensgrootte, leeftijden, gebruik van andere toeslagen en/of uitkeringen
- Inzicht in het gemiddelde energieverbruik van huishoudens, energieprijstrends, trends in vraag en aanbod van energie
- Een beleidsdocument van de rijksoverheid waarin staat welke type huishoudens in aanmerking komen en hoeveel toeslag elk type huishouden zal ontvangen

Slide 3

Ontwerpdoelen en ontwerprestricties (1)

Ontwerpdoel
- Beschrijft wat het te ontwerpen systeem zou moeten bereiken
- Vorm: Het te ontwerpen systeem beoogt …

Ontwerprestricties
- Beschrijft de grenzen aan het systeem
- Vorm: Het te ontwerpen systeem kan niet …

Slide 4

Ontwerpdoelen en ontwerprestricties (2)

1. Individueel: maak een inventarisatie van doelen en restricties
2. Gezamenlijk: beschrijf alle doelen en restricties die jullie belangrijk vinden

Slide 5

Prioritering

De ontwerpers gaan de doelen en restricties gebruiken als leidraad voor hun ontwerp. Door de doelen en restricties te prioriteren, helpen jullie hen bij het maken van ontwerpkeuzes.

1. Leg het belangrijkste doel bovenaan en leg de andere doelen op basis van prioriteit onder elkaar
2. Doe hetzelfde voor de restricties

Slide 6

Toelichting

1. Schrijf een toelichting bij de lijsten met doelen en restricties

Slide 7

B3.2 Sesssion 2

In session 2, the instructor followed the following structure to guide the representative designers through the simulation:

1. Explain the experiment to designers
2. Present the problem formulation to designers
3. Present the normative framework to designers
4. Introduce the five design steps

The five steps were structured as follows:

Step 1: Refine the design assignment
1. Ask designers to study the problem formulation and normative framework
2. Ask designers to formulate a design assignment that they can work with

Step 2: Produce design alternatives
1. Ask designers to identify three alternative ways (of which at least two include an algorithmic applications) to fulfil the design assignment
2. Ask designers to describe the alternatives as a process in two to three sentences

Step 3: Draw up system maps
1. Explain system level design, system map, and system components (including a simple and a complex example of a system map
2. Ask designers to draw up a system map for all three design alternatives by: first identifying the system components; second, illustrating the interactions between components; and, finally, signifying components for which there are different options

Step 4: Reflect on the designs
1. Present the four manifestations of arbitrary use of power
2. Ask designers to indicate the manifestations in their system maps and to determine whether the manifestations are problematic
3. Ask designers to redesign the system maps by adding, changing, or deleting components in order to prevent, mitigate, or correct possibilities for arbitrary conduct

Step 5: Formulate an advice to politicians
1. Ask designers to go back to the problem formulation, the normative framework, and their own design assignment
2. Ask designers to write down:
   An explanation to their system maps
   A description of the most critical design choices and underlying considerations
   Argumentation for a preferred alternative
3. Ask designer to write down an individual opinion on the advice that will be shared with the politicians

The following presentation slides were used for the instructions:

Quasi-experiment ontwerpproces van publieke algoritmische systemen

Slide 1

Instructies

- In deze sessie gaan jullie een systeemkaart maken, kritieke ontwerpkeuzes identificeren en ontwerpalternatieven analyseren
- Een groep raadsleden heeft daarvoor een normatief kader opgesteld op basis van een probleemformulering
- Bedenk daarbij dat het probleem zich leent voor een publiek algoritmische systeem
- Ik geef instructies maar ook niet te veel
- Blijf binnen de grenzen van wat de gemeente mag en kan doen

Slide 2

Probleemformulering

Probleem

Het Rijk heeft gemeentes gevraagd om de verstrekking van een energietoeslag uit te voeren. Deze energietoeslag is gericht op huishoudens die hun energierekening niet meer kunnen betalen door gestegen prijzen. Het is de taak van de gemeente om te zorgen dat de alle daarvoor in aanmerking komende huishoudens de toeslag ontvangen. De gemeente heeft alleen geen goed beeld van de huishoudens die aanspraak kunnen maken op de toeslag.

Middelen

- Financiële middelen om ervoor te zorgen dat de huishoudens in beeld komen
- Financiële middelen om het toeslagenprogramma uit te voeren
- Data over inkomens, huishoudensgrootte, leeftijden, gebruik van andere toeslagen en/of uitkeringen
- Inzicht in het gemiddelde energieverbruik van huishoudens, energieprijstrends, trends in vraag en aanbod van energie
- Een beleidsdocument van de rijksoverheid waarin staat welke type huishoudens in aanmerking komen en hoeveel toeslag elk type huishouden zal ontvangen

Slide 3

Kader vanuit raadsleden

Ontwerpdoelen

Het te ontwerpen systeem beoogt…

1. …toepassing van menselijke maat in ontwerpproces en toepassing van het systeem
2. …het doel van compenseren van huishoudens met een laag inkomen voor een hoge energieprijs
3. …efficiënt, effectief en meerwaarde hebben voor de organisatie/betrokkenen
4. …inzichtelijk te zijn voor belanghebbenden, o.a. ten behoeve van bezwaar & beroep, AVG-recht

Ontwerprestricties

Het te ontwerpen systeem kan niet…

1. …rechten schenden van de betrokken inwoners, niet limitatief: discriminatie, privacy
2. …onaanvaardbare/onevenredige risico's met zich mee brengen voor inwoners, bijvoorbeeld maar niet uitsluitend door het uitvoeren van een DPIA, HRIA, algoritmes impact assessment
3. …voor een ander doel worden gebruikt
4. …uitbesteed worden (dus: wordt door interne organisatie gedaan)

Toelichting

We hebben getracht kaderstellend te zijn en zien graag een voorstel tegemoet, om als raadsbesluit te behandelen

Slide 4

Instructies en stappenoverzicht

Het gaat om:

- Ontwerpen op systeemniveau
- Ontwerpkeuzes inzichtelijk maken
- Advies geven aan raadsleden wat binnen hun kader mogelijk is en welke afwegingen daarbij een rol spelen

Stap 1: ontwerpopdracht opstellen
Stap 2: ontwerpalternatieven produceren
Stap 3: systeemkaarten uitwerken
Stap 4: reflecteren
Stap 5: advies opstellen

Slide 5

Stap 1

Ontwerpopdracht opstellen

Slide 6

Ontwerpopdracht opstellen

1. Neem de probleemformulering en het normatieve kader door
2. Formuleer een ontwerpopdracht waar jullie mee aan de slag kunnen

5 min

Slide 7

Stap 2

Ontwerpalternatieven produceren

Slide 8

Alternatieven

De ontwerpopdracht kan op verschillende manieren (=alternatieven) bewerkstelligd worden

1. Identificeer drie verschillende manieren waarop de opdracht vervult kan worden. Minstens twee van de drie alternatieven moet een algoritmisch systeem zijn.
2. Probeer de alternatieven als een proces te beschrijven. Gebruik 2 tot 3 zinnen.

Algoritmisch systeem: een algoritmisch model dat gebruikt wordt om processen en/of besluitvorming in het openbaar bestuur te ondersteunen, te verbeteren of te automatiseren. Dit algoritmisch model bestaat uit software dat voorspellingen doet, classificaties maakt of uitvoer genereert op basis van vooropgestelde regels of door patronen in grote datasets te ontdekken.

10 min

Slide 9

Stap 3

Systeemkaarten uitwerken

Slide 10

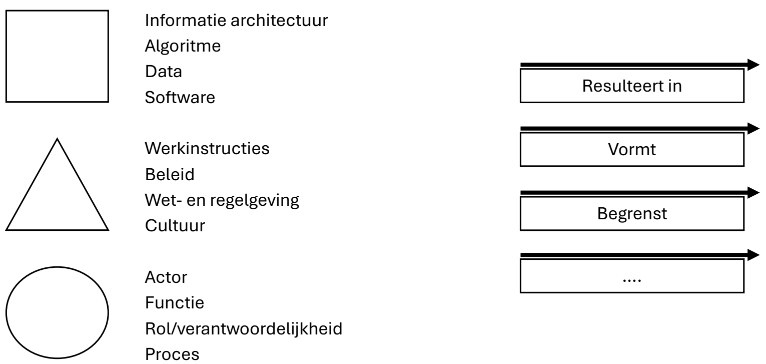
Systeemkaart

Systeemniveauontwerp

- Geeft een overzicht van het volledige publieke algoritmische systeem
- Gaat verder dan de technologie
- Bevat componenten die al jullie disciplines representeren

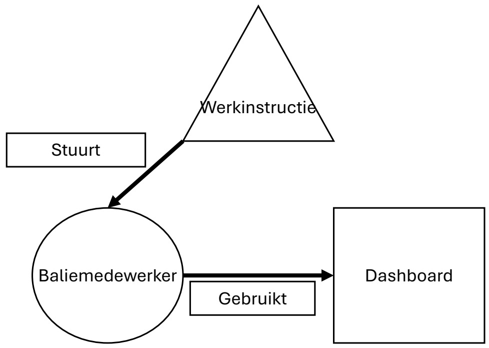
Slide 11

Systeemkaart: 3 componenten



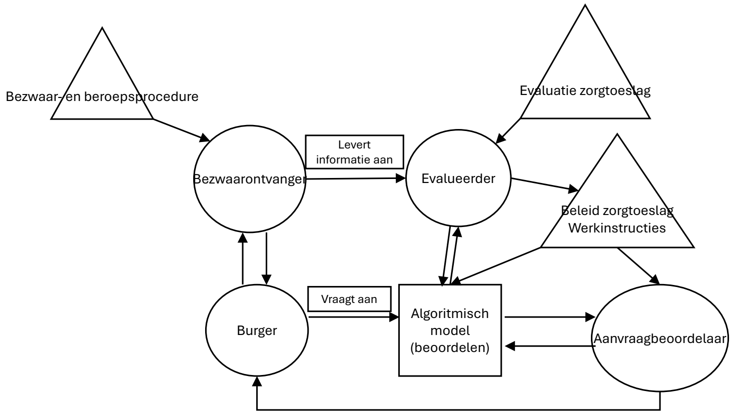
Slide 12

Systeemkaart: voorbeeld 1



Slide 13

Systeemkaart: voorbeeld 2



Slide 14

Systeemkaart: vervaardigen

1. Vervaardig voor elk ontwerpalternatief een systeemkaart
    1. Identificeer de componenten
    2. Verbeeld de interacties tussen de componenten
    3. Maak duidelijk wanneer er meerdere mogelijkheden zijn voor een component

30 min

Slide 15

Stap 4

Reflecteren

Slide 16

Reflecteren op systeemkaart

4 mogelijkheden:

- Besluitvorming op basis van eigen wil of plezier
- Gebrek aan mogelijkheden tot bezwaar en beroep
- Onbegrijpelijke of onvoorspelbare besluitvorming
- Concrete, feitelijke omstandigheden van individuele cases worden niet meegenomen

Zijn deze mogelijkheden te vinden in de kaarten?

Zijn ze problematisch?

20 min

Slide 17

Reflecteren op systeemkaart

Interventies om mogelijkheden te voorkomen, te mitigeren, of te corrigeren

- Voeg componenten toe
- Verander componenten
- Verwijder componenten

20 min

Slide 18

Stap 5

Advies opstellen

Slide 19

Terug naar ontwerpopdracht

1. Ga nog eens terug naar de probleemformulering en het normatieve kader
2. Bepaal welk alternatief het best aansluit bij dit kader
3. Maak eventueel nog wat laatste aanpassingen

5 min

Slide 20

Advies naar raadsleden

1. Beschrijf het werk dat jullie hebben verricht
   1. Geef een toelichting op de systeemkaarten
   2. Beschrijf de meest kritieke ontwerpkeuzes en de achterliggende overwegingen
   3. Beredeneer een voorkeur tussen de alternatieven

20 min

Slide 21

Individuele zienswijzen

1. Ga uit van het advies dat jullie als groep hebben opgesteld
2. Schrijf een individuele zienswijze vanuit je eigen expertise
   1. Welke risico's zie je in de huidige ontwerpen?
   2. Met welke aspecten moeten raadsleden rekening houden?

10 min

Slide 22

Vervolg

- Advies naar raadsleden; zij reflecteren erop
- Daarna kom ik terug bij jullie om hun reflecties te bespreken

Slide 23

B3.3 Session 3

In session 3, the instructor followed the following structure to guide the representative designers through the simulation:

1. Provide a recap of the previous two sessions
2. Present the advice of the system-level designers to the politicians
3. Ask politicians to check whether the system alternatives align with their design goals and design constraints
4. Ask politicians to check whether the goals and constraints should be clarified
5. Ask politicians to adapt the goals and constraints where they think that is needed
6. Ask politicians to formulate questions and/or instructions to the system-level designers

The following presentation slides were used for the instructions:

Quasi-experiment
ontwerpproces van publieke
algoritmische systemen

Sessie 3

Slide 1

Instructies

- Voor de zomer hebben jullie een normatief kader voor een systeem vastgesteld op basis van een probleemformulering
- Een paar weken geleden heeft een groep ambtenaren verschillende syteemalternatieven opgesteld en advies gegeven
- In deze sessie gaan jullie reflecteren op het adviesvan de ambtenaren en jullie eigen normatief kader
- Ik geef instructies maar ook niet te veel
- Blijf binnen de grenzen van wat de gemeente mag en kan doen

Slide 2

Probleemformulering

Probleem

Het Rijk heeft gemeentes gevraagd om de verstrekking van een energietoeslag uit te voeren. Deze energietoeslag is gericht op huishoudens die hun energierekening niet meer kunnen betalen door gestegen prijzen. Het is de taak van de gemeente om te zorgen dat de alle daarvoor in aanmerking komende huishoudens de toeslag ontvangen. De gemeente heeft alleen geen goed beeld van de huishoudens die aanspraak kunnen maken op de toeslag.

Middelen

- Financiële middelen om ervoor te zorgen dat de huishoudens in beeld komen
- Financiële middelen om het toeslagenprogramma uit te voeren
- Data over inkomens, huishoudensgrootte, leeftijden, gebruik van andere toeslagen en/of uitkeringen
- Inzicht in het gemiddelde energieverbruik van huishoudens, energieprijstrends, trends in vraag en aanbod van energie
- Een beleidsdocument van de rijksoverheid waarin staat welke type huishoudens in aanmerking komen en hoeveel toeslag elk type huishouden zal ontvangen

Slide 3

Kader vanuit raadsleden

Ontwerpdoelen

Het te ontwerpen systeem beoogt…

1. ..toepassing van menselijke maat in ontwerpproces en toepassing van het systeem
2. ..het doel van compenseren van huishoudens met een laag inkomen voor een hoge energieprijs
3. ..efficiënt, effectief en meerwaarde hebben voor de organisatie/betrokkenen
4. ..inzichtelijk te zijn voor belanghebbenden, o.a. ten behoeve van bezwaar & beroep, AVG-recht

Ontwerprestricties

Het te ontwerpen systeem kan niet…

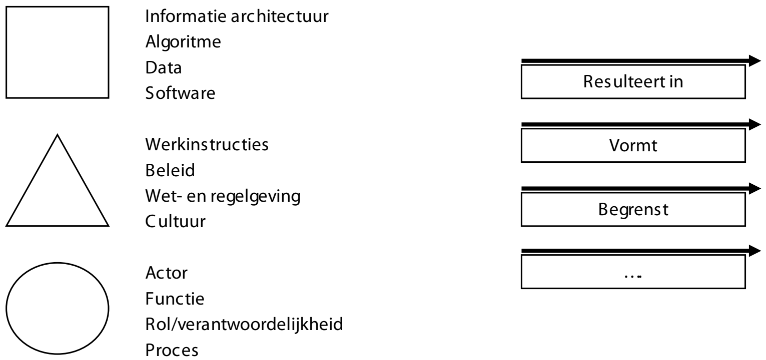
1. ..rechten schenden van de betrokken inwoners, niet limitatief: discriminatie, privacy
2. ..onaanvaardbare/onevenredige risico's met zich mee brengen voor inwoners, bijvoorbeeld maar niet uitsluitend door het uitvoeren van een DPIA, HRIA, algoritmes impact assessment
3. ..voor een ander doel worden gebruikt
4. ..uitbesteed worden (dus: wordt door interne organisatie gedaan)

Toelichting

We hebben getracht kaderstellend te zijn en zien graag een voorstel tegemoet, om als raadsbesluit te behandelen

Slide 4

Uitleg systeemkaarten



Slide 5

Alternatieven

Ontwerpopdracht

Kaders opstellen om de Delftse inwoner op een eerlijke* manier energietoeslag toe te kennen

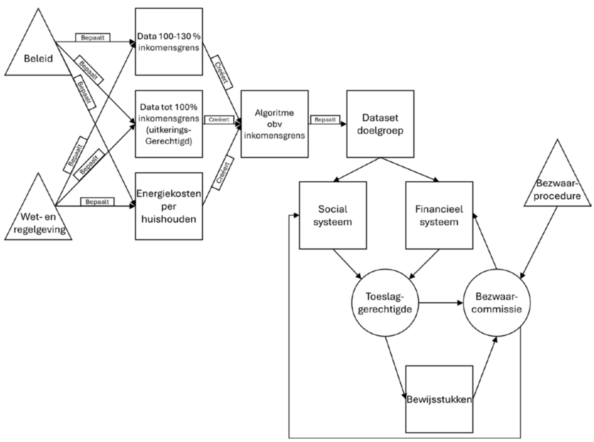
*Laag inkomen; hoge energiekosten

Alternatieven

1. Actief uitkeren op basis van beschikbare informatie
2. Passief uitkeren op aanvraag van de inwoner om te toetsen
3. Deels actief uitkeren en de mogelijkheid open zetten om een aanvraag in te dienen

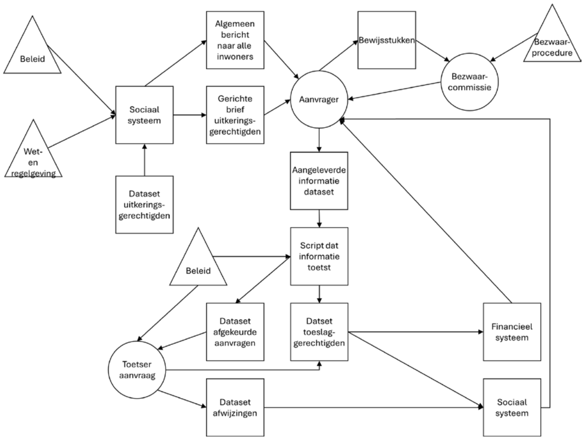
Slide 6

Systeemkaart 1



Slide 7

Systeemkaart 2



Slide 8

Advies

Toelichting systeemkaart

- Laat het algoritmisch systeem zoveel mogelijk JA's eruit filteren → doel efficiency
- Zorg daarnaast voor een aanvraagmogelijkheid + goede bezwaar & beroepsmogelijkheid → doel: menselijke maat & discriminatie voorkomen

Kritieke ontwerpkeuzes

- Omdat de dataset niet 100% betrouwbaar is, moet er altijd een aanvraagmogelijkheid komen
- In het ontwerpproces is geen rekening gehouden met vaste energiecontracten (en dus lage kosten) of met vermogen van inwoners → hierdoor keer je meer uit dan mogelijk wenselijk is

Voorkeursalternatief

Alternatief 3

Slide 9

Het normatief kader

Kennis nemend van het advies van de ambtenaren

- Bespreek:
  - Vallen de systeemalternatieven binnen het normatief kader?

Slide 10

Het normatief kader

Kennis nemend van het advies van de ambtenaren

- Bespreek:
  - Vallen de systeemalternatieven binnen het normatief kader?
  - Kan het normatief kader op verschillende onderdelen verduidelijkt worden?

Slide 11

Het normatief kader

Kennis nemend van het advies van de ambtenaren

- Bespreek:
  - Vallen de systeemalternatieven binnen het normatief kader?
  - Kan het normatief kader op verschillende onderdelen verduidelijkt worden?
  - Wat zouden jullie aanpassen aan het normatief kader?

Slide 12

Het normatief kader

Kennis nemend van het advies van de ambtenaren

- Bespreek:
  - Vallen de systeemalternatieven binnen het normatief kader?
  - Kan het normatief kader op verschillende onderdelen verduidelijkt worden?
  - Wat zouden jullie aanpassen aan het normatief kader?
- Beschrijf de aanpassingen

Slide 13

Instructies/vragen aan ambtenaren?

Kennis nemend van het advies van de ambtenaren

- Bespreek en beschrijf kort:
  - Welke vragen hebben jullie aan ambtenaren?
  - Welke instructies zouden jullie aan de ambtenaren mee willen geven?

Slide 14

Indrukken van simulatie?

Welke inzichten heeft de simulatie jullie opgeleverd?

Slide 15

Vervolg

Ik stuur jullie komende dinsdag een laatste vragenlijst

Slide 16

B3.4 Session 4

In session 4, the instructor followed the following structure to guide the representative designers through the simulation:

1. Provide a recap of the previous three sessions
2. Present the adapted design goals and design constraints, the questions, and the instruction of the politicians to the system-level designers
3. Ask designers to reflect whether the politicians' reaction provides clarification of the design goals and constraints
4. Ask designers to check whether their advice should be clarified
5. Ask designers to adapt their advices where they think that is needed
6. Ask designer to formulate questions and/or instructions to the politicians

The following presentation slides were used for the instructions:

Quasi-experiment
ontwerpproces van publieke
algoritmische systemen

Sessie 4

Slide 1

Instructies

- Voor de zomer hebben de raadsleden een normatief kader voor een systeem vastgesteld op basis van een probleemformulering
- Een paar weken geleden hebben jullie verschillende syteemalternatieven opgesteld en advies gegeven
- Eind vorige week hebben de raadsleden gereageerd op jullie ontwerp
- In deze sessie gaan jullie reflecteren op de reactie van de raadsleden en jullie eigen advies
- Ik geef instructies maar ook niet te veel
- Blijf binnen de grenzen van wat de gemeente mag en kan doen

Slide 2

Probleemformulering

Probleem

Het Rijk heeft gemeentes gevraagd om de verstrekking van een energietoeslag uit te voeren. Deze energietoeslag is gericht op huishoudens die hun energierekening niet meer kunnen betalen door gestegen prijzen. Het is de taak van de gemeente om te zorgen dat de alle daarvoor in aanmerking komende huishoudens de toeslag ontvangen. De gemeente heeft alleen geen goed beeld van de huishoudens die aanspraak kunnen maken op de toeslag.

Middelen

- Financiële middelen om ervoor te zorgen dat de huishoudens in beeld komen
- Financiële middelen om het toeslagenprogramma uit te voeren
- Data over inkomens, huishoudensgrootte, leeftijden, gebruik van andere toeslagen en/of uitkeringen
- Inzicht in het gemiddelde energieverbruik van huishoudens, energieprijstrends, trends in vraag en aanbod van energie
- Een beleidsdocument van de rijksoverheid waarin staat welke type huishoudens in aanmerking komen en hoeveel toeslag elk type huishouden zal ontvangen

Slide 3

Kader vanuit raadsleden

Ontwerpdoelen

Het te ontwerpen systeem beoogt…

1. …toepassing van menselijke maat in ontwerpproces en toepassing van het systeem
2. …het doel van compenseren van huishoudens met een laag inkomen voor een hoge energieprijs
3. …efficiënt, effectief en meerwaarde hebben voor de organisatie/betrokkenen
4. …inzichtelijk te zijn voor belanghebbenden, o.a. ten behoeve van bezwaar & beroep, AVG-recht

Ontwerprestricties

Het te ontwerpen systeem kan niet…

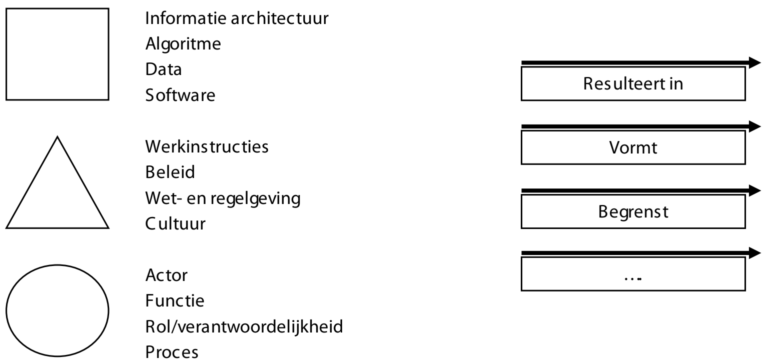
1. …rechten schenden van de betrokken inwoners, niet limitatief: discriminatie, privacy
2. …onaanvaardbare/onevenredige risico's met zich mee brengen voor inwoners, bijvoorbeeld maar niet uitsluitend door het uitvoeren van een DPIA, HRIA, algoritmes impact assessment
3. …voor een ander doel worden gebruikt
4. …uitbesteed worden (dus: wordt door interne organisatie gedaan)

Toelichting

We hebben getracht kaderstellend te zijn en zien graag een voorstel tegemoet, om als raadsbesluit te behandelen

Slide 4

Uitleg systeemkaarten



Slide 5

Alternatieven

Ontwerpopdracht

Kaders opstellen om de Delftse inwoner op een eerlijke* manier energietoeslag toe te kennen

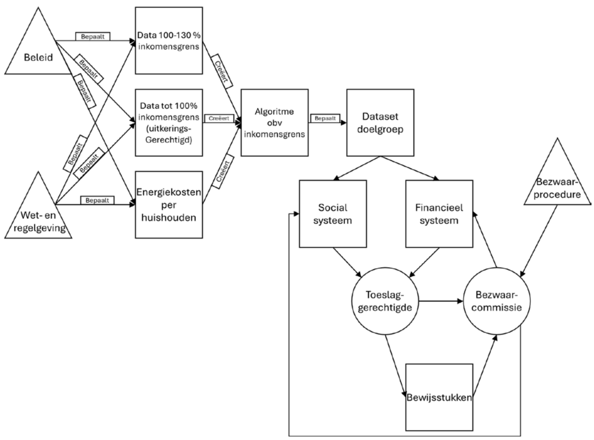
*Laag inkomen; hoge energiekosten

Alternatieven

1. Actief uitkeren op basis van beschikbare informatie
2. Passief uitkeren op aanvraag van de inwoner om te toetsen
3. Deels actief uitkeren en de mogelijkheid open zetten om een aanvraag in te dienen

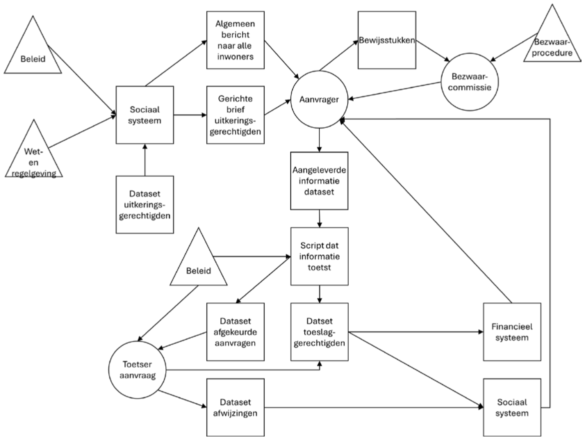
Slide 6

Systeemkaart 1



Slide 7

Systeemkaart 2



Slide 8

Advies

Toelichting systeemkaart

- Laat het algoritmisch systeem zoveel mogelijk JA's eruit filteren → doel efficiency
- Zorg daarnaast voor een aanvraagmogelijkheid + goede bezwaar & beroepsmogelijkheid → doel: menselijke maat & discriminatie voorkomen

Kritieke ontwerpkeuzes

- Omdat de dataset niet 100% betrouwbaar is, moet er altijd een aanvraagmogelijkheid komen
- In het ontwerpproces is geen rekening gehouden met vaste energiecontracten (en dus lage kosten) of met vermogen van inwoners → hierdoor keer je meer uit dan mogelijk wenselijk is

Voorkeursalternatief

Alternatief 3

Slide 9

Reactie raadsleden

- Toelichting per ontwerpdoel/-restrictie; waarom voldoet het aan elk individueel punt?
  - Graag ook een toelichting op non-discriminatie + privacy
- Toelichting op het model?
- Hoe ziet alternatief 3 eruit?
- Model 2: tav "algemeen bericht naar alle bewoners"; hoe ga je dat doen?
- Hoe nauwkeurig is alternatief 1?
- Kan er bij alternatief 1 nog een algemeen bericht uit?
- Kan er nog een feedback loop in alternatief 1?
- Krijgen wij als Raad nog een terugkoppeling over het systeem als het in werking is?
  - Bezwaar, efficiëntie/meerwaarde, etc.

Slide 10

Het normatief kader

Kennis nemend van de reactie van de raadsleden

- Bespreek:
  - In hoeverre geeft de reactie verduidelijking van het normatief kader?

Slide 11

Advies

Kennis nemend van de reactie van de raadsleden

- Bespreek:
  - Welke onderdelen van jullie advies zouden verduidelijkt kunnen worden?

Slide 12

Advies

Kennis nemend van de reactie van de raadsleden

- Bespreek:
  - Welke onderdelen van jullie advies zouden verduidelijkt kunnen worden?
  - Welke onderdelen van jullie advies zouden aangepast kunnen worden?

Slide 13

Advies

Kennis nemend van de reactie van de raadsleden

- Bespreek:
  - Welke onderdelen van jullie advies zouden verduidelijkt kunnen worden?
  - Welke onderdelen van jullie advies zouden aangepast kunnen worden?
- Beschrijf de aanpassingen

Slide 14

Instructies/vragen aan raadsleden?

Kennis nemend van de reactie van de raadsleden

- Bespreek en beschrijf kort:
  - Welke vragen hebben jullie aan de raadsleden?
  - Welke instructies zouden jullie aan de raadsleden mee willen geven?

Slide 15

Indrukken van simulatie?

Welke inzichten heeft de simulatie jullie opgeleverd?

Slide 16

Vervolg

Ik stuur jullie komende maandag een laatste vragenlijst

Slide 17

B4 Simulation: design outputs

B4.1 Session 1

Political instruction

Ontwerpdoelen

Het te ontwerpen systeem beoogt…

1. …toepassing van menselijke maat in ontwerpproces en toepassing van het systeem
2. …het doel van compenseren van huishoudens met een laag inkomen voor een hoge energieprijs
3. …efficiënt, effectief en meerwaarde hebben voor de organisatie/betrokkenen
4. …inzichtelijk te zijn voor belanghebbenden, o.a. ten behoeve van bezwaar & beroep, AVG-recht

Ontwerprestricties

Het te ontwerpen systeem kan niet…

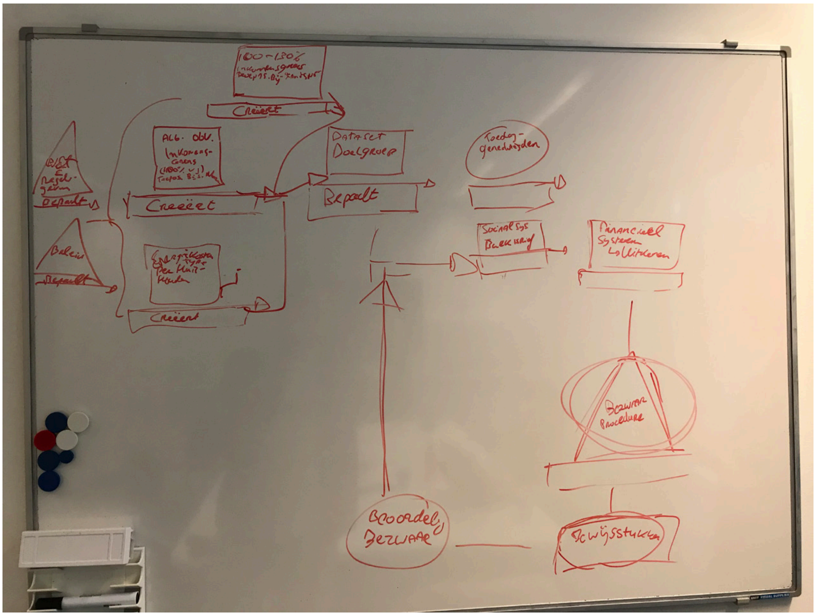
1. …rechten schenden van de betrokken inwoners, niet limitatief: discriminatie, privacy
2. …onaanvaardbare/onevenredige risico's met zich mee brengen voor inwoners, bijvoorbeeld maar niet uitsluitend door het uitvoeren van een DPIA, HRIA, algoritmes impact assessment
3. …voor een ander doel worden gebruikt
4. …uitbesteed worden (dus: wordt door interne organisatie gedaan)

Toelichting

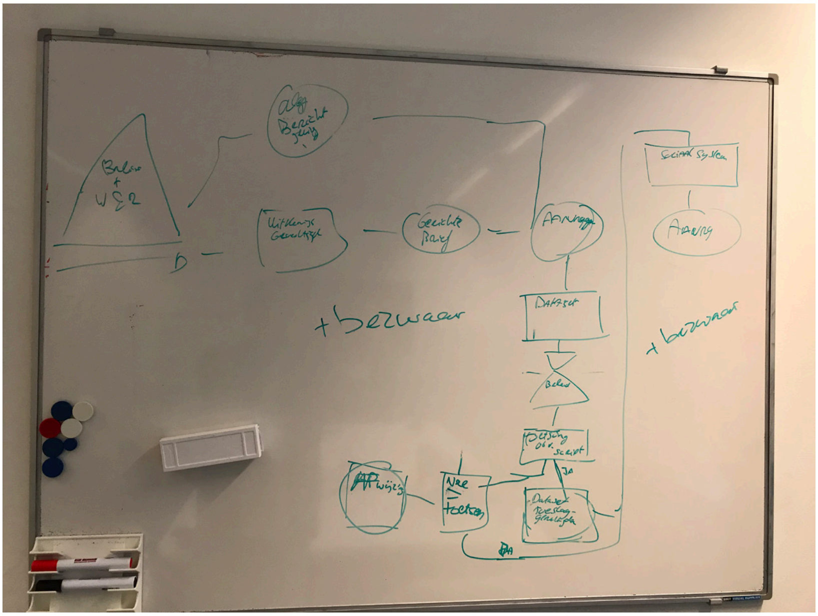
We hebben getracht kaderstellend te zijn en zien graag een voorstel tegemoet, om als raadsbesluit te behandelen

B4.2 Session 2

System map alternative 1



System map alternative 2



Advies

Toelichting systeemkaart

(*) Laat het algoritmisch systeem zoveel mogelijk JA's eruit filteren → Doel efficiency
(*) Zorg daarnaast voor een aanvraagmogelijkheid + goede bezwaar- & beroepsmogelijkheid
↳ Doel: Menselijke maat & discriminatie voorkomen

Kritieke ontwerpkeuzes

(*) Omdat de dataset niet 100% betrouwbaar is, moet er altijd een aanvraagmogelijkheid komen.
(*) In het ontwerpproces is geen rekening gehouden met vaste energiecontracten (en dus lage kosten) of niet vermogen van inwoners.
↳ ~~Dat zijn~~ Hierdoor keer je meer uit dan mogelijk wenselijk

Voorkeursalternatief

Alternatief 3.

B4.3 Session 3

Reaction representative designers to system-level designers

- Toelichting per ontwerpdoel/-restrictie; waarom voldoet het aan elk individueel punt? Graag ook een toelichting op non-discriminatie + privacy
- Toelichting op het model?
- Hoe ziet alternatief 3 eruit?
- Model 2: tav "algemeen bericht naar alle bewoners"; hoe ga je dat doen?
- Hoe nauwkeurig is alternatief 1?
- Kan er bij alternatief 1 nog een algemeen bericht uit?
- Kan er nog een feedback loop in alternatief 1?
- Krijgen wij als Raad nog een terugkoppeling over het systeem als het in werking is? Bezwaar, efficiëntie/meerwaarde, etc.

B5 Questionnaires after simulation

B5.1 Questionnaire after simulation representative designers

Vragenlijst na het quasi-experiment "ontwerpprocessen van publieke algoritmische systemen"

Beste deelnemer,

U heeft nu aan twee sessies van de simulatie deelgenomen. Voorafgaand aan de simulatie heeft u al een vragenlijst ingevuld. Om de simulatie te kunnen evalueren wil ik u nogmaals een vragenlijst in te vullen.

De vragenlijst bestaat uit 12 open vragen. Geef bij elk antwoord een korte toelichting. Invullen neemt ongeveer 20 minuten in beslag.

Dit is het laatste onderdeel van mijn onderzoek. Heel erg bedankt voor uw deelname!

Een deel van de vragen bevatten termen die niet vanzelf spreken of door verschillende organisaties anders worden gedefinieerd. In deze vragenlijst worden de volgende definities gebruikt:

In een ***publiek algoritmisch systeem*** wordt een algoritmisch model gebruikt om processen en/of besluitvorming in het openbaar bestuur te ondersteunen, te verbeteren of te automatiseren. Dit algoritmisch model bestaat uit software dat voorspellingen doet, classificaties maakt of uitvoer genereert op basis van vooropgestelde regels of door patronen in grote datasets te ontdekken. Het algoritmisch model heeft op zichzelf geen functie. Het model krijgt pas een functie als de uitvoer wordt gebruikt door een ambtenaar en wanneer het model onderdeel is van beleid en/of wet- en regelgeving. Dit betekent dat bijvoorbeeld ook de werkinstructies voor de gebruikers van het algoritmisch model onderdeel vormen van het systeem.

Met ***ontwerpen*** bedoelen we in deze vragenlijst alle activiteiten die worden ondernomen om tot een gewenst algoritmisch systeem te komen. Het gaat hier dus niet alleen om de ontwikkeling van de software. Ook het stellen van kaders, het opstellen van werkinstructies en het adviseren over juridische aspecten van een algoritmisch model zijn voorbeelden van activiteiten die onder ontwerpen vallen.

Wanneer er gevraagd wordt naar iets wat u niet meemaakt of herkent, wil ik u vragen dat aan te geven bij het antwoord.

Vragen huidige situatie

Sturing op het ontwerpen van publieke algoritmische systemen

1. Wat zou het invoeren van het normatief kader (de ontwerpdoelen en ontwerprestricties) volgens u veranderen aan het geven van sturing aan het ontwerpen van publieke

algoritmische systemen binnen de gemeente?

Antwoord:

2. Hoe zou u uw invloed op het ontwerp binnen de simulatie karakteriseren? Via welke weg had u die invloed?

Antwoord:

Onderling begrip

3. In hoeverre maakt het advies dat u vanuit de ambtenaren binnen het experiment aangeleverd kreeg inzichtelijk hoe het systeem zou werken?

Antwoord:

4. Welke informatie had u graag gehad in de simulatie? Of welke informatie had verduidelijkt moeten worden?

Antwoord:

Motivering achter ontwerpkeuzes

5. In welke mate gaf het advies u inzicht in de argumentatie achter keuzes die ambtenaren maken in het ontwerpen van publieke algoritmische systemen?

Antwoord:

Zorgvuldigheid bij het komen tot ontwerpkeuzes

6. In welke mate gaf het advies inzicht in de gevolgen voor burgers die publieke algoritmische systemen (kunnen) hebben?

Antwoord:

In staat zijn om keuzes te maken

7. In hoeverre stellen de informatie en de dialoog binnen de simulatie u in staat om uw besluitvormende en controlerende taken als raadslid uit te voeren als het gaat over publieke algoritmische systemen?

Antwoord:

Reflectie

8. In hoeverre gaf het advies van de ambtenaren aanleiding tot het veranderen en/of verduidelijken van uw eigen ideeën over het normatief kader?

Antwoord:

Vragen over gewenste situatie

9. Wat zijn uw algemene bevindingen over de simulatie?

Antwoord

10. Wat zou volgens u in het ontwerpproces in de simulatie moeten veranderen aan de rol van de raad in het ontwerpen van publieke algoritmische systemen?

Antwoord:

11. Wat zou volgens u binnen het proces in de simulatie moeten veranderen aan de rol van de ambtenarij in het ontwerpen van publieke algoritmische systemen?

Antwoord:

12. Heeft de simulatie u beïnvloedt in uw werk als raadslid? Zo ja, hoe?

Antwoord:

Dit is het einde van de vragenlijst

B5.2 Questionnaire after simulation system-level designers

Vragenlijst na het quasi-experiment "ontwerpprocessen van publieke algoritmische systemen"

Beste deelnemer,

U heeft nu aan twee sessies van de simulatie deelgenomen. Voorafgaand aan de simulatie heeft u al een vragenlijst ingevuld. Om de simulatie te kunnen evalueren wil ik u nogmaals vragen een vragenlijst in te vullen.

De vragenlijst bestaat uit 12 open vragen. Geef bij elk antwoord een korte toelichting. Invullen neemt ongeveer 20 minuten in beslag.

Dit is het laatste onderdeel van mijn onderzoek. Heel erg bedankt voor uw deelname!

Een deel van de vragen bevatten termen die niet vanzelf spreken of door verschillende organisaties anders worden gedefinieerd. In deze vragenlijst worden de volgende definities gebruikt:

In een **publiek algoritmisch systeem** wordt een algoritmisch model gebruikt om processen en/of besluitvorming in het openbaar bestuur te ondersteunen, te verbeteren of te automatiseren. Dit algoritmisch model bestaat uit software dat voorspellingen doet, classificaties maakt of uitvoer genereert op basis van vooropgestelde regels of door patronen in grote datasets te ontdekken. Het algoritmisch model heeft op zichzelf geen functie. Het model krijgt pas een functie als de uitvoer wordt gebruikt door een ambtenaar en wanneer het model onderdeel is van beleid en/of wet- en regelgeving. Dit betekent dat bijvoorbeeld ook de werkinstructies voor de gebruikers van het algoritmisch model onderdeel vormen van het systeem.

Met **ontwerpen** bedoelen we in deze vragenlijst alle activiteiten die worden ondernomen om tot een gewenst algoritmisch systeem te komen. Het gaat hier dus niet alleen om de ontwikkeling van de software. Ook het stellen van kaders, het opstellen van werkinstructies en het adviseren over juridische aspecten van een algoritmisch model zijn voorbeelden van activiteiten die onder ontwerpen vallen.

Wanneer er gevraagd wordt naar iets wat u niet meemaakt of herkent, wil ik u vragen dat aan te geven bij het antwoord.

Vragen huidige situatie

Sturing op het ontwerpen van publieke algoritmische systemen

1. Op welke manier hebben de ontwerpdoelen en ontwerprestricties van de raadsleden de keuzes die u heeft gemaakt in de simulatie gestuurd?

Antwoord:

2. Hoe zou u uw invloed op het ontwerp binnen de simulatie karakteriseren? Via welke weg had u die invloed?

Antwoord:

Onderling begrip

3. Wat zou het invoeren van het advies zoals jullie dat in de simulatie hebben opgesteld veranderen aan het politieke debat in de gemeenteraad?

Antwoord:

4. Welke informatie had u, naast het advies, nog meer mee willen geven aan de raadsleden?

Antwoord:

Motivering achter ontwerpkeuzes

5. In welke mate heeft de werkwijze in de simulatie u geholpen bij het vormen van de argumentatie achter de gemaakte keuzes tijdens het ontwerpen van publieke algoritmische systemen?

Antwoord:

Zorgvuldigheid bij het komen tot ontwerpkeuzes

6. In welke mate heeft de werkwijze in de simulatie u geholpen bij het in kaart brengen van (mogelijke) gevolgen van publieke algoritmische systemen voor burgers?

Antwoord:

In staat zijn om keuzes te maken

7. In hoeverre heeft de informatie en de dialoog binnen de simulatie u geholpen om politieke wensen en grenzen in de raad te vertalen naar publieke algoritmische systemen?

Antwoord:

Reflectie

8. In hoeverre gaf de reactie van de raadsleden aanleiding tot het veranderen en/of verduidelijken van uw eigen ideeën over het publieke algoritmische systeem dat jullie hebben ontworpen?

Antwoord:

Vragen over gewenste situatie

9. Wat zijn uw algemene bevindingen over de simulatie?

Antwoord

10. Wat zou volgens u in het ontwerpproces in de simulatie moeten veranderen aan de rol van de ambtenarij in het ontwerpen van publieke algoritmische systemen?

Antwoord:

11. Wat zou volgens u binnen het proces in de simulatie moeten veranderen aan de rol van de raad in het ontwerpen van publieke algoritmische systemen?

Antwoord:

12. Heeft de simulatie u beïnvloedt in uw werk als ambtenaar? Zo ja, hoe?

Antwoord:

Dit is het einde van de vragenlijst

About Curbing Algorithmic Kafka

Public organisations increasingly rely on algorithmic systems for the execution of their tasks and the provision of public services. These public algorithmic systems can inflict harms on citizens by creating Kafkaesque situations. Exemplary cases of Kafkaesque algorithmic systems are the Dutch childcare allowances scandal and the Australian Robodebt scheme. Algorithmic Kafka emerges from possibilities for arbitrary conduct in the constitution of algorithmic systems. Following from the fundamental principles underlying democracy and the Rule of Law, governments should protect citizens from any arbitrary use of power by public organisations. Nonetheless, current design practices of public algorithmic systems do not provide this protection and are lacking democratic legitimacy.

This thesis provides design principles for institutional interventions that restructure design practices of public algorithmic. More specifically, these design principles prescribe how current technocratic and businesslike design practices can be shifted to practices that are embedded in a democratic and Rule of Law context. The institutional interventions facilitate public servants in coordinating the formulation of a socio-technical specification for public algorithmic systems. Furthermore, the interventions reduce arbitrary use of designerly power by strengthening the dialectic between public servants and politicians. The design practices that emerge by establishing the position of system-level designer will curb the emergence of algorithmic Kafka.