Science & Technology Studies

3/2024

Science & Technology Studies

ISSN 2243-4690

Science & Technology Studies

Volume 37, Issue 3, 2024

Articles

Discussion paper

Book reviews

Visit our web-site at

www.sciencetechnologystudies.org

Gender Segregation in the Borderlands of E-Science

Öznur Karakaş

Uppsala University, Sweden; and Norwegian Science and Technology University, Norway/ oznur.karakas@ntnu.no

Gabriele Griffin

Uppsala University, Sweden

Abstract

This article draws on an ethnographic study of an e-science platform in Sweden to analyse how horizontal gender segregation across sciences plays out in e-science, a borderland in which sciences converge around state-of-the art computational technologies for scientific research. While the convergence of sciences in e-science has the potential to open a non-traditional trajectory to attract women to ICTs, we find that this potential remains untapped. Instead horizontal gender segregation is perpetuated through a) restricted mobility of women from scientific fields with higher gender parity to IT, b) gender friction negatively affecting women in cross-disciplinary e-science, c) a gendered developer/user divide permeating e-science collaborations under 'the logic of domains,' and d) perceived self-reliance in computational tool development across sciences acting as 'gendered boundary work' to strengthen the gendered hard/soft divide in sciences.

Keywords: Feminist Science and Technology Studies (FTS), Interdisciplinary Collaboration, E-Infrastructure, Interdisciplinarity, Gendered Asymmetries

Introduction

It is well established in feminist technology studies (FTS) that technology and gender are mutually constructed (Berg, 1994; Berg and Lie, 1995; Faulkner, 2001; Oudshoorn et al., 2002; Montgomery, 2012; Wacjman, 2000, 2004, 2007, 2010). FTS emphasise situated analyses of the ways in which new technologies change how gender is done and how prevailing gender dynamics in turn affect the closure and standardization processes of certain technologies and technological artefacts (Wacjman, 2000). What is at stake is the question of how gender and technology are entangled in context-specific ways. Computerization (Hine, 2006) and digitalization have periodically been sites for technology-related hypes (Ensmenger, 2010). The potential of virtual space and online technologies to produce social changes has been heralded (Plant, 1995; Castells, 2012). Yet, with reference to grid technologies that lay the framework for e-science endeavours, Woolgar and

Coopmans (2006: 3) make clear that "the nature and direction of change is unpredictable". They argue that there is a lack of research regarding the social dynamics of e-science (Woolgar and Coopmans, 2006: 4). This is all the more the case when it comes to research on gender in e-science at a time when earlier e-infrastructure investments have started to give way to data-driven science initiatives, such as the recently founded SciLifelab and the Wallenberg National Program for Data-Driven Life Sciences (DDSL) in Sweden.

E-science is a term used to describe science that involves developing or working with computational methods, tools and applications, very large data sets and distributed network or grid systems to deal with these. Following the publication of the *Revolutionizing Science and Engineering through Cyberinfrastructure Report* in 2003 in the USA, studies on e-science in the STS literature have largely focused on the prospect of distributed collaboration and coordination in e-infrastructure. In hindsight, David Ribes (2019: 520) gives an account of how the role of social science in e-science initiatives was envisaged to be "helping with, for instance, working in a geographically distributed manner, bringing together heterogeneous disciplinary scientists, or examining the difficulties of data sharing". Yet, in these endeavours gender remained either largely understudied, or not dealt with at all. New tools and applications of the rising, male-dominated data sciences *(*Boston Consulting Group Gamma Study, 2019), such as AI and machine learning (ML), have recently been assembled in e-science platforms without due attention to how gender figures within these. STS scholars who study e-science have started to shift their attention to data science (Ribes, 2019; Paine and Lee, 2020; Beaulieu and Leonelli, 2021; Mökander and Schroeder, 2021). With this shift of attention gendered asymmetries which are sunk into the infrastructure might become further invisibilised. Bowker (1994) underlines the need for 'infrastructural inversion'; that is "shifting the emphasis from changes in infrastructural components to changes in infrastructural relations" (Bowker et al., 2010: 99), not least with respect to gender dynamics. This also speaks to the emphasis in the literature on the need to study the 'human infrastructure' of e-science (Lee 2006; Bietz et al., 2010).

This article analyses how existing gender divisions across the sciences, also known as horizontal gender segregation (Corneliussen, 2021), play out in cross-disciplinary e-science with an analysis of the organisational structure of an e-science platform in Sweden (hereafter called the Platform). The under-representation of women in STEM fields which converge in e-science collaborations is well established (Rua-Gomez and Arias-Gaviria, 2020; Zacharia et al., 2020; Santos et al., 2021). Horizontal gender segregation attests to the fact that gender parity also varies across STEM disciplines (Ceci et al., 2014, Begeny et al., 2020; Fisher et al., 2020). The *Gender InSITE Report* records, for instance, higher numbers of women in biological or life sciences (28%) compared to engineering sciences (10%) and mathematical sciences (8%) and a dire under-representation of women in computer sciences (Kelan, 2007; Michell et al., 2017; Zacharia et al., 2020). How do these gendered differences (Ellingsæter, 2014; in Corneliussen, 2021) play out in e-science collaborations where diverse scientific 'disciplines' converge around the use and development of computational tools and methods? In what ways do traditional gendered divisions across math-intensive and non-math intensive sciences (Ceci et al., 2014) play out in cross-disciplinary e-science collaborations? The article responds to these research questions, focusing on the case of a particular e-science platform established in 2010 as a strategic research area by three major universities in Sweden.

Vitores and Gil-Juarez (2016: 670) emphasise the fact that women's engagement in ICTs might at times take certain pathways which are not well-recorded in the literature as they fall outside the scope of a linear computer science career. They invite us to consider other trajectories that emerge or exist at the intersection of computer sciences and other disciplines such as "art and design, cognitive sciences, new media, biology, information science and education or library science, for example" (Vitores and Gil-Juarez, 2016: 673). E-science, being one such area in which diverse sciences converge around ICTs, is a relevant space as a potential pathway to attract women into ICTs.

It is thus useful to analyse existing e-science initiatives to see how horizontal gender segregation plays out in newly emerging research clusters which "conjoin technology, traditionally male dominated, and disciplines that have traditionally been female dominated" (Griffin, 2021: 1).

The article suggests that e-science collaborations act as borderlands where different communities of practice intersect in individuals and groups. These borderlands thus embody the potential to become an area where researchers of more gender equal communities of practice such as biology, medicine and humanities, enter IT-heavy e-science through application-based collaborations. This potential is discussed below in the findings section. In the same section, we also discuss that this potential is not realized in our case study. Instead, existing horizontal gender segregation is perpetuated. We discuss the mechanisms of this below based on data from our analysis of the organisation structure and gender dynamics of the Platform in Sweden. The next section of the article includes our literature review on horizontal gender segregation across the sciences, and a discussion on e-science as borderland. After that we present the case study and the methodology used for the qualitative research. We then present our findings, followed by a discussion and conclusion section.

Horizontal gender segregation across the sciences

When analysing the existing gender dynamics in sciences converging in e-science platforms, an acute and persistent under-representation of women in STEM fields has been diagnosed (Dasgupta and Stout, 2014; Su and Rounds, 2015; Alegria et al., 2016; Sax et al., 2017; Wang and Degol, 2017; Moss-Racusin et al., 2018; Van Veelen et al., 2019; Rua-Gomez and Arias-Gaviria, 2020; Zacharia et al., 2020; Santos et al., 2021). Figures on gender parity in STEM fields attest to this, with different dynamics in math-intensive and non-math intensive fields (Ceci et al., 2014). The International Science Council's *Gender InSITE Report* (2021: 10) analyses gender parity in 85 individual STEM academies from across the world[1] and records the average percentage of women's representation in STEM to be 17% in 2020. This average disguises the fact that gender parity in STEM also varies considerably according to discipline (Barone, 2011; Ceci et al., 2014; Su and Rounds, 2015; Cheryan et al., 2017; Sax et al., 2017; Begeny et al., 2020; Fisher et al., 2020). The *GenderInSITE Report* (2021: xi) states that gender equality varies across disciplines, and it requires a discipline-based action plan. In their comprehensive study on women in academic science in the USA, Ceci et al. (2014) found different patterns within STEM fields both in female representation and in later career attrition rates. Accordingly, they distinguish math-intensive disciplines including geoscience, engineering, economics, mathematics/computer science and the physical sciences - chemistry and physics - (GEEMP) from non-math-intensive disciplines. The latter include life sciences, psychology and social sciences (LPS) (Ceci et al., 2014: 76). Women are clearly under-represented in GEEMP fields. The LPS fields record no gender gap or even over-representation of women at the undergraduate level, while suffering from higher attrition rates at postgraduate level and in moving to associate professorships. In contrast, women who enter GEEMP fields suffer less attrition rates at postgraduate level and in moving to associate professorships when compared to women in LPS (Wang and Degol, 2017: 80).

Gendered computer sciences

The persistent under-representation of women in computer sciences (Lagesen, 2007; Gillard et al., 2007, 2010; Ceci et al., 2014; Vitores and Gil-Juarez, 2016; Cheryan et al., 2017; Sax et al., 2017; Michell et al., 2017, Zacharia et al., 2020) poses a challenge for gender mainstreaming in e-science platforms, given the significant role that computing, computationalisation and ICTs play in these platforms. The *European Parliament Report on Education and Employment of Women in Science, Technology and the Digital Economy* (Zacharia et al., 2020) highlights the under-representation of women in computer sciences in the EU, particularly in artificial intelligence and cybersecurity. The report argues that "even undergraduate female students in computer sciences believe that computer science is a male domain" (Zacharia et al., 2020: 25). It also states that "The percentage of women in

ICT careers still remains relatively low, and it is currently below 2% of women's total share in the European labour market" (Zacharia et al., 2020: 14) and that "the gender gap concerning AI and cybersecurity is the largest among all digital technology domains. The average percentages of females in AI and cybersecurity worldwide are 12% and 20% respectively" (Zacharia et al., 2020: 9). Although the under-representation of women in computer science is not universal and some countries such as Malaysia and India are at odds with this male-dominated picture of the field (Lagesen, 2008; Mellström, 2009; Vitores and Gil-Juarez, 2016: 672), the problem is dire in the more affluent Western world, especially in the Scandinavian countries which also boast higher levels of gender equality. The *Telenor Report on the Gender Gap in Technology in Scandinavia* (2019) states that according to the 2018 OECD Gender Data Portal "only 1 in 5 computer science graduates are women" across 35 European countries. "[T]he gender gap in Norway, Sweden and Denmark is particularly wide" (Telenor, 2019: 9), it adds. Scholars refer to this as the gender equality paradox (Stoet and Greary, 2018; Corneliussen, 2021), or living the contradiction (Griffin, 2022).

One frequent interpretation of the under-representation of women in the field conceives it as a "supply problem" (Vitores and Gil-Juarez, 2016: 670-671) or 'untapped human capital' (Dasgupta and Stout, 2014), generally focusing on wasted human resources. Against this assumption that conceives under-representation of women as a supply problem in the market, we suggest that the problem is rather related to how society is shaped around gendered technological fields of expertise and gendered technologies, feeding into and perpetuating this under-representation (Vitores and Gil-Juarez, 2016). One way in which the problematic of women's under-representation in the sciences might be mitigated is through collaboration across differently gendered sciences, i.e., through collaboration between female-dominated and male-dominated sciences. Such collaboration can potentially occur within e-science platforms that, as explained below, constitute a kind of borderland between diverse sciences.

e-Science as borderland

Following Gloria Anzaldua, Susan Leigh Star (2015: 157) conceptualises the notion of the borderland as the space opened up "when two communities of practice coexist in one person". The concept of borderland enables us to conceive of moving between not only disciplinary boundaries but also gender boundaries across different disciplines. Gendered subjectification acts as an inclusion/exclusion mechanism when it comes to entering certain practices, and climbing the career ladder; it thus contributes to the gendering of sciences as communities of practice. A recent study indicates that "increasing the perceived presence of women in a STEM discipline increases the likelihood that participants would label it a soft science", and "labelling disciplines as soft sciences leads to the fields being devalued, deemed less rigorous, and less worthy of federal funding" (Light et al., 2022: 1). Male-dominated math-intensive disciplines are labelled as hard. One result of this gendered conception of sciences is the so-called math self-efficacy gap, the fact that women exhibit a lower perception of their math competence compared to men. This also plays a role in the under-representation of women in computer science (Cheryan et al., 2017, Fisher et al. 2020; Stearns et al., 2020).[2] Although some argue that there are no longer any math performance gaps between girls and boys (Stearns et al., 2020), math self-efficacy is still low in women. The gendered hard/soft divide across math-intensive and non-math-intensive fields (Ceci et al., 2014) and math self-efficacy feed into each other. Gender stereotypes depicting computer scientists as male geeks or hacker figures serve as another example of how this scientific field is gendered, and how that gendering estranges women from the field (Lagesen, 2007; Reuben et al., 2014; Michell et al., 2017; Sax et al., 2017).

Science and technology studies have contributed a lot to the study of disparities across the sciences, under the rubric of the 'disunity of science' (Galison and Stump, 1996), disciplinary culture (Traweek, 1988), and 'epistemic cultures' (Knorr-Cetina, 1999). But we have yet to understand how the material-semiotic enactments that make up the sciences are laden with distinct gendering mechanisms. This calls for an analysis of horizontal gender segregation across the sciences in terms of epistemic cultures. This article hence integrates (feminist) science and technology studies and studies on women in science, which have largely remained separate (Bauschspies and Puig de la Bellacasa, 2009).

Case study and methodology

The e-science platform discussed here was established by three major Swedish universities in 2010 as a response to the Swedish Government Bill on Research Policy promoting e-science as a strategic research area (SRA). E-science is defined on the official website of the platform as including both the use and development of new computational methods and tools. The potential to collaborate across disciplines in academia and with industry regardless of geographical distance is also emphasised on the website.

Two features, namely distributed collaboration on a shared virtual network and cross-disciplinarity, are presented as the main features of e-science in the literature. STS literature on e-science mostly defines e-science as making use of new information and communication technologies to promote distributed, collaborative, multidisciplinary research (Hine, 2006: vi). It is also defined as "using and processing information in different digital formats to gain new achievements and new scientific insights" (Shokrkhah, 2018: 231), with a special emphasis on collaboration-at-a-distance. E-science is reported to encompass "the use of advanced high-performance computing tools across the sciences" (Schroeder and Fry, 2007: 563), creating new objects, sites and contexts of knowledge (Hine, 2006) in a virtual, hence distributed and connected setting of knowledge production called e-infrastructure. It therefore refers to "the rise of new forms of large-scale distributed scientific enterprises supported primarily through advanced information infrastructures" (Lee et al., 2008: 1). The terms e-science, cyberinfrastructure, e-infrastructure (Ribes and Lee, 2010: 231), as well as grid computing, collaboratories (Lee et al., 2008: 1; Jankowski, 2007: 549), and cyberscience (Nentwich, 2003) are at times used interchangeably to refer to the technological mediation of scientific research within larger collaborative, distributed and multidisciplinary networks supported by ICTs. The revolutionary role of e-science is frequently celebrated and sometimes called the 'the fourth paradigm' (Hey et al., 2009) in which "data-driven, interdisciplinary research is augmenting the existing paradigms of experimental, theoretical and computational science" (Edwards et al., 2011: 67). The *E-infrastructure Report of the Swedish Research Council* defines e-science as "computationally and/or data-intensive science conducted on networked facilities enabling widespread collaboration" (Grönbeck et al., 2014: 19) and a "techno-human ecosystem" (Grönbeck et al., 2014: 22). This report emphasises the human component, the sociality embedded in cyber-structures. It thus gives more space to the 'human infrastructure' (Lee, 2006; Bietz et al., 2010) embedded in the relational ecosystem of the infrastructure (Star and Ruhleder, 1996; Star, 1999).

The immediate effect of the changes introduced with the advent of these techno-human e-science collaborations is stated to be "the redefinition of traditional disciplinary boundaries into vast domains of investigation," also referred to as "the big new sciences" (Ribes and Lee, 2010: 232). Yet it is not clear whether, in practice, disciplinary boundaries are actually blurred or redefined. Van Zundert's (2018: 2) work on Mirador, for instance, focuses on one "open-sourced, web-based, general-purpose image viewer written in JavaScript". It discusses the preservation of data silos partially due to "the institutional makeup of academia and its (grant) funding schemes favour local institution-level digitization and development […] Collaborative development between institutions is often frustrated by funding limitations, and moreover requires significantly more coordination effort than local development" (Van Zundert, 2018: 10). Our research is a contribution to this discussion on 'the institutional makeup of academia' and our results point towards the preservation, rather than reconfiguration of disciplinary silos, which in turn, contributes to the preservation of data silos. Our study, and the particular focus on the organizational logic, focuses on the interdisciplinary collaborations in which many different technologies are produced for scientific research purposes.

Research within the e-science platform we investigate here is classified on its webpage under the larger trans-disciplinary clusters of material science, life sciences, citizen earth and cornerstone technologies, the first three acting as domains to which cornerstone technologies are applied (see Fig. 1). Note that we write Platform with a capital P when we refer to the specific platform investigated, and use a small p when we refer to e-science platforms in generic terms.

While new computational methods, tools and applications are produced under the research cluster of 'cornerstone technologies', the three other research clusters constitute the so-called 'application areas' of these cornerstone technologies. In the Platform the cornerstone technologies are developed by researchers in the respective scientific computing, mathematics and computing science divisions and departments of the three universities involved. Other research clusters include scholars mainly from life sciences, material sciences, and environmental sciences in application-based projects. Application areas rather than disciplinary boundaries alongside computational technologies are thus emphasised in the public presentation of the Platform. Yet, as we shall see later, disciplinarity remains an important boundary that is preserved rather than blurred or reconfigured in the practical enactment of e-science on the Platform.

Methodology

The data discussed in this article come from participant observation of platform activities during the period September 2021-August 2022 and 45 semi-structured in-depth interviews conducted in March 2022-May 2022 with 18 women and 27 men researchers affiliated to the Platform. They ranged from PhD candidates to junior and senior faculty. All the participants were purposively selected according to the criterion that they had to be members of the Platform. They were thus not self-selecting. The first author initially approached the female researchers affiliated with the Platform in University B, the scientific computing program of the IT department of University A, and the female researchers included in the annual report of University C submitted to the Platform in 2021. She then approached all the researchers included in all three lists, and the PIs working for the Platform in disciplines other than scientific computing, computing science and mathematics in University A. The final list of 45 researchers included everybody who responded positively to the request for interview. She also participated in project presentations and seminars of the Platform. The research was conducted based on the premise that e-science is enactment, "pointing

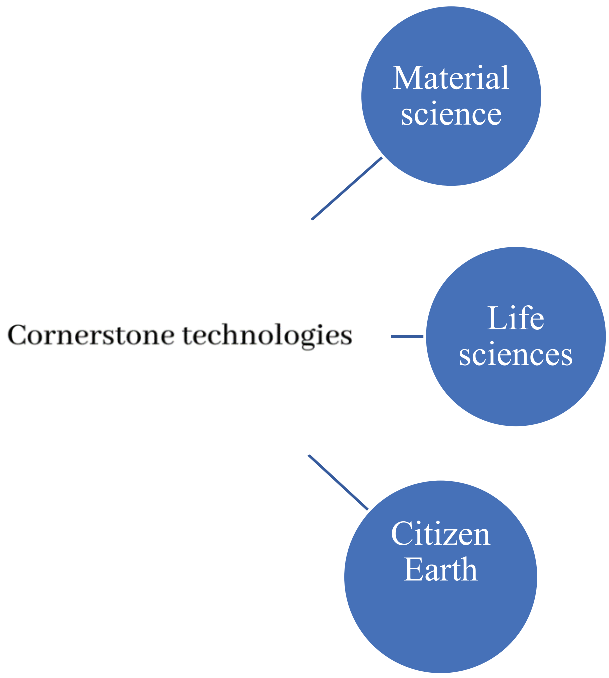

Figure 1. Research clusters of the Platform

to the importance of how socio-material sets of practices achieve and accomplish" e-science "as a meaningful phenomenon" (Bartlett et al., 2018: 2-3). The interviews were conducted either online or in one of the affiliated universities. The research was approved by Swedish Ethical Review Authority (Etikprövningsmyndigheten, No. 2022-00276-01). The interviewed researchers were provided with information sheets about the project and gave written consent for the use of their pseudonymised data in publications. The average length of the interviews was 60 minutes. The interviews covered questions on the interviewees' educational background and how they had entered e-science and the Platform, their interdisciplinary collaborations and associated challenges, the representation of women in their respective disciplines, and in their e-science collaborations, the reasons for the under-representation of women in certain STEM fields, obstacles to women's attraction to and retention in their respective disciplines and in e-science, suggestions to promote gender equality in their respective fields, in e-science, and in the Platform.

One thing that emerged immediately was that the group membership in the Platform was highly ambiguous. The first author observed three ways in which researchers were members of the Platform (see Fig. 2). The first was through working in the research group of a PI who received funding from the Platform, regardless of whether the researchers themselves were funded by the Platform or not. The second was working in the scientific computing division of one of the universities involved (University A), or in the Computing Science Department of another of these universities (University C). Thirdly, they could also be considered a member due to their involvement in a project that was partially or fully funded by the Platform. A considerable number of researchers were themselves unaware of the fact that they were deemed members of the Platform via one of the above affiliations. Due to this, and to the fact that there were no overarching comprehensive statistics on the Platform regarding its personnel, the following data only provide an approximative idea about the number of women involved in Platform activities in the three universities. In 2022 the scientific computing division at the IT Department of University A had 65 researchers in total (13 professors - five women and eight men; 24 teachers and researchers - four women and 20 men; 28 PhD students - nine women and 19 men). In the same year there were 157 researchers (37 women and 129 men) affiliated to the Platform in University B according to the official website of the university. University C had 59 researchers (eight women and 51 men) affiliated to the Platform in 2020 according to the annual report it submitted to the Platform[4]. These figures all indicate a significant under-representation of women on the Platform.

The interviews were audio-recorded and transcribed verbatim by the first author. At this stage the interviewees were pseudonymized by identifying them only by numerals, as we do in this article. The interview transcripts were analysed according to abductive data analysis (Timmermans and Tavory, 2012; Tavory and Timmermans, 2014). Taking a critical distance from grounded theory based on the assertion that "induction doesn't generate theory," (Timmermans and Tavory, 2012: 170); Timmermans and Tavory rather call for abduction. Going beyond the abductive/

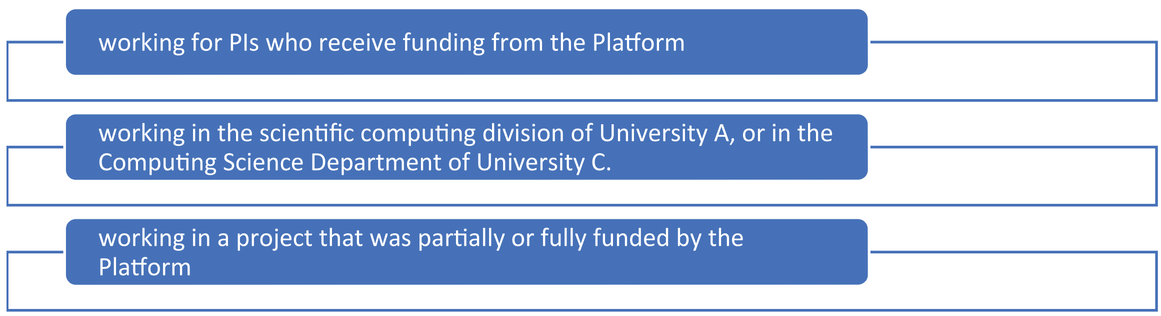

Figure 2. The membership structure of the Platform

deductive dichotomy, abductive analysis resorts to an iterative move between hypotheses on theoretically interesting cases and empirical data at successive stages of the analysis (Vila-Henninger, 2022).

> Once a hypothesis has been formed, deduction helps work out the hypothesis by providing a plausible generalization or causal chain. Induction constitutes the evaluation of the hypothesis because it provides the data that should conform to the deductively delineated premises (Timmermans and Tavory, 2012: 171).

Given this approach, the research started with the hypothesis that e-science collaborations could have the potential to attract female researchers from disciplines where they are better represented to ICTs. The data from the interviews were initially coded under the themes "academic background of researchers," "conceptualization of e-science", "challenges in e-science collaborations," "changes in knowledge production," "level of engagement and sense of belonging in the Platform," "strengths and weaknesses of the Platform," "representation of women in the Platform," "the level of involvement in interdisciplinary collaborations," "collaboration dynamics," "challenges in interdisciplinary collaborations," "gender mainstreaming opportunities in e-science collaborations," "horizontal gender segregation", "vertical gender segregation", "suggestions to promote gender equality in the Platform." These were then extended with codes emerging from the data under the themes of "boundary crossing practices," "the conception of interdisciplinarity", "hard/soft divide across the sciences", "gendered user/developer divide", "prevalence of a technical/engineering conception of collaboration," "limited mobility across sciences", and "the preservation of disciplinary silos". These inductively appearing new codes were then situated within the theoretical explanation in the literatures on women in science and STS, to be checked against the data.

In order to further check the emergent theme of "the preservation of disciplinary silos", co-authorship patterns of researchers at University B were also analysed based on the information provided on the university website. University B was chosen because their e-science collaborations exhibited the greatest disciplinary diversity. Affiliated researchers at University C mostly worked in technology development, with only few projects in other sciences. University A was involved both in technology development with researchers in the scientific computing division at the IT department and in research groups around PIs situated in theoretical chemistry, physics and biology departments. With the departure of the original PIs, University B eventually introduced open calls around projects exhibiting relatively more disciplinary diversity. Based on publications data on the webpage of University B, the disciplines of 113 researchers were compared with the disciplines of the top three researchers they frequently collaborated with. This sampling excluded PhD candidates who do not yet have enough publications for comparison purposes and researchers whose co-authorship data were not provided on the webpage.

FINDINGS

High recognition of horizontal gender segregation

The interviewed researchers articulated a high level of awareness of the horizontal gender segregation across their disciplines, attesting to the higher representation of women in biological sciences and medicine than in engineering, mathematics and IT. They also noted that this gender segregation was reproduced in their e-science collaborations where the so-called application areas of e-science projects involved more women, and the mathematics and IT-heavy work of technology development remained highly male-dominated. As one interviewee, typically, put it: "I've noticed that more men are working with this not application-based things, for some reason, they work on more theoretical stuff, and the applications, for some reason, women seem to think that's more fun" (Interviewee 15, woman, junior researcher).

The majority of the women researchers, as illustrated in the quotes below, thought that the reasons for this division were the gendered perceptions of scientific fields and gendered expectations regarding individual career trajectories:

I mean, I don't believe that women don't like technology. Because I know that many women do. Clearly, but I think that expectations play a larger role than what women actually want to do. I mean, because people don't expect you to want that. So, it's not so easy. I think the resistance from society, from your peers, I think this is a big factor at least. […] And maybe women feel more attracted to environmental sciences also because it's expected that it will be more attractive to women (Interviewee 4, woman, senior researcher).

I think it's the same thing [in computer science], similarly to math where it's seen as a boy's thing, when you are kids in the whole society. […] there is this perception around you all the time that even if people may not notice so much constantly tells you, that's not for you. That's not for you. You shouldn't do that. You're not good at that (Interviewee 39, woman, junior researcher).

E-science, a potential space for gender equality?

The majority of the interviewees thought that e-science, as a new area of research at the intersection of different disciplines and disciplinary cultures entails opportunities to break away from preconceived gendered ideas around math-intensive fields. One interviewee highlighted the potential of this newness to build more gender-equal communities of research:

In some sense, I think that the focus areas have shifted more towards data driven science. And maybe there are more women there because that's new, a newer science. I think that there is a tendency for the old areas to be more male-dominated. […] There are more opportunities [in newer science]. And, the hierarchies haven't formed themselves yet. So that's why I think there could be more [women] (Interviewee 6, woman, senior researcher).

Another interviewee referred to the prevailing notion of 'proper computer science' as a male gendered domain for people involved in computer hardware and software, and stated that application-oriented e-science does not fit this conception and therefore has the potential to attract more women.

I would say there would be more of an equal gender balance in the e-science than there would be in some of the, quote, more core kind of nerdy computer science, if you want to call it that. […] Because there'd be a little bit more breadth of sort of people involved and so on (Interviewee 31, man, senior researcher).

Applications with societal value such as environmental sustainability or cancer research were also thought to be of importance when it comes to attracting more women to e-science. An interviewee told the first author how physics' claims in the past, as a discipline, mostly focused on solving 'hard' problems and that might have deterred women, adding:

I think that if we put more emphasis on the value of sustainable technologies, and we need e-science people in material science, I think we will get more and more women to apply to our program (Interviewee 17, man, senior researcher).

E-science collaborations funded by the Platform indeed bring together researchers coming from and/or intersecting different communities of practice around similar problems. Hence, they have the potential to be spaces where more than one community of practice, both in terms of gender and disciplinary belonging, co-exist in their participants in a way that might alter the gendered perceptions and expectations around disciplines. Some of the interviewed researchers were involved in boundary-crossing practices through interdisciplinary collaborations, being specialized in computational sub-divisions of their disciplines, and/or changing disciplines. Most of the interviewees dated their entrance into e-science, either as users or developers of computational methods, back to their postgraduate studies. As exemplified below, they typically thought that e-science collaborations have the potential to attract more women scientists to computationally heavy areas of research, especially from disciplines with higher representations of women such as biological sciences and medicine:

One thing that in particular comes to mind is exactly life sciences. So, what happens now is that medicine uses increasingly computational

> techniques. And in medicine, […] I believe it's slightly more women than men who study medicine. But I mean, medicine definitely doesn't have a gender problem with regards to women. And so there and as well in biology. So traditionally, fields that would use computational science were engineering, physics, these types of things, then increasingly chemistry, and only let's say, in the last 20 years has it seriously started in biology and medicine. And so therefore, there is, I would say, at least in e-science in Sweden there I see an increase of women because of that (Interviewee 20, man, senior researcher).

Perpetuation of horizontal gender segregation

Yet, the potential hinted at above remained untapped in the Platform, and the existing horizontal gender segregation across the sciences was, as stated above, not bridged, but reproduced. There were four reasons, discussed below, for the perpetuation of horizontal gender segregation in our case study.

Limited mobility across disciplines with higher gender parity to male-dominated computational technology development

Interdisciplinary collaboration[5] was a key concern for the Platform which functioned mainly as a research funding distribution hub. Distribution of funding was mostly centred on a few PIs situated in their disciplines who also manifested strong engagement in Platform activities. They were the principal investigators (PIs) who collaboratively applied for the government's call to establish an e-science SRA (strategic research area). These PIs located in their respective disciplines used Platform funding to hire PhDs and postdoctoral researchers. Only relatively recently with the departure of the original PIs, did University B start to involve new PIs around e-science projects which exhibited a strong focus on application areas in cognitive sciences, life sciences, and environmental sciences.

In this context interdisciplinarity was enacted primarily in two ways. Firstly, funding was granted to PhD candidates and postdocs from new disciplinary constellations and application areas in e-science projects. This resulted in individual boundary-crossing across disciplines. Secondly, technology developers and domain specialists worked in parallel, doing their bits of works in collaborative e-science projects. This being the case, the mobility of the researchers from disciplines with higher gender parity to technology development in the disciplines of scientific computing and mathematics was very limited.

Of the 45 interviewed researchers, only five reported mobility across biological sciences or medicine and male-dominated fields of IT and engineering. Of these five researchers three were men who had entered IT-intensive fields such as bioinformatics and scientific computing from biological sciences and medicine (molecular biology and genetics, physiotherapy and public health, biology). The two women who were in this category, on the other hand, were situated in evolutionary biology with a background in mechanical engineering, and in scientific computing with a background in biology respectively. It is thus clear that women's mobility across biological sciences and medicine to IT-intensive fields remained very limited in the case of the interviewed Platform researchers.

Gender friction in e-science collaborations

The female researchers involved in the Platform experienced not only science friction, interoperability problems that occur when two or more disciplines work together on similar problems (Edwards et al, 2011), but also what we might call gender friction. Here we use the term gender friction to refer to the gendered aspect of interoperability problems in e-science interdisciplinary collaborations which adversely affect women. The literature on women in interdisciplinary STEM collaborations records certain gendered challenges. Zippel (2019), for instance, reports that existing institutional and organizational gender inequalities also permeate interdisciplinary STEM collaborations, and a gendered imaginary prevails in interdisciplinary collaborations, marked through terms such as 'patronizing helper', 'exploiter', 'partner' and 'friend'. This imaginary "reproduce[s] inequalities through symbols and practices" (Zippel, 2019: 1802). Griffiths et al. (2022: 233), on the other hand, state that "a survey of STEM faculty at a large public research university found that faculty from under-represented groups - in terms of gender, race, and sexual orientation - had more negative experiences with department-level research collaborations."

In our case study, the above-discussed nature of interdisciplinarity in the Platform, along with the need for constant self-training in the use of ICTs, posed certain challenges, especially for contingent junior female faculty. These challenges prominently included how much one actually needed to know about other disciplines one engages with in an interdisciplinary context. In an environment where interdisciplinarity was mostly conceived as technology developers' and domain specialists' working in parallel, it was rather ambiguous as to what it meant to be competent in a new discipline. The interviewee below talked of her impression of interdisciplinary work as "intruding on" another area of expertise, which was perceived to be harder for women.

> I think it's common that people in interdisciplinary topics, and maybe especially women, feel a little bit like they are intruding in someone else's field. […] like as an engineer, in a medical application, you feel that you don't know anything about medicine, and then you don't have anything to say about things there (Interviewee 1, woman, senior).

This was observed to go hand in hand with a high sense of self-responsibility especially among contingent female faculty who tended to overperform in e-science projects. "I think there has been a lot of this fear of not doing well enough," said Interviewee 13 about her work in her research group (PhD candidate). Interviewee 12, a PhD candidate with a background in engineering who started working in an evolutionary biology department as part of her e-science project, expressed issues she experienced mostly because the biology department which hired her was not well set up to conduct cross-disciplinary projects. "But me being on the fringe, I know that it's going to cause an issue because at every meeting we have on my progress, there's new information and new directives and new things that are applied," she said, adding "the issue is mine because I need to learn where I am right now." This becomes a challenge when interdisciplinary work mostly relies upon such individual cross-boundary action, and the disciplinary organizational structure of universities is sometimes not yet ready to accommodate such boundary-crossing actors (Griffin, 2022). In our case, especially junior women researchers internalized these issues which were not necessarily about them. They exhibited great degrees of self-responsibility, anxiety and stress.

Limited female mobility from biological sciences and medicine to IT fields, and gender friction restricted e-science borderlands' capacity to meaningfully alter asymmetrical gender divisions across disciplines in e-science projects. Furthermore, as we shall see below, even when there was gender parity in cross-disciplinary e-science collaborations, the gendered developer/user divide permeated research groups. Computational system/tool/method developers mostly were men, and computational self-reliance across disciplines acted as 'gendered boundary work' (Pereira, 2019; Vuolanto and Kolehmainen, 2020) to further strengthen the hard/soft divide across the sciences depending on their perceived proximity to mathematics and IT.

Gendered user/developer divide

The logic of domains and the user/developer divide

The above-mentioned conception of interdisciplinary work was operative under a particular organizational logic, namely the logic of domains, described by Ribes et al. (2019: 281) as "a de facto organizing principle for science policy and technology development". According to this logic, application areas in the Platform were classified as specific domains of action in which research was supported through 'cornerstone technologies' (see Fig. 1 above). Ribes et al. (2019) state that this logic envisages a 'domain independent' area of expertise, namely computing, information sciences and more recently data science, presumed to be universal and general. On the Platform website, the research cluster called cornerstone technologies bore this attribute of domain-independence, as the technologies were described as "transcendent" in relation to the domains of material sciences, life sciences and citizen earth.

Tool developers in the Platform also enacted this logic in how they developed generic models from particular datasets and/or vice versa. This was described by one interviewee as "tweaking aspects of the model so that they can latch on to this data" (Interviewee 24, woman, senior researcher). This logic was also apparent in how the interviewees conceptualized the need for the domain-independence of computational tools:

> Well, I mean, e-science is broader than what we're doing in computational sciences, what we're doing in chemistry for instance, because then we are sort of focused on methods that give chemistry results. And then, of course, in mechanics, they focus on things that sort of solve mechanics problems. So, the methods are quite distinct, there are similarities, but they are doing sort of different things. E-science collects all of these, and also puts the focus on the methods rather than the discipline. So that's a new thing about e-science. It sort of creates a network above the disciplines, a full umbrella zone of the disciplines, and connects people (Interviewee 15, woman, junior researcher).

The discourse of supporting sciences through e-science whose computational tools remain generic, domain-independent and beyond scientific disciplines prevailed among the tool developers in scientific computing, mathematics and computing sciences. "We try to support emerging science" said a senior researcher, adding "So, it is part of our mission to make sure that all sciences can access computational resources that are needed" (Interviewee 4, woman, senior researcher).

This organizational logic, along with the particular enactment of interdisciplinarity mostly relying on working in parallel, within one's disciplinary boundaries, around a common problem, perpetuated a developer/user divide within interdisciplinary e-science collaborations. It was common to observe that technology developers referred to scientists in application areas as their 'users':

> It's difficult to characterize what exactly people need, you know, I mean, when your user comes and says, "I need this to work." "Okay, what do you mean, by saying work?" And it's difficult for people who don't know how this works (Interviewee 3, woman, senior researcher).

This showed that computational technology development was conceived as an engineering problem, and e-science as a form of technology

Table 1. The 10 researchers at University B who frequently co-authored with researchers from another discipline.

Researcher	Department	Frequently co-authors with (as per top 3 collaborators)	Discipline
Researcher (male)	Economics and Management	1) Researcher, (male) 3) Researcher, (male) The 2nd collaborator is from the same department.	1. Medicine 3. Clinical Chemistry and Pharmacology
Researcher (male)	Mathematics (Faculty of Science)	1) Researcher, (male), • 2nd and 3rd most frequent collaborators are from the same department	1. Economics
Researcher (male)	Astrophysics	1) Researcher, (male) • 2nd and 3rd most frequent collaborators are from the same department	1. Mathematics
Researcher (male)	Mathematics (Faculty of Science)	All three male researchers	Astrophysics
Researcher (female)	Geology	All three male researchers	Astrophysics
Researcher (male)	Bioinformatician, at the Faculty of Medicine	1) Researcher (male) 2) Researcher (male) 3) Researcher (female)	**1.** IT **2.** Electrical and Information Technology **3.** Physics
Researcher (male)	Scientific and technical computing	All three male researchers	Structural mechanics
Researcher (male)	Mathematical statistics	All three male researchers	Physical geography and ecosystem science
Researcher (male)	Data scientist at the Faculty of Medicine	3 male researchers	Biochemistry and structural biology
Researcher (male)	Mathematics (Faculty of Engineering)	Two male, one female researchers	Communications engineering

transfer to support computationalisation trajectories of scientific disciplines rather than a research innovation which reconfigured disciplinary boundaries. The impression was

> that the level of ambition is not about bringing the disciplines into e-science, or [bridging] that gap that we were talking about [between scientific disciplines and e-science], but rather facilitating the use of e-science across disciplines, but still within their disciplinary silos. So, [this platform] isn't really providing a platform for, you know, dissolving the boundaries between those disciplinary silos, but rather, it's about increasing the accessibility of e-science within each discipline (Interviewee 32, man, senior researcher).

The preservation, rather than reconfiguration, of disciplinary silos was also visible in the co-authorship patterns of affiliated researchers. The top three collaborators of the vast majority of 113 researchers at University B (see the section on methodology above for selection criteria), were from their own departments or centres. Only 10 researchers, of whom only three were female, were recorded to frequently co-author with researchers from other disciplines (see Table I). 7 researchers were involved in co-authorship practices with researchers from other disciplines to a lesser extent.

This also led to the preservation of disciplinary cultures as attested by the interviewee cited below.

> So also in [this Platform], in e-science platforms, do you think that scientists inherit the culture of their own disciplines?
>
> I think so. Yes. Or how do they blend? Do they change each other? Do they interact? Perhaps a bit, but I think which department you are in is important. And then of course, it depends if you are dominating the department or the department is dominating you. So, it depends on the size of the groups also. But yes, I think the culture is more tied to domains than to what you actually do (Interviewee 4, woman, senior researcher).

Although e-science brings together scientific disciplines with varying degrees of gender diversity, hence has the potential to act as a borderland in which communities of practice intersect in one person and researchers are exposed to one another's 'disciplinary' (Traweek, 1988) or 'epistemic' (Knorr-Cetina, 1999) culture, this potential was not realized in this Platform. Disciplinary silos were largely preserved and e-science solutions were mostly conceived in terms of engineering problems around technology transfer. One result of this is that a gendered user/developer divide permeated these e-science collaborations, which reflected the traditional gender divisions across the disciplines.

Gendered technology user/developer divide in the platform

The enactment of the logic of domains, the nature of interdisciplinarity in the Platform, and the resulting preservation of disciplinary silos all meant that the already existing horizontal gender segregation across the disciplines was reproduced. A closer look at e-science projects in this Platform not only in terms of the numeric representation of women but also, and especially, the type of work conducted within the interdisciplinary research projects showed that existing gender divisions across the disciplines permeated the e-science projects. The interviewees typically reported that the task of computational technology development which involved the theoretical work of numerical analysis and computational simulation, among others, remained highly male-dominated. Interviewees from quite different fields of research involved in e-science collaborations, such as the examples below, all stated this.

> I know plenty of women in astronomy, who get involved with sort of e-science and big data. And they are quite happy with it, and they do fine. But also, I know that, within astronomy, studies that are more focused on stars [...] tend to be a much friendlier field [for women], rather than, say, cosmology, or [working on] things that are very distant in the universe, or things like cosmological simulations, which are just theoretical computer simulations of the universe [...] that field is a little bit more male-dominated, and I guess a little less friendly than, say, fields using stellar data. [...] *Observation of stars, you know, requires a lot of work, but you're also sort of limited to the data that you*

> *get from stars that are available or, you know, the instruments that you use.* And so, I think somehow, it's not as personal. The result doesn't reflect what you think. And so, in that way, *the theoretical fields are the fields where you create these huge simulations*. I think it tends to build an environment that is much more protective of your own data. And, and a little bit more guarded of your own science [Interviewee 9, woman, senior researcher, emphasis added].

Here, the interviewee drew on the distinction between the work of mere observation, the "use" of observed data, as well as instruments, and the theoretically heavy work of designing computational simulations. We see how within the same discipline, the work relying on the 'use' of data and computational tools, and the development/design of these tools remains gendered.

> Yeah, we have like, groups, I'm in the systems [system development] group. So, we are 8 people and there are two women if I recall well, yeah, and then there is a bioinformatics group, where it's four people and there are no women, and then there are like more lab-oriented groups, which I don't know as much because I don't interact with them as much. But there, I think, there are many more women (Interviewee 35, man, junior researcher).

Here again, we see that in the same life sciences centre - life sciences being a STEM field with relatively higher gender parity - the work of technology development [system development] and bioinformatics remained highly male-dominated.

One interviewee who thought that e-science collaborations have limited capacity to contribute to an increase in the number of women in computational tool development referred to the problem of their inclusion in e-science collaborations as users and not as developers:

> [Women] have to learn something because they're using. But they will never become a developer. They may say to the developer, "Look, here, you have done lousy work, change it, because we don't use this" and things like that. […] Some people from computer science will teach the biologists. Yes, sure. But this will not lead to more people, female people in computer science. Of course, synergies are great, there will be something, there will be some people who learn biology and vice versa. And start programming and so on, sure, but it's not going to solve the major problem (Interviewee 7, woman, senior researcher).

Thus, even when there was gender parity in an interdisciplinary e-science research group, the traditional gender division across the math-intensive and non-math-intensive divide (Ceci et al., 2014) seemed to be reproduced and not mitigated in e-science collaborations.

Gendered boundary work around self-reliance in computational tool development

The term boundary work was originally developed to refer to rhetorical tools used by scientists to distinguish science from non-science in a time when modern sciences aspired to distinguish themselves from religion and technical know-how for claims of authority (Gieryn, 1983). In time, the term came to be used also to define practices and discourses that serve to create distinctions and boundaries across and within sciences. There are also studies which discuss boundary work that occurs as a response to new technologies for scientific research (see Burri, 2008; Reyes-Galindo L., 2016, among others). Recently, the gendered character of boundary work has started to be analysed (Pereira, 2019; Vuolanto and Kolehmainen, 2020). Below, we discuss how distinctions made across sciences with respect to their perceived computational self-reliance acted as gendered boundary work in the Platform which solidified the gendered hard/soft divide.

It was common among male interviewees to classify disciplines along a scale depending on the disciplines' proximity to mathematics and computing. Physics and theoretical chemistry were two disciplines which were perceived to be close to mathematics and computing. The presumed self-reliance regarding computational tools and methods development especially in physics, but also in theoretical chemistry, served as a boundary work for the scientists to draw boundaries around their disciplines to reinforce their authority. For example, the requirement to have discipline-specific knowledge was very much accentuated in the case of physics; it was cited as the reason why it is physicists themselves who need to develop their computational tools.

One interviewee (Interviewee 19, man, senior researcher) told the first author how a computer scientist who was hired by the research group to do programming to address their scientific problems failed to be efficient, as he did not know the problems in the field, and could not write hundreds of lines of codes at once. He then added the story of when CERN opted to buy commercial software instead of asking physicists to do the programming:

> […] This was especially true at CERN, because at CERN, actually it's a bit of a funny story, but it's like 30 years ago almost, right? When they started to plan for this new collider, they said, "We don't want to program things at CERN, we want to buy commercial software." And so, for 10 years, they had a big investment in commercial software, because they said, "We don't want physicists to write the program." But in the end, it turned out that this commercial software didn't really deliver. Because they didn't understand the problem we faced. And so, RUTH, this program that we use today, was really started as kind of like a renegade project, it was not really sanctioned by the management, they really looked down upon it for many years. But the problem was that they knew exactly what we needed, right? So, they made a program that could do all the things we needed, whereas other people made maybe more beautiful programs, but they couldn't do what we needed to do, right? (Interviewee 19, man, senior researcher).

The same requirement for discipline-specific knowledge was not as much highlighted in biological sciences and in medicine. There, just as in the example below, the emphasis was on the researchers' dependence on tool developers outside of their discipline.

> So, that's where I think a platform … could fulfil an important role because we may have quite uneven formal training and uneven knowledge of [computational] methods and, previously your research was normally more focused and now we are forced to do research that is a lot more complex and you need to be quite good at almost everything, but you're not very good at anything, you are kind of more superficial sometimes (Interviewee 5, woman, senior researcher).

In this discourse of varying levels of computational self-reliance and confidence across disciplines, it was observed that the gendered hard/soft divide between the sciences was reinforced. Hence, the presumed computational savviness and self-reliance of a discipline was used as gendered boundary work to underline how hard or soft a discipline was.

Some male interviewees drew the boundary between physics and biology as to how deterministic or stochastic their computational models were. In the quote below, the presumed precision of correspondence between real-life interactions and computational models in physics - i.e., deterministic over stochastic - which enables the "staging" of a physical action (Knorr-Cetina, 1999: 34)- was used as boundary work between physics and biology:

> Computationalisation of scientific disciplines is related to how deterministic or stochastic their models are, how much noise they incorporate. Models are more deterministic in physics and less so in biology, also leading to how suitable their problems are to being computationally simulated (Interviewee 26, man, senior researcher).

Another male interviewee associated the different pace of diverse sciences in adopting mathematical and computational models, or their computationalisation, to how hard/soft they were supposed to be, reformulating the boundary in terms of the hard/soft divide:

> If you put like all kinds of sciences, so to say, like on a scale with the hard sciences at the bottom and the soft, softer sciences at the top, you could see, along this scale, people started to use more and more mathematical models, and that's what I would qualify as e-science, this use of maths to model a problem (Interviewee 35, man, junior researcher).

Researchers were aware that biology was labelled as 'not hard'. A junior woman researcher stated that in high school, biology was considered a "loss for science because it's not a hard science" and the overall feeling she got was that "biology was a bit deprecated as a science, it was not a pure, hard science" (Interviewee 13).

A male interviewee was quick to associate the higher number of women in biology with the discipline being less math-intensive:

> So why do you think we have more women in biology?
>
> I don't know. It's less maths maybe. If I would say that. So, if one would label, say this physics and maths, they are more male-oriented disciplines, then biology would be the opposite of it (Interviewee 30, man, junior researcher).

Overall, as opposed to physics and theoretical chemistry, computational competence in biology was in general perceived to be low:

> [In biology] they are kind of in a less privileged situation. In physics, we could help ourselves [in computational tool development], while in biology, they probably can't, so the more dire need for this kind of organization falls in those departments (Interviewee 3, woman, senior researcher).

While chemistry was also deemed less math-intensive and 'softer', there was clear gendered boundary between laboratory work and computational chemistry. "For some reason, theory is not attractive [for women]", said an interviewee, adding that "it could be that what attracts females to chemistry is sort of the chemistry of doing things with your hands in a way, working with sort of practical things" (Interviewee 2, man, senior researcher). He stated that chemistry in that sense was closer to biology and "kind of a softer subject", and added that it could be the reason why they needed to recruit PhD students to work on e-science projects from physics and other departments. This was somehow in conflict with his previous statements on the self-reliance of chemistry when it comes to developing computational tools to solve problems in the field. Yet, it is illustrative of the way in which perceived computational self-reliance is used to draw boundaries both between biology and chemistry, the first otherwise stated to be close to chemistry, and between laboratory work in chemistry and computational chemistry.

In the case of the Platform, the perceived or real computational self-reliance of a discipline was mobilized to draw boundaries both across the sciences and between theoretical/computational and wet-lab work within the same science, all in line with the hard/soft divide. Given the gendered nature of the hard/soft divide across the sciences, which goes hand in hand with the gender division both within and across the sciences, computational self-reliance also acted as gendered boundary work for claims of authority. This gendered boundary work around disciplinary computational self-reliance hence strengthened the gender division across the scientific disciplines.

Discussion and conclusion

In this article, we attend to situated practices and meaning-making around technology development and use in a particular e-science platform in Sweden to account for the extent to which existing horizontal gender division across the sciences is enacted in the borderland opened by e-science collaborations around the use and development of the state-of-the-art computational tools. This can also be reframed in terms of the more general question of "shifts in power relations around knowledge" (Wouters et al., 2013: 3) and the possibility of social transformation within existing inequality regimes (Acker, 2006). These inequality regimes have become persistent and resistant to change, especially in the case of gender imparity in computer sciences and ICTs, and gender divisions across the sciences.

Although there exist discussions and data on the possible "levelling effect" of new e-science technologies (Hine, 2006: xvi) when it comes to enhancing female participation in scientific endeavours and the workforce (Palackal et al., 2006; Oleksy, 2012; Oladejo et al., 2021), and the career advancement of female academics (Ojo et al., 2015), the findings from the Platform offer a grim response to the question of whether the position of women in science is changing with this new technology (Kretschmer and Aguillo, 2005). They attest to the perseverance of the traditional horizontal gender segregation, and gender inequalities across disciplines, including within e-science collaborations.

Although Platform members articulated the notion of a potential for e-science collaborations to attract more women to computationally supported research, one could see that in practice this potential remained largely untapped and the existing horizontal gender segregation was perpetuated through the following mechanisms which we also used to structure this article: a) mobility across disciplines with asymmetrical gender divisions remained limited; b) gender friction, gender-specific problems suffered by women in interdisciplinary collaborations, took its toll on female researchers; c) traditional gendered divisions across scientific disciplines permeated e-science collaborations and perpetuated the gendered technology developer/user divide where developers mostly remain men[6]; and d) different levels of self-reliance in technology development across disciplines and the perception of scientific fields' proximity to IT and maths acted as 'gendered boundary work' (Pereira, 2019; Vuolanto and Kolehmainen, 2020). All this reinforces the gendered hard/soft science divide. Disciplinary silos were preserved rather than blurred or reconfigured. Technology development was deemed an engineering problem, and e-science computational technology transfer, rather than a reconfiguration of disciplinary boundaries. This all mitigated the potential of boundary crossing to alter existing gender asymmetries in the sciences. How can we account for this persistence of gender asymmetries in new technology formations such as e-science?

It is well known in STS that new technologies are "built on an installed base" (Star, 1999: 382), a base that also includes existing asymmetrical social relations across 'the human infrastructure' (Lee, 2006; Bietz et al., 2010). This needs to be kept under consideration, especially in the case of technologies which enable disembodied and distributed communication in the virtual or cyber space. Virtual space is closely tied to existing inequalities in the broader social world and it supplements rather than completely replaces real-life interactions (Woolgar, 2002). That space "reflect[s] and reinforce[s] existing social orders, expressing and materializing hierarchical relations" (Davis et al., 2021: 1). Hence the belief in e-science's potential to mitigate gender asymmetries across sciences has, in the Platform under study, turned out to be the 'cruel optimism' that Lauren Berlant (2011) invokes to characterize the failed promises held out to women of the possibilities of inclusion under changing conditions.

Our case study turns our attention to the fact that new technologies are assembled and become embedded in the existing techno-human infrastructure, and do not create a ground-zero for social emancipation. To be able to tap into the potential of cross-disciplinary e-science collaborations to meaningfully bridge the gender gap across the sciences, we need concerted efforts and collective positive action at organisational level. These efforts will inevitably need to address the gendered technology user/developer divide and the way interdisciplinarity is practically enacted. They will need to analyse the repercussions of the prevalence of the logic of domains as an organisational principle in e-science initiatives, and tackle the ways in which computational self-reliance in the sciences with respect to computational tool development acts as gendered boundary work. This points towards the requirement for future studies on the extent to which disciplinarity remains important in e-science collaborations, on the gendered aspects of epistemic cultures (Knorr-Cetina, 1999), and on how and whether the computationalisation of the sciences alters existing epistemic cultures, or creates new ones, with possibly different gender relations.

References

Acker J (2006) Inequality regimes: Gender, class and race in organizations. *Gender & Society*, 20(4): 441-464.

Alegria S, Anderson-Knott M, Berheide C, et al. (2016) *Pathways, Potholes and the Persistence of Women in Science: Reconsidering the Pipeline.* London: Lexington Books.

Barone C (2011) Some things never change: Gender segregation in higher education across eight nations and three decades. *Sociology of Education* 84(2):157-176.

Bartlett A, Lewis J, Reyes-Galindo L and Stephens N (2018) The locus of legitimate interpretation in big data sciences: lessons for computational social science from -omic biology and high-energy physics. *Big Data & Society* (1): 1-15.

Bauschspies W and Puig de la Bellacasa B (2009) Feminist science and technology studies: A patchwork of moving subjectivities. An interview with Geoffrey Bowker, Sandra Harding, Anne Marie Mol, Susan Leigh Star and Banu Subramaniam. *Subjectivities* 28: 334-344.

Beaulieu A and Leonelli S (2021) *Data and Society: A Critical Introduction*. London: Sage Publications.

Begeny CT, Ryan MK, Moss-Racusin CA and Ravetz G (2020) In some professions, women have become well represented, yet gender bias persists - perpetuated by those who think it is not happening. *Science Advances* 6(26): eaba7814. DOI: 10-1126/sciadv.aba7814.

Berg AJ (1994) A Gendered socio-technical construction: The smart house. In: Cockburn C and Furst-Dilic R (eds) *Bringing Technology Home: Gender and Technology in a Changing Europe*. Buckingham: Open University Press, pp. 165-180.

Berg AJ and Lie M (1995) Feminism and constructivism: Do artifacts have gender? *Science, Technology & Human Values* 20: 332-351.

Berlant L (2011) *Cruel Optimism*. Durham: Duke University Press.

Bietz MJ, Baumer EPS and Lee C (2010) Synergizing in cyberinfrastructure development. *Computer Supported Cooperative Work* 19: 245-281.

Boston Consulting Group Gamma Study (2019) *What's keeping women out of data science?* Available at: https://www.bcg.com/publications/2020/what-keeps-women-out-data-science (accessed 12 February2022).

Bowker GC (1994) *Science on the Run: Information Management and Industrial Geophysics at Schlumberger*, 1920-1940. Cambridge, MA: MIT Press.

Bowker GC, Baker K, Millerand F and Ribes D (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger J, Klastrup L and Allen M (eds) *International Handbook of Internet Research*. Dordrecht, Heidelberg, London and Newyork: Springer Science and Business Media BV, pp. 97-117.

Burri RV (2008) Doing distinctions: Boundary work and symbolic capital in radiology. *Social Studies of Science* 38(1): 35-62.

Castells M (2012) *Networks of Outrage and Hope: Social Movements in the Internet Age*. Cambridge, Massachusetts: Polity Press.

Ceci SJ, Ginther DK, Kahn S and Williams WM (2014) Women in academic science: A changing landscape. *Psychological Science in the Public Interest* 15: 75-141.

Cheryan S, Ziegler SA, Montoya AK and Jiang L (2017) Why are some STEM fields more gender balanced than others? *Psychological Bulletin* 143(1): 1-35.

Corneliussen HG (2021) Unpacking the Nordic gender equality paradox in ICT research and innovation. *Feminist Encounters: A Journal of Critical Studies in Culture and Politics* 5(2), art. 25.

Dasgupta N and Stout JG (2014) Girls and women in science, technology, engineering and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences* 1(1): 21-29.

Davis JL, Williams A and Yang MW (2021) Algorithmic reparation. *Big Data & Society* 8(2). https://doi.org/10.1177/20539517211044808 (accessed 9 June 2023).

Edwards PN, Mayernik MS, Batcheller AL, Bowker GC and Borgman CL (2011) Science friction: Data, metadata, and collaboration. *Social Studies of Science* 41(5): 667-690.

Ensmenger N (2010) *The Computer Boys Take Over: Computers, Programmers and The Politics of Technical Expertise*. Cambridge, MA: MIT Press.

Faulkner W (2001) The Technology Question in Feminism: A View from Feminist Technology Studies. *Women`s Studies International Forum* 24(1): 79-95.

Fisher CR, Thompson CD and Brookes RH (2020) Gender differences in the Australian undergraduate STEM student experience: A systematic review. *Higher Education Research & Development* 39(6): 1155-1168.

Galison P and Stump DJ (1996) *The Disunity of Science: Boundaries, Contexts and Power*. Redwood City, CA: Stanford University Press.

GenderInSITE (2021) *Gender Equality in Science: Inclusion and Participation of Women in Global Science Organizations*. Gender in Science, Innovation, Technology and Engineering Report, September, ISBN: 9788894405446.

Gieryn TF (1983) Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review* 48(6): 781-795.

Gillard H, Howcroft D, Mitev N, Richardson H and Ferneley E(2007) Shaping the global economy: Gender, ICTs and development research. In: *Proceedings of the 9th International Conference on Social Implications of Computers in Developing Countries*, Sao Paulo, May 2007, Brazil.

Gillard H, Howcroft D, Mitev NN and Richardson H (2010) 'Missing women': Gender, ICTs and the shaping of the global economy. *Information Technology for Development* 14(4): 262-279.

Griffin G (2021) Feminizing innovation: Challenges in science and technology studies (STS). *Feminist Encounters* 5(2): art. 24.

Griffin G (ed) (2022) *Gendered Inequalities in Tech-Driven Research and Innovation: Living the Contradiction.* Bristol: Policy Press.

Griffiths EE, Mickey EL and Dasgupta N (2022) A 'chillier' climate for multiply marginalized STEM faculty impedes research collaboration. *Sex Roles* 86(3): 233-248.

Grönbeck H, Hulth P, Johnsson L et al. (2014) *Swedish Science Cases for E-Infrastructure*. Stockholm: Vetenskapsrådet.

Hey T, Transley S and Tolle K (2009) *Fourth Paradigm.* Microsoft Research.

Hine CM (2006) *New Infrastructures for Knowledge Production: Understanding E-Science.* Hershey: Information Science Publishing.

Jankowski NW (2007) Exploring e-science: An introduction. *Journal of Computer-Mediated Communication* 12: 549-562.

Kelan EK (2007) 'I don't know why': Accounting for the scarcity of women in ICT work. *Women's Studies International Forum* 30: 499-511.

Knorr-Cetina K (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Kretschmer H and Aguillo IF (2005) New indicators for gender studies in web networks. *Information Processing and Management* 41: 1481-1494.

Lagesen VA (2007) The strength of numbers: Strategies to include more women into computer science. *Social Studies of Science* 31(1): 67-92.

Lagesen VA (2008) A cyberfeminist utopia? Perceptions of gender and computer science among Malaysian women computer science students and faculty. *Science, Technology & Human Values* 33: 5-27.

Lee CP (2006) The human infrastructure of cyberinfrastructure. In: Alberta B (ed) *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work*, Canada, 4-8 November 2006, pp. 438-492.

Lee CP, Bietz M and Ribes D (2008) Designing cyberinfrastructure to support science. In: *Computer Supported Cooperative Work Conference Workshop*, San Diego, USA.

Light AE, Benson-Greenworld TM and Diekman AB (2022) Gender representation cues labels of hard and soft sciences. *Journal of Experimental Social Psychology* 98: 104234.

Mellström U (2009) The intersection of gender, race and cultural boundaries, or why is computer science in Malaysia dominated by women? *Social Studies of Science* 39(1): 885-907.

Michell D, Szorenyi A, Falkner K and Szabo C (2017) Broadening participation not border protection: How universities can support women in computer science. *Journal of Higher Education Policy and Management* 39(4): 406-422.

Mökander J and Schroeder R (2021) AI and social theory. *AI & Society*, http://doi.org/10.1007/s00146-021-01222-z (accessed 9 June 2023).

Montgomery CM (2012) Making prevention public: The Co-production of gender and technology in HIV prevention research. *Social Studies of Science* 42(6): 922-944.

Moss-Racusin CA, Sanzari C, Calouri N and Rabasco H (2018) Gender bias produces gender gaps in STEM engagement. *Sex Roles* 79: 651-670.

Nentwich M (2003) *Cyberscience: Research in the Age of the Internet*. Vienna: Austrian Academy of Sciences.

Ojo JO, Odunlade RO and Adekanye EA (2015) Impact of information communication technologies (ICTs) on career advancement of women academics. *Information Impact: Journal of Information and Knowledge Management* 6(3): 34-50.

Oladejo AI, Nwaboku NC, Okebukola PA and Ademola IA (2021). Gender difference in students' performance in chemistry: Can computer simulation bridge the gap? *Research in Science & Technological Education*. Available at: http//doi: 10.1080/02635143.2021.1981280 (accessed 10 November 2021).

Oleksy W, Just E and Zapedowska-Kling K (2012) Gender issues in information and communication technologies (ICTs). *Journal of Information, Communication and Ethics in Society* 10(2): 107-120.

Oudshoorn N, Saetnan AR and Lie M (2002) On gender and things: Reflections on an exhibition on gendered artifacts. *Women`s Studies International Forum* 25(4): 471-483.

Paine D and Lee C (2020) Coordinative entities: Forms of organizing in data intensive science. *Computer Supported Cooperative Work* 29(2): 335-380.

Palackal A, Anderson M, Miller BP and Shrum W (2006) Gender stratification and e-science: Can the internet circumvent patrifocality? In: Hine CM (ed) *New Infrastructures for Knowledge Production: Understanding E-science*. Hershey: Information Science Publishing, pp. 246-272.

Pereira MM (2019) Boundary work that does not work: Social inequalities and the non-performativity of scientific work. *Science, Technology & Human Values* 44(2): 338-365.

Plant S (1995) The future looms: Weaving women and cybernetics. *Body & Society* 1(3-4): 45-64.

Reuben E, Sapienza P and Zingales L (2014) How stereotypes impair women's careers in science. In: *Proceedings of the National Academy of Sciences* 111(2). Available at: http://doi.10.1073/pnas.1314788111 (accessed October 2021).

Reyes-Galindo L (2016) Automating the horae: Boundary-work in the age of computers. *Social Studies of Science* (46)4: 586-606.

Ribes D (2019) STS, meet data science, once again. *Science, Technology & Human Values* 44(3): 514-539.

Ribes D and Lee CP (2010) Sociotechnical studies of cyberinfrastructure and e-research: Current themes and future trajectories. *Computer Supported Cooperative Work* 19: 231-244.

Ribes D, Hoffman AS, Slota SC and Bowker GC (2019) The logic of domains. *Social Studies of Science* 49(3): 281-309.

Rua-Gomez C and Arias-Gaviria J (2020) *Why so Few? The Gender Representation Gap in Academic STEM Fields, a System Dynamics Approach*. In: 38th International Conference of the System Dynamics Society (Virtual), 19-23 July.

Santos EDD, Albahari A, Diaz S, and De Freitas ES (2021) Science and technology as feminine: Raising awareness about and reducing the gender gap in STEM careers. *Journal of Gender Studies* 31(4): 505-518. Available at: http//doi:10.1080/09589236.2021.1922272 (accessed 5 October 2021).

Sax LJ, Lehman KJ, Jacobs JA et al. (2017) Anatomy of an enduring gender gap: The evolution of women's participation in computer science. *The Journal of Higher Education* 88(2): 258-293.

Schroeder R and Fry J (2007) Social science approaches to e-science: Framing an agenda. *Journal of Computer-Mediated Communication* 12: 563-582.

Shokrkhah Y (2018) Cross border e-science and e-search partnership: Bridging the gap between science and media. *Journal of Cyberspace Studies* 2: 229-237.

Star SL (1999) The ethnography of infrastructure. *American Behavioural Scientist* 43(3): 377-391.

Star SL (2015) Misplaced concretism and concrete situations: Feminism, method, and information technology. In: Bowker JC, Timmermans S, Clarke AE and Balka E (eds) *Boundary Objects and Beyond: Working with Leigh Star*. Cambridge, MA: The MIT Press, pp. 143-171.

Star SL and Ruhleder K (1996) Steps towards an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research* 7(1): 111-134.

Stearns E, Bottia MC, Giersch J, et al. (2020) Do relative advantages in STEM grades explain the gender gap in selection of a stem major in college? A multimodel answer. *American Educational Research Journal* 57(1): 218-257.

Stoet G and Geary DC (2018) The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science* 29(4): 581-593.

Su R and Rounds J (2015) All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology,* Available at: http://doi: 10.3389/fpsyg.2015.00189 (accessed 20 October 2021).

Suchman L (2002) Located accountabilities in technology production. *Scandinavian Journal of Information Systems* 14(2): 91-105.

Tavory I and Timmermans S. 2014. *Abductive Analysis: Theorizing Qualitative Research*. Chicago: University of Chicago Press.

Telenor (2019) *Programmed out: Gender gap in technology in Scandinavia*. Telenor Report, Plan International Norway and Telenor Group, 2019. Available at: https://www.telenor.com/wp-content/uploads/2019/09/The-Gender-Gap-in-Technology-in-Scandinavia_Full-report.pdf (accessed 9 June 2023)

Timmermans S and Tavory I (2012) Theory construction in qualitative research from grounded theory to abductive analysis. *Sociological Theory* 30(3): 167-86.

Traweek S (1988) *Beamtimes and Lifetimes: The World of High Energy Physicists*. Cambridge, MA: Harvard University Press.

Van Veelen R, Derks B and Endedijk MD (2019) Outnumbered and negatively stereotyped threatens career outcomes of women in STEM. *Frontiers in Psychology* 10(150) Available at http://10.3389/fpsyg.2019.00150 (accessed 3 October 2021).

Van Zundert J J (2018) On not writing a review about Mirador: Mirador, IIIF, and the epistemological gains of distributed digital scholarly resources. *Digital Medievalist* 11(1): 1-48.

Vila-Henninger L, Dupuy C, Van Ingelgom V et al. (2022) Abductive coding: Theory building and qualitative re(analysis). *Sociological Methods & Research* 0(0): 1-34.

Vitores A and Gil-Juarez A (2016) The trouble with 'women in computing': a critical examination of the deployment of research on the gender gap in computer science. *Journal of Gender Studies* 25(6): 666-680.

Vuolanto P and Kolehmainen M (2020) Gendered boundary-work within the Finnish skepticism movement. *Science, Technology & Human Values* 46(4): 789-814.

Wacjman J (2000) Reflections on gender and technology studies: In what state is the art? *Social Studies of Science* 30(3): 447-464.

Wacjman J (2004) *Technofeminism*. Cambridge: Polity Press.

Wacjman J (2007) From women and technology to gendered technoscience. *Information, Communication and Society* 10(3): 287-298.

Wacjman J (2010) Feminist theories of technology. *Cambridge Journal of Economics* 34:143-152.

Wang MT and Degol JL (2017) Gender gap in science, technology, engineering and mathematics (STEM): Current knowledge, implications for practice, policy and future directions. *Educational Psychology Review* 29(1): 119-140.

Woolgar S (2002) *Virtual Society? Technology, Cyberbole, Reality*. Oxford: Oxford University Press.

Woolgar S and Coopmans C (2006) Virtual witnessing in a virtual age: A prospectus for social studies of e-science, In: Hine CM (ed) *New Infrastructures for Knowledge Production*. Hershey: Information Science Publishing, pp. 1-25.

Wouters P, Beaulieu A, Scharnhorst and Wyatt S (2013) *Virtual Knowledge: Experimenting in Humanities and the Social Sciences*. London: MIT Press.

Zacharia ZC, Hovardas T, Xenofontos N and Pavlou Y (2020) *Education and Employment of Women in Science, Technology and Digital Economy, Including AI and Its Influence on Gender Equality*. Brussels: Report for the European Parliament, PE 651.042, April 2020.

Zippel K (2019) Gendered ımages of ınternational research collaboration. *Gender, Work & Organization* 26(12): 1794-1805.

Notes

1 The report uses a complex methodology covering all continents (see Gender InSite, 2021: viii).

2 There may, of course, be other reasons for this under-representation, such as fear of harassment from male colleagues or a sense of lack of social safety. For the purposes of the article, we focused on the ones which have appeared during the data analysis.

3 The project which led to the research whose findings are discussed in this article was conceived by Griffin who also secured the funding for Karakaş's postdoc at Uppsala University. Karakaş designed the research and conducted the ethnographic field study. She analysed the data from the fieldwork, the interviews, and online information on researchers' biographies and co-authorship practices. She drafted the article, and attended to its revision during the peer-review process.

4 As seen above, different sets are used to give an approximative idea about the number of women in the three universities of the Platform. This is related both to the ambiguity around group membership in the Platform and to the lack of availability of a list of affiliated researchers. Only University B had a list of the staff affiliated to the Platform, therefore this list was the most exact document to rely upon. Group membership in University C and A depended on the affiliation to the afore-mentioned departments and research groups of PIs funded by the Platform. University C submitted an annual report to the Platform, while University A did not. Here, we resorted to the list of researchers in the annual report of University C, and the number of researchers affiliated to the scientific computing division in University A. The number of women in the research groups of PIs funded by the platform therefore couldn't be counted in the latter case.

5 Interdisciplinarity as a concept has many meanings. For some it means working together across sub-areas of one academic field. However, the type of interdisciplinary that is of concern for us here involves broader interdisciplinary work across disciplines, e.g. between more gender-equal sciences, such as medicine, biological sciences and environmental sciences *and* the male-dominated engineering and IT.

6 See Suchman (2002) for the demystification of the designer/user opposition in technology production.

Back to the Present of Automated Mobility: A Typology of Everyday Use of Driving Assistance Systems

Gisle Solbu

Department of Interdisciplinary studies of culture, Norwegian University of Science and Technology, Norway/ gisle.solbu@ntnu.no

Tomas Moe Skjølsvold

Department of Interdisciplinary studies of culture, Norwegian University of Science and Technology, Norway

Marianne Ryghaug

Department of Interdisciplinary studies of culture, Norwegian University of Science and Technology, Norway

Abstract

This article focuses on how car drivers domesticate technologies of automation and the way this might inform our understanding of potential shifts to a more automated mobility system. The current literature on automated mobility has mainly addressed drivers' roles in terms of their attitudes towards—and acceptance of—an anticipated shift to high-level driving automation. In this article, however, we take a step back from expectations around automated mobility to explore the domestication of driving assistance technologies and systems already in use. The analysis is built on qualitative interviews with drivers of private cars in Norway. Based on our findings, we develop a typology of user-technology characterisations highlighting three themes of the drivers' use (comfort, safety, and novelty) as well as two modes of engagements (modulation and non-use). Our analysis suggests that automation is likely to be an incremental and gradual process and that its eventual application depends on the specificities of the practices that it seeks to disrupt. Moreover, we argue that the governance of automated mobility needs to be attentive to the dynamic and unpredictable roles technology will have in processes of socio-technical change. In this context, we highlight the key roles of users in shaping processes of appropriation of both new technologies and broader innovations and argue that knowledge about technology domestication provides important insights to changes towards automation in our current mobility systems.

Keywords: Automated Driving, Domestication Theory, Driving Assistance Systems, User Studies

Introduction

Expectations that automation and digitalisation will transform current mobility systems are high, both amongst policy makers and the transport industry (Bergman et al., 2017; Haugland and Skjølsvold, 2020; Ryghaug et al., 2022). In European policy, a language centred on ideas like smart mobility, digitalisation, connectivity and automation has become integral to the articulation of mobility futures (EC, 2020). In industry, most car manufacturing companies are pursuing projects related to technologies for automation, and Big Tech has joined in. Apple is pursuing the development of its own electric car aimed towards full self-driving capabilities (Gurman, 2021), while Google's self-driving car project (now called Waymo) has communicated an ambition to enable mobility without "anyone in the driver's seat"[1] (Waymo, n.d.) for more than a decade (Poczter and Jankovic, 2014). If realised, such a transition to automated mobility might be the most significant change to the mobility system since the introduction of the combustion engine (Hopkins and Schwanen, 2017), not least due to a series of proposed benefits such as increased access to mobility for older adults and children, improved road safety and sustainability.

The research community has also embraced these expectations for automated mobility, often through mapping the road towards increased automation by addressing technical and social barriers for innovation, adoption and use of automated mobility (see Milakis et al., 2017; Hermann et al., 2018). Within earlier studies of automated mobility, major streams of work also include research into the complexity of the technological infrastructure needed to enable automated driving (see Ryghaug et al. 2022; Marti et al. 2019; Liu et al., 2020; Lipson and Kurman, 2016) and issues related to the societal organisation of automated vehicles (see Milakis et al., 2020; Milakis et al., 2017; Mladenovic et al., 2020; Stilgoe, 2020; Stilgoe and Cohen, 2021; Cohen et al., 2018). A substantial body of literature has also developed in relation to the ethics of automated vehicles, mainly focusing on analysing issues concerning accidents with self-driving cars (see Dogan et al., 2020; Nyholm and Smids, 2016; Wolkenstein, 2018; Manchon et al, 2021). Within these literatures, the use of existing technologies has only been scarcely described.

In this article, we take a step back from the grand visions of automated vehicle futures which might arguably overshadow much-needed attention to the ways that ongoing processes of automation are already shaping mobility practices in important ways. Expectations for automated mobility futures are characterised by ideas of radical change and disruption. However, as innovations meet the challenges of practical implementation, they can lose their momentum. The successful introduction of new technology requires the alignment of a broad set of interests and actors (Pinch and Bijker, 1984; Bijker 1997; Bijker and d'Andrea, 2009). Thus, understanding the roles of users of a new technology in shaping processes of its appropriation (Ryghaug and Toftaker 2014; Anfinsen et al.2019; Berker et al., 2006) as well as broader innovation processes (Schot et al., 2016) is central to theorising the potential for changes towards automation in current automobility systems (Cohen et al., 2020). Our objective is therefore to give an account of how actors work to appropriate new technologies and the struggles and frictions that can appear when new technologies meet established routines and driving patterns.

When Tesla launched its new Model S featuring an improved autopilot function in 2015, it was described as "not a car, but a sophisticated computer on wheels," pointing to the external cameras, ultrasonic sensors and robust processing power available to assist the driver in controlling the vehicle (Hirsch, 2015). It seems as though cars are being gradually transformed to reduce the gap to self-driving futures. Indeed, the Insurance Institute for Highway Safety claims that "the building blocks for that technology [driverless cars] are already out on the road."[2] Regardless of whether this vison will come into fruition, new car models are increasingly being moulded and engineered to fit visions of future self-driving vehicles. Over the last decade, drivers have been exposed to new technologies such as advanced systems that might disrupt established driving patterns, intervene in decision-making processes and interfere with individual driving preferences. This means that today´s driving practices are already being influenced and changed by driver-related systems and technologies. Despite this, little is known about the ways driving practices are changed by the integration of such new technologies, which new practices emerge or what roles users play in making new technologies part of road-based mobility. The current transport-oriented literature largely seems disinterested in the links between contemporary technology and practice, gradual change and the automated mobility futures so often discussed.

In this article we ask how contemporary drivers domesticate (Silverstone and Haddon, 1996; Lie and Sørensen, 1996; Berker et al., 2006) technologies of automation, as well as how this might inform our analysis of potential shifts to a more automated mobility system. We do so by studying the mundane ways that automation has become part of driving over the past decade and by exploring ways that drivers make sense of, use or resist automation technologies that are already mainstream in contemporary car models. We focus on systems that automate specific functions in cars, generally referred to as Driving Assistance (DA) systems. DA systems are integrated into most modern cars to varying degrees of technological sophistication. They range from basic functions such as rear-view parking cameras and (adaptive) cruise controls that assist in maintaining a steady speed to advanced systems such as the Tesla autopilot which combines multiple sensor technologies to automate or assist driving tasks (e.g., speed adjustment, lane centring, road sign reading, parking) and actively intervene with the driver's control of the vehicle (Bengler et al., 2014). Systems with a high degree of complexity are frequently referred to as Advanced Driving Assistance Systems (ADAS), but for matters of simplification from here onward we use "DA systems" as an umbrella term capturing both the simpler systems and ADAS. Focusing on the configuration of available DA systems, we ask: how can we characterise the processes through which users appropriate such systems, and what can these characterisations teach us about the roles that users can play in a transition towards increased automation in driving?

Answering these questions provides an important corrective to mainstream narratives which suggest that a driverless society lies just around the corner. Our account instead suggests that automation is likely to be an incremental and gradual process, and that its eventual application will depend on the specificities of the practices it seeks to disrupt.

Theoretical perspective: Domestication

Our analysis is grounded in a socio-technical perspective (Sovacool et al., 2020), meaning that we are interested in exploring how relationships between technologies and their users are formed and how both the users' understanding of the technologies and their driving practices are shaped through appropriating new technologies. More specifically, our analysis mobilises domestication theory (Silverstone and Haddon, 1996; Lie and Sørensen, 1996; Berker et al., 2006), which enables a detailed understanding of the micro-practices of technology appropriation. Our analytical strategy thus stands in contrast to most social scientific research on automated mobility that does not focus on actual practices but is rather more futures-oriented and focuses on analysing systemic changes, addressing overarching transformations in mobility practices and anticipating issues related to their wider economic, technological, societal and ethical implications.

Domestication theory allows for an analysis of DA systems that zooms in on the use of technology. The approach focuses on how technological objects are transformed from something 'wild' into something 'tame' as users construct shared understandings of the technologies and how they are supposed to be used (Silverstone and Haddon, 1996; Sørensen, 2006: 46). Importantly, it evolves around an open-minded analytical process that is attentive towards the unexpected outcomes produced through technology appropriation. That is, the use and meaning of a technology are not taken for granted in the analysis but understood as something co-produced through interactions between users and technologies. The approach also renders both the technology and the social organisation surrounding it as malleable entities. Hence, new technology does not only create social change, but technologies themselves gain different meanings depending on their context of use and the specific practices they are appropriated into (see Ryghaug and Toftaker, 2014; Ryghaug et al., 2018; Aune, 2001; Næss, 2021).

One of the main aims of a domestication analysis is to capture the complex and extensive work that is done to stabilise the use of a new technology and to then unpack how new skills, practices and meanings are produced through this work (Sørensen, 2006). Domestication theory invites an analytical focus on three generic sets of features: 1) the formation of *new practices* in relation to the technology, such as the establishment of routines for using the technology or the development of institutions to support and regulate its use; 2) the development of new skillsets and *cognitive processes* related to taking part in the new practices; and 3) the construction of *meanings* of the technology, including the role the technology may have to the production of identities of the actors involved (Sørensen, 2006; Sørensen et al., 2000).

We approach our analysis of DA systems in a similar way. From the perspective of car manufacturing companies, DA systems have become a way to offer certain benefits such as comfort and safety to car buyers. Put differently, technology design contains a technology script (Akrich, 1992) that also includes the imagination of the user and expectations about the technology's intended use. Our interest lies in exploring what happens when these scripts meet actual users and actual uses. When the DA systems are domesticated, we should assume that the scripts are negotiated and even re-configured by users and that alterations of the intended use or entirely new and unexpected forms of use might emerge. Our focus is thus on addressing what we can theorise as the formation of new networks between technology, user and their environment that will occur when a technology is enacted in specific user contexts (Latour, 1996; Callon, 1986; Akrich, 1992). The inherent flexibility in the use-potential of technologies can be explained as the outcome of this dynamic process of network formation, which depends on already existing practices and is therefore difficult to predict or generalise (Sørensen, 1994; Pantzar, 1997).

By doing this we provide an alternative account of users compared to much of the literature on automated mobility. Previous studies' accounts of users are predominantly divided into simulator studies of driving conducted from a psychological perspective, such as using eye tracking to document attention shifts related to automated driving features (de Winter et al. 2014; Merat et al. 2014), and studies of public expectations, attitudes and acceptance of automated vehicles (König and Neumayr, 2017; Kyriakidis et al., 2015, Xing et al., 2021). In contrast, we ascribe a more decisive role to users as we focus on the active role they may take in appropriating new technologies and, relatedly, in shaping societal sustainability transitions (Ingeborgrud and Ryghaug, 2019; Ryghaug and Skjølsvold, 2019; Ryghaug et al., 2019; Sørensen, 2006;. Schot et al., 2016). Users can play important roles both as facilitators and critics of change, but this hinges on understanding what actually happens when technologies are being put to use and creating an openness to all the ways that DA systems become part of driving practices.

Methods and data

This study draws on empirical material consisting of 37 qualitative interviews with drivers using DA systems. By using qualitative interviews, we were able to collect in-depth information about users' experiences and perceptions of DA systems. The interviews were semi-structured, allowing for comparison across the material, but focused on open-ended questions that encouraged participants to share their personal experiences. While this choice of method has the limitation of not allowing direct observation of user-technology interactions, it has been valuable in understanding not only the practical aspects of using DA systems, but also the underlying assumptions, non-use practices, and sense-making processes that are crucial to the domestication of technology. Ultimately, our methodology enabled us to provide detailed insights into the complexities of using DA systems.

The interviews were conducted with people who owned cars in which one or more systems for DA were installed. The simplest cars, technologically speaking, had functions for cruise control in addition to antilock braking systems (ABS) and electronic stability programs. The most sophisticated cars had several new systems for "function specific automations" working in combination with each other, such as adaptive cruise control, lane assisting and lane centring technologies, an automatic braking system, a rear-view camera, parking sensors, parking assistance, and road sign reading. The interviewees' cars represented a broad variety of brands and a wide range in model years, with the oldest car having been produced in 2009 and the newest in 2020. The interviewees were selected to represent a diverse demographic profile, including a variety of professional backgrounds, years of education and ages (from 26 to 75 years). An important consideration was to recruit from urban and rural areas, as we wanted to explore variations in appropriating the DA systems across different geographical and infrastructural contexts. Table 1 gives an overview of the interviewees with information about age, sex and car used. For two of the interviewees age information has not been retrieved. For the cars, manufacturer and model names are listed and also model year when this information was available to the interviewees.

We conducted qualitative analysis of the material, following an approach inspired by grounded theory (Charmaz, 2006). The objective

Table 1. List of interviewees

Interviewee nr	Age, sex	Car model (year)
IW1	40s, Male	Toyota Rav4 (2011)
IW2	50s,Male	Tesla Model S P85D (2015)
IW3	30s, Female	Tesla Model X
IW4	50s, Male	Tesla Model 3 (2019)
IW5	30s,Female	Peugeot Rifter
IW6	60s, Male	Volvo XC60 (2017)
IW7	70s, Female	Renault Zoe
IW8	40s, Male	Tesla Model 3 (2019), Mercedes B electric
IW9	60s,Male	Audi A4 (2017)
IW10	60s,Male	Peugeot 208e (2020)
IW11	30s,Male	Toyota Avensis (2015)
IW12	20s,Male	Volkswagen eGolf (2020)
IW13	60s, Male	Audi E Tron (2020)
IW14	50s, Male	Peugeot 208 (2009)
IW15	60s, Female	Toyota Avensis (2018)
IW16	60s, Female	Kia Soule (2020)
IW17	60s, Male	Audi E-tron (2019)
IW18	30s, Male	Toyota Auris Touring (2016)
IW19	40s, Female	Tesla Model S (2016)
IW20	60s, Male	Tesla Model S (2014)
IW21	50s, Male and female	BMW Hybrid
IW22	Female	Peugeot 508 (2012)
IW23	Male	Nissan Leaf (2016)
IW24	50s, Male	Tesla Model 3 (2019)
IW25	40s, Male	Opel Ampera (2017)
IW26	20s, Male	BMW 318 (2016)
IW27	70s, Male	Mitsubishi Outlander (2017)
IW28	50s, Female	Jaguar F pace (2019)
IW29	20s, Male	Tesla Model 3 (2019)
IW30	60s, Female	Mitsubishi Outlander (2017)
IW31	40s, Male	Nissan Leaf (2018)
IW32	60s, Female	Nissan Qashqai (2015)
IW33	20s, Male	Jaguar I-Pace
IW34	40s, Male	VW e-Golf (2020)
IW35	50s, Male	Tesla Model X
IW36	50s, Male	Opel Ampera E (2018)
IW37	70s, Female	Mitsubishi Outlander

of our analysis was to develop an empirically grounded understanding of the users' interactions with the DA systems that could provide a basis for theorising processes of automated driving and the users' roles in these processes. The analysis was conducted in two main steps. The first step was geared towards developing a rich empirical characterisation of how the drivers were using the DA systems. We asked descriptive analytical questions: In relation to what aspects of their driving were DA systems used? How did users see the benefits and challenges of using DA systems? How was the use of DA systems shaping their driving? We also focused on developing insight into the relations that were formed between the users and the technologies. This was done by identifying dimensions in the interviewees' use of DA systems that cut across the identified main themes. The second step of the analysis was aimed towards generating an understanding of the domestication process as a whole in relation to DA and cars, focusing on the main features of technological domestication highlighted by Sørensen (2006).

A typology of DA and user interactions

In the following analysis, we have first developed a typology of DA systems and user interactions based on themes identified in the users' accounts (comfort, safety and novelty) and the modes of use identified across those themes (modulation and non-use). Second, we discuss our findings in terms of key features of the domestication process, asking: What new practices can we observe in the use of DA systems? What types of skillsets are developed to stabilise the use of the DA systems? How do drivers ascribe meaning to the technologies, and in what way do the technologies play a role in the formation of identities among the drivers?

DA for comfort

Unsurprisingly, comfort was one of the most frequent themes in the interviewees' accounts of using DA. When talking about DA in this context, the drivers made sense of the technologies in relation to mundane, everyday aspects of their driving. The technologies were "just there" as parts of their vehicle, and they used them whenever they felt the systems could offer increased comfort. For most of the drivers, DA systems had not been important in the decision of which car to buy; neither did they feel that the DA systems impacted how, when or why they used the car. Rather, their car use was generally presented as routinised, based on what they presented as stable driving practices and travel habits. The applicability of DA technologies was thus dependent on the technology being integrated into this existing landscape of everyday use. For example, adaptive cruise control and lane assistance were typically used to make long distance driving less tiring by delegating certain driving tasks from the driver to the car. One driver explained:

> […] it just makes driving a bit more comfortable. I have noticed that when I use the accelerator in a normal car [without DA], my foot actually starts to hurt, but if you use DA you will relax much more. You get less tense and you have a more comfortable driving experience. (IW2: 50s, male, Tesla Model S P85D 2015)

Another way DA was used to increase comfort was to assist in situations where the technologies alleviated the driver's stress, like using a rear-view camera to park the car accurately:

> I have become completely dependent on using the rear-view camera; you know using the sensors for backing the car, the rear-view camera, the alert system for parking and things like this when you are driving and are about to park. The combination of the alert system for parking, the sensors measuring the distance and the rear-view camera makes parallel parking so much easier. (IW6: 60s, male, Volvo XC60 2017)

As the quote illustrates, domesticating DA as comfort was dependent on the driver experiencing a benefit to their driving from the systems. Also, there needed to be a good match between the DA's formatting and the driver's existing driving patterns. To be used, the DA systems thus needed to have a low degree of disruption and build upon the ways that the drivers were already using the car. In sum, this means that the DA systems played a subordinate role in the driving practices and was rather experienced as technologies that provided incremental improvements to the comfort of driving. Importantly, this type of mundane use was prominent among users of older and simpler DA systems as well as advanced DA systems.

Comfort: Modulation and non-use

We are not only interested in exploring the common themes in the drivers' DA use but also in understanding the domestication process in light of the drivers' engagements with the technologies. That is, the drivers' accounts also show how this domestication drew on a continuous interaction between the user and the technology. Moreover, the drivers were constantly modulating the technology to make sense of it and adapt its use according to individual preferences, driving patterns and driving environments. Many pointed to obvious limitations in certain DA systems because of their dependency on clearly visible road-surface markings and stable driving patterns. This finding is also supported by previous research demonstrating the diverse attachments of autonomous vehicles, like reliance on infrastructure (see eg., Stilgoe, 2018; Tennant and Stilgoe, 2021; Ryghaug et al., 2022). Using DA was thus dependent on the driver's ability to interpret the driving environment and the driving patterns of other cars. For the general use of DA, this meant that the drivers also had their own individual preferences between when they saw the benefit of using DA and when they felt the technologies generated more annoyance than support. For example, some drivers actively chose to use adaptive cruise control for most of their driving, while others often avoided it, such as in situations where the road infrastructure was poor or where driving patterns could be unpredictable and create difficult situations. The threshold was quite low for when some drivers experienced the DA systems as generating more dis-comfort than comfort, showing they expected DA systems to be easily matched with their existing ways of driving. The following quote illustrates this form of "on and off" use:

> If you are driving on a country road and you reach a turn, then the car wants to slow down and I find it difficult to adjust the level of the system to something that feels comfortable. I would say if I am driving on a country road, and driving long distances, I would only use adaptive cruise control if there is traffic and I am stuck behind other cars. (IW17: 60s, male, Audi E-tron 2019)

The use of DA systems was also dependent on the driver's interest, willingness and ability to engage with the technologies. Some interviewees would present themselves as "too lazy" or not "interested" or "curious enough" to figure out how the technologies worked or how to operate them. As a result, they chose instead to dismiss them and to "only use a small portion of what the car has to offer" (IW28; 50s, female, Jaguar F pace), arguing that they were not willing to invest the time needed to figure things out when they could drive "just fine" without them. In these instances, the more advanced the DA systems were the more unapproachable they could appear. This means that comfort was not only relevant for the drivers' domestication of the technology but was also used as an argument for dismissing it.

This type of selective use—or rather, non-use—could also be the result of a driver experiencing a mismatch between their personal driving style and how DA systems enforced a certain structure upon their driving. One rural-based interviewee explained that she liked to drive slowly to enjoy the landscape and that DA, like cruise control, would not allow for such idiosyncrasies in her driving pattern:

> We have so much beautiful nature and then I slow down and I want to be able to look at the nature, so I drive I little bit like that…I cruise around for myself and enjoy the nature. My driving is not straight forward in the same speed, with a lot of traffic on a long highway. (IW15: 60s, female, Toyota Avensis 2018)

The drivers in these cases did not incorporate the DA systems into their driving because of an experienced friction between automation and individual preferences. Such instances of non-use point to diverse valuation practices among the drivers that were difficult to generalise and accommodate through automation. While the drivers experienced that automation algorithms prioritised efficiency, they gave accounts of how driving could fulfil roles beyond the practical task of transporting them from A to B and rather be a source of enjoyment (see e.g., Edensor, 2003). It is worth noting that interviewees ascribing to established gender identities could also be shaping the domestication process in this respect, like stereotypical claims about masculine fascination for technology. Our material provides support for this, e.g., one female interviewee stated that she did not use certain DA functions, but then added "but my husband always does" (IW3, 30s, female, Tesla Model X) . However, recent studies on the adoption of electric vehicles in Norway suggest that traditional gender roles in relation to cars are evolving (Anfinnsen et al., 2019). Furthermore, automation is expected to further influence the gendering of cars, underscoring the need for a nuanced investigation of the relationship between cars and gender (Weber and Kröger, 2018).

DA as safety

Safety was the second main theme identified from our interviews. Some drivers explained that the DA systems outperformed their own cognitive abilities, and so DA was understood as augmenting their driving to improve their safety. This was often attributed to more advanced DA systems that could automatically stop the car or slow down its speed when sensing an obstacle such as a pedestrian stepping onto the road or a car in front performing a rapid brake without warning. In these situations, many referred to the DA as an "added layer of safety" responding faster than themselves, without distraction. One interviewee explained enthusiastically how his car was "reading situations on the road better" than himself (IW24, 50s, male, Tesla Model 3 2019), while another told about a situation where the car had saved her from a potentially big accident:

> I was looking away from the road for a few second to adjust the radio, and then the car suddenly beeped very loudly! Someone had braked in front of me, without me noticing it. If it had not beeped I very well could have driven right into it […] It adds a feeling of safety to know that the car will tell me if something is in front of it (IW19; 40s, female, Tesla Model S 2016)

The use of DA systems was also experienced as enhancing safety by freeing drivers from tasks such as keeping the speed limit by checking the speedometer, adjusting the speed, and changing gears. This allowed drivers to pay full attention to the road. Some claimed that this made them substantially "less tired" from long distance driving when using complex DA systems like installed in the Tesla models (IW8, 40s, male, Tesla Model 3 2019). However, one interviewee also explained how a quite simple function like automated light adjustments allowed her to focus on road conditions which increased safety.

> I can concentrate on the road conditions. When it is dark outside during winter in Norway it is very tiering to keep switching on and off the lights while focusing on driving on narrow roads at the same time […] and if you control it manually perhaps you want to keep the full beam light on for longer than you should (IW21: 50s, female, BMW Hybrid)

For many, the question of delegating tasks to technology also sparked reflections on how using DA affected their own attention. This led to more critical remarks that the technologies decreased their attention towards the road, as DA systems enabled them to multitask (applying lipstick or unpacking lunch and eating) while still feeling safe. One driver warned about this practice:

> I do not become a better driver, I become a more passive driver. It is not like you increase the level of driving by using new technologies, you can instead become less aware of actually driving. (IW18: 30s, male, Toyota Auris Touring 2016)

As shown here, when speaking about safety, the interviewees often engaged in an explicit reflexive process where the drivers showed awareness of benefits of the DA systems to their driving and also potential down-sides of using the technologies. Interestingly, this points to how the users experience of safety was based on balancing the need for support and their own participation in driving when using more advanced DA systems.

Safety: Modulation and non-use

Just as the comfort provided by DA resulted from the drivers' modulating efforts, the feeling that DA improved safety tended to emerge as a result of driver-technology interaction over time. Responsibilities were little-by-little delegated to the DA systems, gradually building trust in the technology. For example, many interviewees described being anxious when using lane assist technologies or adaptive cruise control for the first time, as well as feeling a lack of control or that the car had a will of its own when it automatically adjusted itself on the road. Several also experienced episodes where the car either "phantom braked" seemingly without any visible obstacles in the road or steered them into potentially dangerous situations because of poor road marking. As a result of such experiences or just hearing others talk about them, many had a conscious relationship with the technologies in which they gradually learned when and how to use them and were alert to adjust the technology whenever needed. In these instances the interviewees talked about "thinking ahead of the car" (IW4, 50s male, Tesla Model 3 2019), or like one interviewee explained

> I never keep my foot far away from the pedals, and if I feel like it does something strange I adjust it myself […] We Tesla drivers are part of gigantic product development process where they track everything we do so the systems can learn from it (IW8: 40s, male, Tesla Model 3 2019)

Another example of gradually building trust concerned the use of a camera to assist in parking and reversing the car. Several interviewees explained how they found it difficult to completely trust the camera; some often felt the need to complement it by looking in the mirrors or turning their head, just to make sure they were clear of any obstacles. In this way, the use of the DA systems did not replace but rather added elements to existing practices and using the technologies was expressed as a processes of "learning to trust them" through gradual adaption (IW16, 60s, female, Kia Soule 2020) and also gradually incorporating new elements into the ways they had been trained to drive. These aspects of building trust, experience and overlapping practices highlight the temporal dimension of domestication and provide challenges for researchers who seek to understand the acceptance of automation because this, too, can be assumed to emerge over time with experience. Systems that had been a part of cars for many years, such as ABS, triggered less reflection; they were perceived as natural parts of the car and were relied on by the interviewees in their driving. Their own skills to complement these technologies, such as cadence breaking or correcting a skid, were perceived as gradually degrading.

Importantly, trust in their car's DA systems was not established for all the drivers, and we also observed non-use in relation to safety. Some chose to not use DA systems because they were not willing to "share" their control of the vehicle with the technology or they felt like the technology was not accurate enough to be of practical use. The perception of control has been presented as an important component in drivers' attitudes towards DA systems, often referred to as the "loss of control" argument in human-machine cooperation. However, it has also been difficult to confirm this argument empirically (see Weyer et al., 2015). While our data shows that the drivers in general felt at ease with using DA systems and that most issues were related to practical applicability instead of safety, the loss of control was still mentioned as a source of concern. On a related note, many felt that they did not understand how the advanced technologies made decisions, and therefore would not delegate control to them. This indicates relations between the level of DA advancement, technical understanding, practical experience and trust. One driver, who was normally using an older car, but had been trying out a newer model with advanced DA described this feeling of insecurity in using advanced DA systems as such:

> It was very special to let go of the car and leave everything to the car all the time. You do not know what lays behind the choices that the car makes, what if the car does something that you would not want to happen? (IW11: 30s, male, Toyota Avensis 2015))

Even though many of the drivers were actively using DA systems in ways that automated certain aspects of their driving, their perception of responsibility remained stable. Building practical experience with the technologies and becoming sensitised to their shortcomings seemed to produce a reflexivity among drivers concerning the relationship between themselves and the car. Most interviewees strongly believed that the responsibility was still the driver's, even if a malfunction in the DA system contributed to an accident:

> The responsibility always lays with the driver, there is no discussion about that in my opinion. You cannot blame the technology when you are driving a car. It is the driver who is responsible, these are only tools meant to assist you (IW7, female, 70s, Renault Zoe)

In this way, the driver´s understanding of responsibility is linked to a practical understanding of the technologies in use, the contexts in which it is used and how it affects the driving. Moreover, their understanding of responsibility in relation to new DA technologies is derived from existing practices of responsibility for car driving that are stabilised through traffic laws and other institutions, such as traffic schools. Importantly, we observed that both drivers of cars with advanced and simpler DA systems posed strong claims about the driver being responsible in case of accidents.

DA as novelty

The third main theme in the drivers' accounts of using DA systems was novelty. Through the drivers' discussions of novelty, we can see that their domestication of DA was linked to ideas of technological progress—that DA was something new, exciting and cutting edge—and this accordingly added value to their driving experience and their car. This means that the domestication of DA was not only linked to practical aspects, such as DA systems making driving easier or more comfortable, but also to sensations like excitement and enjoyment of driving (Næss et al., 2023).

In focusing on the aspect of novelty, the interviewees foregrounded the technologies, portraying them as futuristic elements that distinguished their current car from their previous cars. Novelty was also closely tied to a more general interest in technological development, and the drivers´ accounts of using the most advanced DA systems were sometimes accompanied by an expressions of fascination for technological progress:

> It is almost out of this world, you know. I hope every car gets safety functions like this, not necessarily the autopilot function, but, so many accidents could have been avoided if the machine had taken over. It is absolutely genius. I am not skeptical at all, I know a lot of people are skeptical, but I am not. I think that the machine, all things considered, always is smarter than you, who can be a bit unfocused, or maybe you are getting a text. (IW24: 50s, male, Tesla Model 3 2019)

Our data on the DA-user interactions thus show many similarities to previous literature's identification of important features among early technology adoption and the particularly prominent role of affective aspects of technology use in early phases of technology domestication (see Schot et al., 2016; Pantzar, 1997). Moreover, the phase-in of new technologies in Norway, such as energy-saving technologies in households, has drawn strongly on policy strategies focusing on the importance of such early adaption in stimulating market adaption.

Novelty: Modulation and non-use

As in the case of comfort and safety, domestication through novelty was based on an interaction between the user and the technologies. However, the interaction was described more explicitly as an active involvement with the DA systems and based on a feeling of participation. For example, several interviewees experienced their car doing unexpected things that potentially could bring them into a dangerous situation, but they showed a marked leniency towards these malfunctions, explaining that they knew they were dealing with immature technology that had to be developed further.

> I have to confess that something could have gone terribly wrong in the situation where the car got confused about the road markings in the overtaking line. […] But nothing happened: the wheels ended up outside the lane, but I corrected the car, and got a very big skid so the car almost drifted sideways, but then the car corrected itself. My heart beat a little bit faster, but that was a reminder; ´you cannot relax here´, this is beta testing, beta car…. the technology is beta and you cannot expect that everything works smoothly. (IW4, 50s, male, Tesla Model 3 2019)

Some drivers also saw themselves as having an explicit role in this development, taking on a role not unlike what Schot, Kanger and Verbong (2016) have described as 'user-innovators.' These drivers noted that their experiences could be crucial for improving the systems. Systems like those installed by Tesla that continuously collect data from the drivers were important in making this role of participation explicit, and thus also making the driver attentive and reflexive about how they used the technologies and their malfunctions.

The interaction between the users and the technology was also shaped by how the drivers imagined the intelligence of the DA systems; some related to the technology almost like a pet or a child, talking about "teaching" the car how to behave properly in certain situations. These interviewees referred to the process of using DA as a continuous process of learning, for themselves and for the car, or explained that dangerous situations could be "spooky" but also "part of the game" when using new technology:

> So unwanted and unexpected situations do happen, like when the car turns too much to the right and then panics when it detects it, and then all of a sudden make a sharp turn to the left again. Then I use my hand actively and help the car to understand and I sense that each time, before a new update resets the process, the car is learning. The car understands that it is supposed to keep to the left at that exact GPS point on the map. (IW4: 50s, male, Tesla Model 3 2019)

Interestingly, while a fascination for the new was an important theme in domesticating DA, the aspect of novelty was also presented as a reason for choosing not to use the technologies. In these cases, the drivers argued that they were not skilful enough, too old or not interested enough to introduce new elements into their already routinised ways of driving. Moreover, some argued that they preferred the "proper" fully manual way of driving, like they had been taught in traffic school. In discussing this aspect, interviewees highlighted the more emotional or tactile aspects of driving, such as explaining how they enjoyed shifting gears themselves and being fully in control of their vehicle or that driving could be something joyful and not simply a means of transportation. One interviewee described the feeling of using a car without DA as:

> It is that fantastic feeling for a man in my age, to be able to feel that I am back in a car that really is a car, and not a computer. There are feelings of joy connected to driving something where you sense that it is you who are making the decisions and not a computer that tells you how to drive at all times […] You drive the car; the car does not drive you. (IW10: 60s, male, Peugeot 208e 2020)

Such reflections illustrate the different meanings a car can have in people's lives and the social dimensions influencing car driving that can reach far beyond a narrow understanding of the car as a mean of transportation (se eg. Pearce, 2017; Jain and Lyons, 2008; Edensor, 2003). The introduction of new technologies to automate driving in this context seems to be understood as a source of detachment from driving and experienced as something cold and emotionless.

A characterisation of technology and driver interactions

Synthesising across the three themes discussed above, we generated a typology of technology-driver interactions. These interactions can be categorised according to their thematic focus (comfort, safety and novelty), with two dimensions (modulation and non-use) cutting across the three themes. This means that the domestication process is linked to certain thematic areas but also dependent on the driver's active involvement in modulating the technology. Moreover, the same thematic areas of domestication can be negated and presented as causes of non-use. The two dimensions in the typology thus point to the dynamic character of DA use, as the users show different modes of interacting with the DA systems within each of the themes. Table 2 presents an overview of this typology.

The typology demonstrates the diversity in DA use and illustrates that domestication of DA depends on more than the driver's acceptance of the technology; it depends on the formation of a complex set of relations between drivers, technology, road infrastructure and natural environments.

Table 2. A typology of technology-driver interactions

Themes	Modulation	Non-use
<u>Comfort:</u> Using DA to make driving more comfortable and less stressful in certain situations. Using DA to relieve the body from physical and mental stress. Match between automation and individual preferences.	Adapting DA to specific driving contexts / individual preferences. Modulating DA through active user involvement. Perceiving functionality as a product of DA and user interaction.	Dismissing DA because of annoyance, poor functionality or technology avoidance. Experiencing mismatch between DA formatting and personal driving preferences. Drivers not interested in investing time into the process of adapting the technology.
<u>Safety:</u> Using DA as an added layer of safety to existing driving practices. Experiencing DA as augmenting the driver's skillset. Drivers present a belief in DA's capabilities to outperform human cognition. Using DA to moderate personal driving patterns like staying within road speed limits.	Users negotiating relationships of trust with DA systems in terms of delegating driving tasks. Placing importance on human presence and intervention to back up technological malfunctions and serve as precautionary measures. Gradual processes of adaption with overlapping practices of DA use and pre-DA habits.	Experiencing a lack of autonomy while driving and limited insight into the car's decision-making process that is perceived as scary. Drivers not willing or interested in delegating control or engaging with the technology in order to adapt the use of DA to specific driving contexts.
<u>Novelty:</u> Using DA as something that adds excitement and enjoyment to car driving. A focus on the technologies in themselves, and linking their use to processes of exploration. Drivers demonstrate curiosity related to the functionality DA can provide.	Experimenting with the technology and testing its limits to map functionality. Perceiving malfunctions as opportunities for improvement. Drivers perceive themselves as active participants in the technological development.	Dismissing new technology based on nostalgic ideas of driving. Perceiving new technology as something that interferes with established driving practices. Displaying a focus on driving as a cultural practice and highlighting the enjoyment of non-automated processes. Perceiving DA as something too difficult to learn or not necessary given how they use the car.

Domestication of DA systems: Practices, skills and meaning

In this section, we discuss use-technology interactions by focusing on the three main features of domestication processes: emergent practices, skillsets and meaning-making (Sørensen, 2006).

Emerging practices

Building on the analysis presented above, we can argue that the use of DA did little to intervene with the interviewees' routinised ways of using their cars, in terms of when they used the car or what role the car played in their daily life. These routines were stabilised through the general organisation of their life, such as the distance they lived from their workplace, if they had a cabin outside of town or if they had kids that needed to be driven to different types of activities. To the extent that the use of DA affected these routines, it was only in minor ways, such as making daily car rides more comfortable or resulting in the car being the preferred travel mode for long distance travel due to increased comfort. Some also described a new practice in which they found themselves driving just for the sake of the excitement of testing new technologies.

Thus, rather than influencing practical aspects of using the car, the use of DA systems was closely tied to existing practices of driving; its appropriation was built upon these established practices. This meant that the changes introduced through DA were experienced as a continuation of established driving practices. The DA introduced new technological elements that incrementally modulated their driving behaviour, such as cruise control functions making their driving style less aggressive. Importantly, the user's active involvement to support the technologies was a key element in stabilising the use of DA. The use of DA can thus be described as a hybrid practice, partly consisting of the driver's own cognition and driving skillset and partly the technological automation enabled through DA systems.

Skillsets

As part of this hybrid practice, the skillsets needed to successfully appropriate the technologies largely entailed merging already-existing driving skills with the processes of automation provided by DA systems. The drivers rarely problematised the process of learning about the functionality of the technologies, instead presenting it as a straight-forward process based on an intuitive "learning-by-doing approach" that only demanded a certain level of willingness to engage with the technologies. The more important skillset for appropriating the technology was thus a meta-cognitive skill: an ability to evaluate the contexts in which DA systems could be used. This included making individual judgments about advantages and disadvantages based on driver preferences, learning what driving situations could be difficult for the DA systems to interpret and identifying situations where the technologies would need "help" from the driver. Through using the DA technologies and being exposed to a variety of situations, this meta-skillset is developed by drivers and becomes a key component in building a relationship of trust between the drivers and the DA systems.

Meaning-making

For the most enthusiastic users, the DA technologies were part of performing an identity as a progressive and technology-optimistic driver. This shaped the way the drivers ascribed meaning to the technologies, focusing on them as beta technologies and highlighting that some technological features were immature. This way of giving meaning to the technologies was important for how these drivers positioned themselves in relation to the technology and how the technologies were domesticated. To understand the technologies as "still in development" gave room for a permissive attitude towards the malfunctions of the technologies and also for understanding themselves as participants of what they presented as an ongoing "experiment" towards increased automation in driving. Importantly, such willingness to participate in the technology development process, to become what Schot, Kanger and Verbong (2016) and Kanger and Schot (2016) refer to as user-producers, has been shown to be a crucial resource in the dispersion of new technologies. Through their tinkering, user-producers play a key role in adapting technologies to specific user contexts. This type of behaviour also aligns with Panztar's (1997) description of early encounters with technology as often characterised by sensations of joy and happy experiments; the enthusiasm expressed by some of the drivers appears to be an important resource for domesticating new DA systems.

This fascination with the novelty of the technologies is not only important for establishing an emotional tie between the driver and the technology, but can also be seen as a resource for more general reflections on driving practices and human-technology interactions. For the interviewees, experiencing new DA systems controlling the car in certain situations sparked reflections on the limitations of their own cognitive abilities. This reflexive process was also prominent among the drivers who did not share this enthusiasm for the new technologies and had more cautious or anxious approaches to appropriating the technologies. Using the technologies and experiencing a lack of trust in the technologies' decision-making or friction between the automation's formatting and their individual driving preferences became a disruptive element in a routinised way of driving. For our interviewees, this generated reflections on the technological complexity of automating driving, the differences between human and technological decision-making and the role of infrastructure in driving.

Conclusion

Many European countries are now impatiently trying to set new trajectories for their mobility systems. In this context, visions of automation have become powerful attractors for policy makers in search of solutions to issues related to sustainability, safety, mobility access and local industrial development. However, transitions always entail grappling with trade-offs and unintended consequences (Kemp et al., 2007; Skjølsvold and Coenen, 2021). Hence, there is a need for careful governance approaches to address the wider societal implications of automation as well as how drivers and passengers will be affected by such developments (Hopkins and Schwanen, 2018). Our study represents an effort to take a step back from the promises of automation and direct our gaze on automation technologies already in use. By doing this we have, first, wanted to distance ourselves from hyped industry visions of automation's rapid upheaval of the transport system and rather work to understand the incremental steps being made towards automation and, second, highlighted how these incremental changes are gradually becoming part of life on the roads in different ways. Our study points towards a set of concrete findings concerning the technologies we have studied, but also towards a set of generic processes involved in socio-technical change. It seems unlikely that large-scale transport automation will be able to by-pass such processes in which domestication plays a key role.

By exploring the use of DA systems through the lens of domestication theory, we have shown how a driver's understanding of DA systems and, accordingly, their use patterns are shaped by the specific and complex contexts the technologies are adopted into. This is an important empirical insight that should be integral to how we understand and aim to initiate processes of change. This perspective also opens the way for a more fundamental argument about the unpredictability of implementing new technologies for driving automation. Domestication processes are shaped by the socio-technical arrangements that the technologies become part of, often causing unexpected outcomes. As policy makers and industry now seek to transform transportation through automation, this raises the central question of how one could try to predict potential unintended consequences.

In our study, the drivers' accounts of using DA systems show rather modest impacts on how the car was used in everyday life. This is an important observation in relation to the popular visions of driving automation that often point to possibilities to fundamentally disrupt current mobility practices. These visions stand in stark contrast to how our study shows DA systems are used today. Rather than disrupting user behaviours, DA systems are today aimed towards sustaining existing practices by making them more comfortable, less cognitively challenging, less stressful and safer.

Interestingly, our study also shows that the tolerance for friction between DA systems and the driver's established ways of driving is rather low. Too much interference or change often had the consequence of the technologies not being used. The domestication process was accordingly dependent on a good match with established ways of using the car. To further introduce automation into driving as a mean to reach sustainability goals, it is thus important to find ways that can facilitate more substantial disruptions in today's mobility practices. There is considerable political potential in mobilising DA systems to make a constructive contribution in facilitating more sustainable driving practice but this would require new design practices as well as engaging actively with existing environmentally-oriented elements of contemporary societies.

As an overarching observation, this study highlights the importance of facilitating experimentation when introducing new technologies, allowing users to build relations with the technologies. The trust some of the drivers developed towards the DA systems was not a given but rather the result of the drivers gaining experience through their use. Moreover, the use of the DA systems also appeared as a source of reflexivity among the drivers, both in relation to the capabilities and responsibilities of the technologies and to their own driving behaviours. Users can in this way be seen as important resources in a transition to new forms of mobility; analysis of the use itself provides crucial knowledge for policy makers that can unpack and create awareness towards the unpredictable ways that technologies for driving automation can become part of the ways we drive. In sum, the domestication perspective employed in this study leads to increased attention in how aspects related to experimentation and use over time can be key elements to explore in the transition towards automated mobility, not only to map fault lines and drivers for processes of change but also to actively enable new and productive relations between drivers, technologies and environments.

Acknowledgements

We want to thank Lina Ingeborgrud, Niri Kvammen Forberg and Bård Haugland Torvetjønn for participating in the DRIVERS research project and conducting interviews that provided the basis for this paper. The research has been funded by the Norwegian Research Council under the project name "DRIVERS: Digitalization of the road sector and its consequences: the role of driving". Grant number: 283354

References

Akrich M (1992) The De-scription of Technical Objects. In: Bijker W E and Law J (eds) *Shaping Technology, Building Society: Studies in Sociotechnical Change*. Cambridge, Massachusetts and London, England: The MIT Press, pp. 205-224.

Aune M (2001) Energy Technology and Everyday Life – the Domestication of Ebox in Norwegian Households. *Summer Study Proceedings for the European Council for an Energy Efficient Economy*, 5–16. Available at: https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2001/Panel_4/p4_1/ (accessed March 31, 2020).

Anfinsen M, Lagesen VA and Ryghaug M (2019) Green and gendered? Cultural perspectives on the road towards electric vehicles in Norway. *Transportation research part D: transport and environment* 71: 37-46.

Bengler K, Dietmayer K, Farbe B, Maurer M, Stiller C and Winner H (2014) Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent transportation systems magazine* 6(4): 6-22.

Bergman N, Schwanen T and Sovacool BK (2017) Imagined People, Behaviour and Future Mobility: Insights from Visions of Electric Vehicles and Car Clubs in the United Kingdom. *Transport Policy* 59: 165-173.

Berker T, Hartmann M, Punie Y and Ward K (2006) *Domestication of Media and Technology*. Berkshire: Open University Press.

Bijker WE (1997) *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, Massachusetts: MIT Press.

Bijker WE and d'Andrea L (2009) *Handbook on the Socialisation of Scientific and Technological Research. A tool for promoting science and technology socialisation policies addressed to policy makers, research and innovation actors and stakeholders.* Brussels: EU.

Callon M (1986) The Sociology of an Actor-Network: The Case of the Electric Vehicle. In: Callon M, Law J and Rip A (eds) *Mapping the Dynamics of Science and Technology*. London: The Macmillan Press, pp. 19-34.

Charmaz K (2006) *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London: Sage.

Cohen T, Stilgoe J and Cavoli C (2018) Reframing the Governance of Automotive Automation: Insights from UK Stakeholder Workshops. *Journal of Responsible Innovation* 5(3): 257-279.

Cohen T, Stilgoe J, Stares S, et al. (2020) A Constructive Role for Social Science in the Development of Automated Vehicles. *Transportation Research Interdisciplinary Perspectives* 6:100133. https://doi.org/10.1016/j.trip.2020.100133.

de Winter JC, Happee R, Martens MH and Stanton NA (2014) Effects of Adaptive Cruise Control and Highly Automated Driving on Workload and Situation Awareness: A Review of the Empirical Evidence. *Transportation Research Part F: Traffic Psychology and Behaviour* 27: 196–217.

Dogan E, Costantini F and Le Boennec R (2020) Ethical Issues Concerning Automated Vehicles and Their Implications for Transport. In: Milakis D, Thomopoulos N and van Wee B (eds) *Policy Implications of Autonomous Vehicles.* Cambridge, Massachusetts: Elsevier, pp. 215-233.

Edensor T (2003) Defamiliarizing the mundane roadscape. *Space and culture* 6(2): 151-168.

European Commission (EC) (2020) *Sustainable and Smart Mobility Strategy—Putting European Transport on Track for the Future.* Brussels: European Commission.

Gurman M (2021) Apple Accelerates Work on Car Project, Aiming for Fully Autonomous Vehicle. *Bloomberg, US Edition*, November 18. Available at: https://www.bloomberg.com/news/articles/2021-11-18/apple-accelerates-work-on-car-aims-for-fully-autonomous-vehicle (accessed June 23, 2022).

Haugland BT and Skjølsvold TM (2020) Promise of the Obsolete: Expectations for and Experiments with Self-Driving Vehicles in Norway. *Sustainability: Science, Practice and Policy* 16(1): 37-47.

Herrmann A, Brenner W and Stadler R (2018) *Autonomous Driving: How the Driverless Revolution Will Change the World*. Bingley, UK: Emerald Publishing.

Hirsch J (2015) Elon Musk: Model S Is Not a Car But a Sophisticated Computer on Wheels. *Los Angeles Times,* March 19, 2015. Available at: https://www.latimes.com/business/autos/la-fi-hy-musk-computer-on-wheels-20150319-story.html (accessed July 5, 2023).

Hopkins D and Schwanen T (2017) Governing the Race to Automation. In: Marsden G and Reardon L (eds) *Governance of the Smart Mobility Transition*. Bingley, UK: Emerald Publishing, pp. 65-84.

Hopkins D and Schwanen T (2018) Automated Mobility Transitions: Governing Processes in the UK. *Sustainability* 10(4): 956. https://doi.org/10.3390/su10040956.

Ingeborgrud L and Ryghaug M (2019) The Role of Practical, Cognitive and Symbolic Factors in the Successful Implementation of Battery Electric Vehicles in Norway. *Transportation Research Part A: Policy and Practice* 13: 507–516.

Jain J and Lyons G (2008) The gift of travel time. *Journal of transport geography* 16(2): 81-89.

Kanger L and Schot J (2016) User-Made Immobilities: A Transitions Perspective. *Mobilities* 11(4): 598-613.

Kemp R, Rotmans J, and Loorbach D (2007) Assessing the Dutch energy transition policy: how does it deal with dilemmas of managing transitions? *Journal of Environmental Policy & Planning* 9(3-4): 315-331.

König M and Neumayr L (2017) Users' Resistance Towards Radical Innovations: The Case of the Self-Driving Car. *Transportation Research Part F: Traffic Psychology and Behaviour* 44: 42-52.

Kyriakidis M, Happee R and de Winter JC (2015) Public Opinion on Automated Driving: Results of an International Questionnaire Among 5000 Respondents. *Transportation Research Part F: Traffic Psychology and Behaviour* 32: 127-140.

Latour B (1996) On Actor-Network Theory: A Few Clarifications. *Soziale Welt* 47(4): 369-381.

Lie M and Sørensen KH (eds) (1996) *Making technology our own?: domesticating technology into everyday life*. Oslo: Scandinavian University Press.

Lipson H and Kurman M (2016) *Driverless: Intelligent Cars and the Road Ahead*. Cambridge, Massachusetts: MIT Press.

Liu R, Wang J and Zhang B (2020) High Definition Map for Automated Driving: Overview and Analysis. *The Journal of Navigation* 73(2): 324-341.

Manchon JB, Bueno M and Navarro J (2021) From Manual to Automated Driving: How Does Trust Evolve? *Theoretical Issues in Ergonomics Science* 22(5): 528—554.

Marti E, de Miguel MA, Garcia F and Perez J (2019) A Review of Sensor Technologies for Perception in Automated Driving. *IEEE Intelligent Transportation Systems Magazine* 11(4): 94-108.

Merat N, Jamson AH, Lai FC, Daly M and Carsten OM (2014) Transition to Manual: Driver Behaviour When Resuming Control from a Highly Automated Vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour* 27: 274-282.

Milakis D, Thomopoulos N and van Wee B (eds) (2020) *Policy Implications of Autonomous Vehicles*. Advances in Transport Policy and Planning, vol. 5. Cambridge, Massachusetts: Academic Press.

Milaki D, van Arem B and van Wee B (2017) Policy and Society Related Implications of Automated Driving: A Review of Literature and Directions for Future Research. *Journal of Intelligent Transportation Systems* 21(4): 324-348.

Mladenović, M, Stead D, Milakis D, Pangbourne K and Givoni M (2020) Sociotechnical Imaginaries of Connected and Automated Vehicle Technology: Comparative Analysis of Governance Cultures in Finland, Germany, and the UK. Bridging Transportation Researchers (BTR), 10-11 Aug 2020, Online Conference.

Næss R (2021) El-sykkel som utvidelsesteknologi: Nødvendigheten av å studere et brukerperspektiv i en grønn omstillingsprosess. *Norsk Sosiologisk Tidsskrift* 5(4): 1-16.

Næss R, Heidenreich S and Solbu G (2023) Sensory and emotional dimensions of domesticating new technology: an experiment with new e-bike users in Norway. *Mobilities*. Epub ahead of print 10 July 2023.

Nyholm S and Smids J (2016) The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem? *Ethical Theory and Moral Practice* 19(5): 1275-1289.

Pantzar M (1997) Domestication of Everyday Life Technology: Dynamic Views on the Social Histories of Artifacts. *Design Issues* 13(3): 52-65.

Pearce L (2017) 'Driving-as-Event': re-thinking the car journey. *Mobilities* (12)4: 585-597. DOI: 10.1080/17450101.2017.1331007

Pinch TJ and Bijker WE (1984) The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science* 14(3): 399-441.

Poczter SL and Jankovic LM (2014) The Google Car: Driving Toward a Better Future? *Journal of Business Case Studies* 10(1): 7-14.

Ryghaug M, Haugland BT, Søraa RA and Skjølsvold TM (2022) Testing Emergent Technologies in the Arctic: How Attention to Place Contributes to Visions of Autonomous Vehicles. *Science & Technology Studies* 35(4): 4-21. doi: 10.23987/sts.101778

Ryghaug M, Ornetzeder M, Skjølsvold TM and Throndsen W (2019) The Role of Experiments and Demonstration Projects in Efforts of Upscaling: An Analysis of Two Projects Attempting to Reconfigure Production and Consumption in Energy and Mobility. *Sustainability* 11(20): 5771. https://doi.org/10.3390/su11205771.

Ryghaug M and Skjølsvold TM(2019) Nurturing a Regime Shift Toward Electro-Mobility in Norway. In: Finger M and Audouin M (eds) *The Governance of Smart Transportation Systems.* Cham, Switzerland: Springer, pp. 147-165.

Ryghaug M, Skjølsvold TM and Heidenreich S (2018) Creating Energy Citizenship Through Material Participation. *Social Studies of Science* 48(2): 283-303.

Ryghaug M and Toftaker M (2014) A Transformative Practice? Meaning, Competence, and Material Aspects of Driving Electric Cars in Norway. *Nature and Culture* 9(2): 146-163.

Schot J, Kanger L and Verbong G (2016) The Roles of Users in Shaping Transitions to New Energy Systems. *Nature Energy* 1(16054). https://doi.org/10.1038/nenergy.2016.54.

Silverstone R and Haddon L (1996) Design and the Domestication of Information and Communication Technologies: Technical Change and Everyday Life. In: Mansell R and Silverstone R (eds) *Communication by Design: The Politics of Information and Communication Technologies*. Oxford, UK: Oxford University Press, pp. 44-74 .

Skjølsvold T M and Coenen L (2021) Are rapid and inclusive energy and climate transitions oxymorons? Towards principles of responsible acceleration. *Energy Research & Social Science* 79: 102164.

Sørensen KH (1994) Technology in Use. Two Essays on the Domestication of Artefacts, STS Working Paper, No. 2/94. Trondheim, Norway: NTNU, Centre for Technology and Society.

Sørensen KH (1996) Learning Technology, Constructing Culture. Socio-Technical Change As Social Learning, STS Working Paper, No. 18/96. Trondheim, Norway: NTNU, Centre for Technology and Society.

Sørensen KH (2006) Domestication: The Enactment of Technology. In: Berker T, Hartmann M, Punie Y and Ward K (eds) *Domestication of Media and Technology*. Berkshire, England: Open University Press, pp. 40-61.

Sørensen KH, Aune M and Hatling M (2000) Against Linearity: On the Cultural Appropriation of Science and Technology. In: Dierkes M and von Grote C (eds) *Between Understanding and Trust: The Public, Science and Technology*. Amsterdam: Harwood Academic Publishers, pp. 237-257.

Sovacool BK, Hess DJ, Amir S, et al. (2020) Sociotechnical Agendas: Reviewing Future Directions for Energy and Climate Research. *Energy Research & Social Science* 70: 101617. https://doi.org/10.1016/j.erss.2020.101617.

Stilgoe J (2018) Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science* 48(1): 25-56. https://doi.org/10.1177/0306312717741687

Stilgoe J and Cohen T (2021) Rejecting Acceptance: Learning from Public Dialogue on Self-Driving Vehicles. *Science and Public Policy* 48(6): 849-859.

Stilgoe J (2020) Who's driving innovation. *New Technologies and the Collaborative State*. Cham, Switzerland: Palgrave Macmillan.

Tennant C and Stilgoe J (2021) The attachments of 'autonomous' vehicles. *Social Studies of Science* 51(6): 846-870.

Weber J and Kröger F (2018) Introduction. *Transfers: Interdisciplinary Journal of Mobility Studies* 8(1): 15-23. 10.3167/trans.2018.080103

Waymo (n.d.) About Waymo, Waymo homepage. Available at: https://waymo.com/about/ (accessed May 4 2022).

Weyer, J, Fink RD and Adelt F (2015) Human–Machine Cooperation in Smart Cars: An Empirical Investigation of the Loss-of-Control Thesis. *Safety Science* 72: 199-208.

Wolkenstein A (2018) What Has the Trolley Dilemma Ever Done for Us (And What Will It Do in the Future)? On Some Recent Debates About the Ethics of Self-Driving Cars. *Ethics and Information Technology* 20(3): 163-173.

Xing Y, Lv C, Cao D and Hang P (2021) Toward Human-Vehicle Collaboration: Review and Perspectives on Human-Centered Collaborative Automated Driving. *Transportation Research Part C: Emerging Technologies* 128: 103199.

Notes

1 See the Waymo website, https://waymo.com, for information on the Google self-driving project.

2 See communication from Russ Rader, Senior vice president at the Insurance Institute for Highway Safety retrieved from https://abcnews.go.com/blogs/business/2013/09/the-next-step-to-driverless-cars

Ethical Plateaus in Danish Child Protection Services: The Rise and Demise of Algorithmic Models

Helene Friis Ratner

Danish School of education, Aarhus University, Denmark/ helr@edu.au.dk

Ida Schrøder

Aarhus University, Denmark

Abstract

This paper analyses how controversies shape an emerging field of AI in Danish child protection services. In a context of high controversiality, we examine how algorithmic systems evolve in conjunction with changing ethical stakes. Empirically, we report a study comprising all Danish attempts (*n=4*) to develop algorithmic models for child protection services. These attempts were never fully implemented and have been either cancelled, paused or changed significantly since their outset. Combining the notion of 'ethical plateaus' with insights from valuation studies, we propose that public controversies shape how organisations enact their algorithms as ethically 'good'. Our findings demonstrate how valuations of ethically contestable algorithms involve the very distribution of agency across humans and algorithms, i.e., how much power and agency should be delegated to algorithmic models. In the case of Danish child protection services, this moves towards reducing their agency.

Keywords: Artificial intelligence, Ethics, Controversy, Child Protection Services

Introduction

Heralded with promises of increased scope, speed, and precision, algorithmic systems based on machine learning are making their way into child protection services (Andrejevic, 2017; Meilvang and Dahler, 2022; Redden et al., 2020). At the same time, algorithmic technologies have spurred an avalanche of ethical concerns regarding the powers of algorithms (Beer, 2009; Mittelstadt et al., 2016). Such concerns do not only exist in academic literature but are increasingly put on public display through medialised controversies (Kristensen, 2022). However, as Noortje Marres (2021) argues, public controversies are not simply democratic contestations of emerging technologies but increasingly figure as trial grounds for innovation and product development. Addressing such an interplay between medialised controversy and the development or closure of algorithmic models, we are interested in its implications for changing ethical boundaries of AI[1] in child protection services.

Internationally, policy makers and managers have begun to investigate predictive algorithms as tools to support social workers' assessments and decision-making. Yet, the critiques of the ethicality of algorithmic models in child protection services abound. Rather than help children, the critiques go, algorithmic models instead result in biased and non-transparent decisions, increased inequality, and a de-humanised administration based on datafied surveillance of the poor (Dencik et al., 2019; Eubanks, 2018; Jørgensen, 2021). The high level of controversy in the field of child protection renders the question of ethics especially prevalent and makes it a good setting for exploring empirically how the ethics and value of algorithms are developed and change in conjunction with public value contestation and criticism.

We propose to analyse this emerging and changing field of 'ethical AI' in child protection through the concept of 'ethical plateaus' (Fortun and Fortun, 2005; Seaver, 2021). This concept highlights the instability of an ethical terrain with moving boundaries of what is ethically possible. To examine how organisations negotiate the value and ethics of their algorithmic model we further draw on insights from valuation studies (Helgesson and Muniesa, 2013). Rather than assuming 'value' to be an inherent property of an object, valuation studies emphasise value as *enacted*, i.e., as a practical accomplishment and attribution. Together, these concepts offer useful sensibilities for the analysis of how the ethical stakes of publicly contested algorithmic systems change over time and how this impedes the very valuation of algorithmic models. Thus, in examining how public controversy, ethical stakes and valuations of algorithmic models shape one another in 'changing ethical plateaus', we ask how algorithmic models are enacted as valuable, how they are rendered controversial and why they, eventually, were cancelled or changed.

Empirically, we report a qualitative study of four attempts to develop algorithmic models for public sector child protection agencies in Denmark. The four algorithmic models comprise all Danish attempts to design, develop, test, and implement algorithmic models in this area of work and they took place during a six-year period (2017-2022). Although predictive algorithms used to profile children and families at risk are gaining ground internationally, especially in anglo-saxon countries (Cuccaro-Alamin et al., 2017; Redden, 2018; Russell, 2015), these systems are deemed highly controversial in Denmark, and none of the four algorithmic models have been implemented. Indeed, the examples we examine in this paper never made it beyond the status of project designs, limited trials, or research. This is curious for a country, which is otherwise ranked as the number one country in digital government across the world (United Nations Department of Social and Economic Affairs, 2022). In this sense, the Danish case provides a rather unique entry point into understanding how ethical plateaus emerge and undergo change in conjunction with organisational attempts at developing and rendering algorithmic models in child welfare valuable and legitimate. Furthermore, studying algorithms that have been cancelled or changed significantly provides a fruitful entry point into valuation practices as the actors developing the algorithms, first, have been called to explicate how and why their (envisioned) algorithm is ethically good, and second, justify or explain why they had to close or change it.

The paper is structured as follows. Taking our inspiration from anthropology and Science and Technology Studies (STS), we provide a brief review of some important discussions here. This is followed by a presentation of our conceptual approach, the combination of 'ethical plateaus' and 'valuation', as well as a methods section, introducing the context of child protection services and accounting for the empirical data assembled and analysed for this paper. In the subsequent analytical sections, we present our analysis of the four Danish cases, identifying how and why the different organisations embarked on using AI for child protection, how valuations changed over time, sometimes in response to public controversies, as well the events that led to cancelling, pausing or changing the algorithmic projects. In a concluding discussion, we develop a timeline across the four cases, visualising how ethical plateaus change over time and how public controversies and valuations intersect. Over time, the algorithmic models are granted less agency and power. This indicates that ethical plateaus are closely entwined with negotiations of how agency and power should be distributed across humans and machines, figuring the politics of what it might mean to live with AI and what we mean by ethical AI.

Ethics as socio-material and processual accomplishments

This paper builds on the work of scholars associated with STS and anthropology who are beginning to examine ethics as a processual, relational and practical accomplishment. Puig de la Bellacasa, for instance, conceptualises ethics as "concrete [socio-material] relationalities in the making" rather than "normative morals" (Puig de la Bellacasa, 2010: 152). From this follows that we should not presume knowing in advance what ethical AI should look like, but instead, examine its ongoing, uncertain and situated making. Many scholars in STS have followed this call. In his study of music recommender algorithms, Seaver (2021), for instance, explores how the makers of recommender systems engage with popular critiques of algorithms as powerful, computerised agencies replacing 'careful human judgment' of music. Seaver (2021: 512) analyses developers' reasoning of these critiques as the business of actively "making ethics, trying to understand, evaluate, and reconfigure the field of possible choices". Ethics, in this view, are enacted in the developers' framing of how competing values can co-exist.

Another example is Douglas-Jones (2017) who has conducted an ethnographic study of the ethics review committees for biomedical research in the Asia-Pacific region. Exploring ethics as a material practice, she examines how universal ethical standards are negotiated in encounters with situated and socio-material circumstances such as office spaces. This focus allows her to study the mundane work of building infrastructures for ethical review and universal standards, concluding that universal ethics emerge "as a site of ongoing attention and negotiation, standard making and aspiration" (Douglas-Jones, 2017: 28). Finally, Ziewitz (2019) has studied ethics as a practical accomplishment in a context of SEO consultants work with search machine optimisation, aiming to understand 'how people organise themselves as ethical in the absence of the ontological security that professional ethicists and some philosophers presume'. Like Puig de la Bellacasa (2010), Ziewitz emphasises the need to leave behind the inclination to decide whether a practice is ethically correct and instead accomplish a deeper understanding of how "'being ethical' [is] (…) implicated in and organize the (…) experiences of people" (Ziewitz, 2019: 713). This approach, thus, allows a close look into the practical work of establishing, negotiating and distributing ethics in a context of its increasing contestation.

In our analysis of how the valuation of algorithms change in conjunction with changing ethical stakes, we build on these sensitivities. Examining the interplay between medialised controversies and algorithmic projects in Danish child protection services, we contribute with an inclusion of public controversy as an important factor in the development of what counts as ethical AI in child protection.

Ethical plateaus and valuation studies

The notion of 'ethical plateaus' helps us conceptualise changing ethical boundaries of what is possible in techno-scientific situations fraught with dilemmas and ethical contestations. Defining ethical plateaus as a site "where multiple technologies interact to create a complex terrain or topology of perception and decision making" (Fischer in Fortun and Fortun, 2005: 47), the concept allows one to examine the intersection and co-evolvement of different ethical concerns. For our purpose, we do this across four Danish algorithmic models in child protection services, attending to how the developers attempt to manage the horizons of possible ethical issues posed by controversial algorithms and how the different projects relate to one another, e.g., through practices of 'un-ethicizing' (Tønnesen, 2009).

The concept's geological metaphor brings about the image of a complex socio-technical landscape made up by interactions between, in our case, algorithmic models, administrative apparatuses and public media controversies. Like geological plateaus, formed through processes such as volcanic activity, tectonic uplift, or erosion, ethical plateaus constitute a dynamic terrain of changing and competing concerns and values.[2] In this context, we might see public controversies as volcanic 'ruptures' shaping the formation of plateaus, insofar as they make organisations re-value and modify their algorithmic models significantly. The ethical plateaus, then, "foregrounds the tectonics of the ethical sphere—everything that supports and constrains the range of ethical possibility, without making a strong distinction between the 'hard' constraints typically associated with technology and the 'soft' ones associated with society" (Seaver, 2021: 513). Following from this, we are not interested in assessing the ethicality of the algorithmic models we study by benchmarking them against established ethical principles. Instead, we explore how valuations of the ethicality of algorithms change over time in conjunction with public controversy, understanding this process as changing ethical plateaus and their ongoing figurations of what is ethically acceptable to do with AI in Danish child protection services.

As part of this, we draw on the insights from Valuation Studies, allowing us to depart from the idea that the value of an algorithm is an inherent attribute or quality of the algorithm itself. Instead, this approach emphasizes value to be the result of a situated and practical endeavour to explicate what the algorithm is good for (Helgesson and Muniesa, 2013). I.e., if we are to learn what is valuable about an algorithm, we must look for the situated valuations of algorithms. In this view, value conflicts do not occur between pre-established ideas of what is good and valuable in a society, but rather as practices of negotiating, adjusting, and reconceptualising the algorithms. Analysing changing valuations of what counts as the ethically good algorithm in conjunction with public controversies allow us to draw the contours of emerging ethical plateaus of AI in Danish child protection services.

Empirical resources

Our study comprises of all (four) Danish attempts to develop algorithmic models for child protection: Three municipal development projects and one research project named RISK ("Underretninger i fokus"). RISK and one of the municipalities, Gladsaxe, were subject to public controversy. As Gladsaxe and RISK have been under much public scrutiny, pseudonymization is impossible here. The other two municipalities are pseudonymised.

For all cases, we conducted interviews with key stakeholders (e.g., project leaders, data scientists, municipal directors) amounting to 39 interviews. We collected 63 public- and non-public documentations (e.g., project descriptions, minutes, legal assessments, power point presentations) and traced public media-debates concerning the roles of algorithms in Danish public administration (source: Infomedia), including 45 media articles. For all interviews, we had our interviewees make a timeline and identify important turning points. For those projects subject to public media controversy, we also asked our interlocutors to reflect on selected critical media quotes.

In analysing the empirical material, we developed timelines for all cases, identifying 1) the initial justifications for developing the algorithmic models and initial ethical considerations, 2) public and non-public contestations and critiques of their ethicality as well as project members' responses to these, and 3) why they were eventually closed or changed considerably. For these moments, we analysed valuations. i.e., their enactment of the (ethical) 'goodness' of the algorithmic models. We further made note when project agents explicitly related their own algorithmic model to the other cases, to examine how they made sense of their own algorithm in comparison to the other Danish examples. As our aim is to identify emerging ethical plateaus, we have chosen the valuations that we deemed to be most dominant and influential in the development (and termination) of the algorithms. Our analysis thus does not reflect all valuations detected in our mapping and, thus, we make no claims of completeness.

The context of Danish child protection services

Child protection services are tasked with the difficult but crucial task of preventing and stopping child maltreatment. Identifying children at risk of maltreatment, however, is a complex task, fraught with uncertainty and severe consequences for the families involved if the wrong assessments are made (Villumsen and Søbjerg, 2020). Even in Denmark, where the universalistic welfare model entices professional collaborations across core welfare services such as daycare, healthcare, and school, and where the Social Service Act demands early interventions with family-oriented services, four percent of children are at some point during their childhood placed in out of home institutions or foster families due to neglect, maltreatment or other situations which impedes on the child's wellbeing and abilities to develop alongside with his or her peers (VIVE, 2022).

How to break these statistics has been a pivotal concern in the past 20 years of reformation of the Danish child welfare system. Most importantly the welfare professionals' obligation, as well as private citizens' access, to notify the authorities about children they are concerned about was immensely expanded from 2011 and onwards. Correspondingly, the number of notifications about children failing to thrive has increased ever since – from 97.288 notifications in 2015 to 138.099 notifications in 2021 (Statistics Denmark, 2022). Combined with municipalities' legal demand to assess the severity of notifications within 24 hours upon receival, the system of notifications has become pivotal in the organising and innovation of Danish Child protection agencies. As we will see in the following, the four algorithmic projects entail rather different 'problematisations' (cf. Callon, 1986) of this situation, two (the municipalities) targeting the massive amounts of notifications, one (RISK) targeting the assessment of these notifications and, finally, one (Gladsaxe) trying to pre-empt notifications through early intervention (cf. Ratner and Elmholdt, 2023).

The Gladsaxe model: Algorithmic prediction to pre-empt unwanted futures

The first attempt to develop a predictive algorithm in the field of child protection was undertaking by Gladsaxe municipality in 2017-2018. In 2017, the civil servants and politicians of Gladsaxe municipality formulated the idea of developing the algorithmic tool for "data-driven early detection". This model was to solve a problem of being notified of children's problems too late. In Denmark, child protection services learn about children's maltreatment through notifications sent to them, i.e., concerns about children's wellbeing. As the then leader of the child protection services explained, it was not simply a problem of welfare professionals failing to notify them but a problem of the municipality not linking up data already held by different welfare departments:

> I mean, how do we reach these families when their children are infants? (…) At some point, our leader of the employment services remarks that they are the first to be advised when a long-term unemployed mother is pregnant. (…) But the problem is that this information stays with the employment services because there is no 'forward information-button', you know? We are not notified in the child protection services so we can't begin working with these mothers. (…) As it is now, we are simply waiting for the children's symptoms – that something is wrong in the family – to emerge, instead of acting on our knowledge of the risks being present in a family. (Interview, June 2021, leader of Child protection services)

Illustrating how long-term unemployment is considered a 'risk', the leader reflects on how such data on risk resides with other welfare departments but rarely reaches the child protection officers. This realization made them think about how to bypass notifications being sent 'too late'. An algorithmic merging of data from different welfare departments, the idea was, could serve as an alternative mode of detecting children *before* symptoms would emerge. The algorithm was envisioned to merge data on known risks such as parents' employment status and history, substance abuse, absence from appointments with the dentist or health nurse, earlier notifications to child protection services. The idea was to use the algorithm for detection of at-risk families, after which a case worker would make contact and offer help on a voluntary basis. This valuation of the goodness of the algorithm relies on a distinction between risk (here attributed to e.g. parents with long-term unemployment) and symptoms. Rather than acting on 'symptoms' on children's maltreatment, of which they are notified in notifications, the algorithm is valued for its capacity to detect risks and thereby pre-empt children's possible maltreatment through anticipation.

In valuing the algorithm as ethically good, the developers further emphasised that only the computer would access citizens' data. A municipal leader said it like this in an interview:

> I mean, all these data would be in a black box which neither I nor other employees could access. (…) We don't need to see all these [the municipality's corpus of] citizens' data. We are not

> interested in this. They [the citizens] need to live their lives out there. It is just the small share [of children] which we believe we can pull out (…). And these are the ones we want the algorithm to find. (Interview, June 2021, leader of Child protection services)

With this description of the algorithm, they established a distinction between the computational analysis of all citizens' data and the caseworkers' access to the select citizens profiled by the algorithm. They visualised the ethicality of the algorithm as a 'statistical black hole', emulating the algorithm's analysis of citizens' data as non-visible to employees (figure 1), demarcating this as more ethical than case workers looking through all this data.

This valuation invokes the employees' non-access to citizens' data as an ethically valuable property of the algorithmic model. It also enacts surveillance as a human endeavor, i.e., algorithmic profiling would not count as surveillance unless a human caseworker is being informed of the identity of the profiled citizen.

For the algorithm to be tested and implemented, the municipality applied to the Ministry of Interior to be exempted from privacy regulation requiring citizens' consent to merge data. To the municipality's' surprise, the Ministry of Interior rejected their application with the argument that they really liked their idea and wanted to propose the government to change to this legislation rather than granting one municipality exemption. As the data scientist noted: "I mean, it really surprised us that they didn't want us to test it [in just our municipality] before granting all municipalities legal permission to do this" (interview, June 2021). At this point, the emerging ethical plateaus of AI in child protection, are thus rather wide. The algorithmic model is valued for enabling the "early detection of (…) risk factors in the parents before symptoms of maltreatment appear with the child, and hereby secure an earlier and more effective prevention of vulnerability" (Internal document[3]).

Controversy I: Algorithmic surveillance (2018-2019)

In March 2018, the Danish Government referred explicitly to Gladsaxe's application in their policy initiative to combat so-called 'ghettos'. One of the initiatives was to detect minority children assumed to be in extra need of protection from "parents [who] are affiliated with countries with other parenting traditions where violence is legal" and the algorithm was mobilized as a tool to identify these children (Danish Govern-

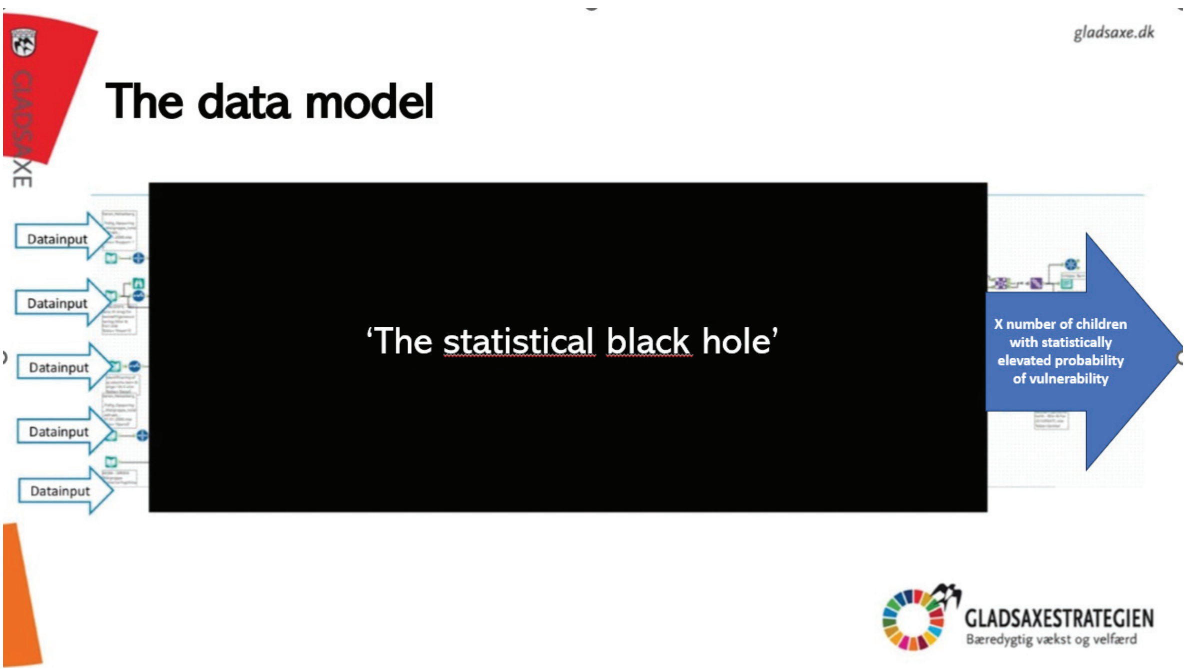

Figure 1. The statistical black hole, from Gladsaxe municipality power point presentation, 2018

ment, 2018: 30). This racialised figuration of the algorithm resulted in massive media attention to Gladsaxe's detection model. With headlines about "data-surveillance of families with children", a data-ethical controversy articulated issues of surveillance, biased decisions, a lack of transparency, risk of data misuse (Kjær, 2018). This public scandal introduced an important rupture to the emerging ethical plateaus of AI in child protection services, narrowing its wide planes contoured by techno-optimism to one establishing Gladsaxe's algorithmic model as unethical. As a result of the public controversy, several political parties withdrew their support to the profiling initiative in the Government's 'ghetto plan' and Gladsaxe had neither an exemption nor general legislation to support their tool (cf. Kristensen, 2022). With no legal mandate, the municipality had to put the algorithm on indefinite hibernate. Thus, the project was terminated before the algorithm could be fully developed and implemented.

The Gladsaxe model, and its racialised offspring as a 'ghetto plan', narrowed the ethical boundaries of what was possible to do with AI for child protection considerably. In this sense, Gladsaxe's project is also pivotal in the emerging ethical plateaus of AI for child protection services. Indeed, a search in Infomedia, a Danish newspaper- and magazine archive, revealed no hits on the terms child* AND algorithm* prior to 2018, whereas the number of articles continuously rose thereafter, most of them critical and focusing on Gladsaxe. This suggests that, at least in the media generated public, the coining of algorithms and child protection was not a matter of public attention prior to 2018. Today, the 'Gladsaxe model' has become a common point of reference in public and informal conversations about 'what can go wrong' in the use of predictive algorithms, to the extent that it has even become a hashtag (#Gladsaxemodel) in Twitter debates (the social media now known as X). Thus, we view the Gladsaxe experiment, the first attempt to develop AI for child protection services in Denmark, as the early formations of ethical plateaus in this field.

RISK (II): Algorithmic decision-support model to improve case workers' risk assessments

During the same period as the Gladsaxe experiment, a group of interdisciplinary researchers embarked to develop different predictive algorithm, here for the purpose of examining whether a decision-support tool could help social workers assess notifications. The potential value of this algorithmic model was envisioned in terms of hindering child maltreatment. As they wrote in a draft research article, "Child maltreatment has significant costs to its victims and, more generally, to society. Unfortunately, identifying cases of child maltreatment is a difficult task for Child Protective Services" (internal document). This difficulty was elaborated in their project description where they highlight a context of a growing number of notifications (from 97.288 in 2015 to 137.986 in 2019) (Project description: 2). The decision-support model was thus made valuable as a potential to help children but also a as research project assessing the efficiency of such a tool. Although sharing with Gladsaxe the objective of improving child protection, RISK's model also differs in important ways (Ratner and Elmholdt, 2023). First, rather than predicting risk *before* symptoms occur, RISK aimed predicting risk *after* symptoms had been notified. Second, rather than merging data from different welfare areas, they only wanted to use data that social workers could already legally access. Third, given their emphasis on the need to also research the value of such a tool from a social work and family perspective, their valuation of the algorithmic model as helpful was deliberately kept as an open question to be explored through research.

In the wake of the first rupture (I) in 2018, RISK published several reports on various aspects of their research endeavour, amongst them a report on their 'ethical considerations', dated October 2018. They sum up the ethicality of RISK, including testing it in social worker's practice, using the following words:

> It is ethically sound to test the tool in practice. Every day, assessments of notifications are made. The judgments and decisions, based on the assessments, are complex and they entail vast amounts of information and a series of ethical

> dilemmas, regardless of whether a statistical tool is being used or not. The ambition is to help children and young persons at risk in the best possible way. We adjudicate that the implementation of a statistical tool as a support for the qualitative assessment to be ethically sound and potentially improving the protection of vulnerable children and young persons" (Report on ethical considerations, October 2018).

Drawing the distinction between the complex, qualitative human assessment and the quantitative assessment offered by the algorithm, the role of the algorithm as a supplementary tool is emphasised. Here, the algorithm is valuable as a support in an already complex decision-making situation. With this valuation, the predictive algorithm is enacted as 'just' a tool, which will not render ethical dilemmas mores complex – because they are already complex. As a tool, in contrast, it might have the ability to improve how children and young persons are helped. Thus, the algorithmic processing of data was emphasised as ethically valuable, with the complexity of qualitative risk assessment already being ethically difficult supporting this valuation.

Controversy II: Algorithmic bias (2020-2021)

During the fall of 2019, as RISK went public with preliminary results from their first testing of the statistical model in the child protection departments of two municipalities, journalists and data scientists started scrutinising RISK and its algorithm. In January 2020, the algorithm was called out as a "shadow version of the Gladsaxe model" in a tech-magazine (Andersen, 2020), and in social-media platforms, it was criticised in harsh terms for the mere idea of developing algorithms in the field of child protection. In a comments section, it was for instance called a case of "contempt for professional knowledge, incapsulated in tech-utopianism' and 'an assault on the population". When a magazine through the help of a science student found biases in the algorithm, the controversy shifted towards a more technical debate.

The article describes the risk of two kinds of biases. One has to do with the reproduction of biases from the data input – here 'age bias' – which makes the algorithm assume that "the severity of a neglect increases with the age of the child" (Kulager, 2021).[4] The other bias, 'automation bias', has to do with the impact of using the algorithm as decision-support, when or if the caseworkers (uncritically) adjust their decisions to match the risk scores of the algorithm. In the corresponding media debates scientists, professors, social worker, and lay persons dispute the technicalities of the algorithm, the datasets, the validity of the research, the caseworkers' decision-making among several other aspects that touches upon the ethicality of even testing algorithms in the field of child protection. One professor simply judged RISK to be "irresponsible" (Andersen, 2021b). These problematisations thus disputed both the value of the algorithmic model but also the very idea that a decision-support tool could add value to child protection.

The members of RISK reacted to the ethical controversy by participating in the media debates – explaining, arguing and providing answers about how and what they have done, and which iterations they were considering for the second version of their model. They emphasised their willingness to create "openness and transparency about such a difficult subject as the use of machine learning in social work." (Andersen, 2021b). Meanwhile, in June 2020, they marked their publicly available reports with a stamp saying: "Temporary brief. Expired June 2020" and wrote new reports for the next phase of the project. Whereas the expired versions of reports are written in Danish, the new versions are framed in an open, English and theoretically grounded language, cementing their scientific ambitions.

During this revision, the dual valuation of helping children and researching the value of algorithmic decision-support shifted towards the latter. As the project manager explained: "Our hope is that we will be used as a knowledge base, somewhere in the debates, and not be put in a corner as those proposing that we should assess notifications by running them through an algorithm" (interview, project manager, March 2022). This valuation, on the one hand, enacts the algorithm as a valuable object of scientific scrutiny and, on the other hand, it positions RISK as ethically more legitimate than the other ongoing algorithmic experiments since it is the only one being done as research. As a project member noted during a conversation: "algorithms are too dangerous to simply let loose in practice, without research about the consequences" (field notes, June 2022). Thus, although the algorithm might be dangerous when tested on real cases, the valuation here emphasises the importance of researching algorithms before they are 'let loose'.

Controversy III: Legal uncertainty and lack of legitimacy 2021-2022

During summer of 2021, a scholar of social- and administrative law laid out her take on the legality of employing machine learning algorithms to support child protection services in the media (Andersen, 2021a). Scrutinising the legality of RISK's aim to test the algorithm on real cases, she disputed their legality report, written by the state attorney who had approved the algorithm as legal. A main objection concerns the legal requirement to individually consider which information about a child to collect from a principle of data minimisation; a requirement rendered impossible by the standardised algorithmic collection and assessment of various data points.

In view of this critique, the research group reached out to the state attorney and asked them to re-assess their own audit. In March 2022, the state attorney reaffirmed their earlier assessment, i.e., that they had legal backing for testing the algorithm on real cases. However, rather than finding solace in the authority of the state attorney, the project members at this point began to emphasise a different aspect of this report, leading them to doubt the legality of testing the algorithm on real cases. The project manager explains:

> Our judgement is that it [the state attorney's approval] is simply not solid enough ground for the project to stand on. (…) They [the state attorney] leave several doors open – for instance 'under the conditions of agreement in the field' – and we don't see that [agreement]. Our judgement is, firstly, that we will not continue with the project if there is any doubt about the legality. (…) Another position would be to say: 'well, if the state attorney says it is legal, then there is no doubt about the legality'. But then we know, we will be the object of even more criticism than we have already been, right? We do not wish to be in that position again. (Interview, project manager, June 2022)

As the quote also illustrates, the ongoing critique of their research project points to the lack of legitimacy and due to this, it is not clear cut whether there is a legal basis for testing the decision-support model on real cases. In this regard, the legal rupture in 2021-2022 affords a change in the tectonics of the ethical plateaus of AI for child protection where disagreement about the legal basis becomes the starting point for new interpretations of what is possible to do with algorithms, even if this happens under the label of research.

Even though RISK ends up not testing the algorithm in practice, due to its doubtful legality, the algorithm itself continues to live in a new version where it will only be tested in what the project manager calls "safe environments" – i.e., with artificial data in experimental workshops with child protection caseworkers. This cements a valuation where the algorithm is purely a research object and is delegated the role of acting as guinea pig in a laboratory like setting. Without access to 'real life' cases, it will have no influence on children's lives. The (human) project members, in turn, are delegated the role as researchers, constructing and controlling the artificial setting in which the algorithmic model is to be examined. This marks a shift in the ethical plateaus where algorithms are considered too dangerous to be used for decision-support.

Municipality X: The algorithmic detection of acute notifications

In 2019, two other municipalities began algorithmic developments, in the wake of the Gladsaxe controversy. Even though they were not subject to public controversy, the very existence of these, as we will see, shaped their ideas about what was ethical to do with algorithms. Here, we focus on the algorithmic development in municipality X.

The purpose of the third algorithmic model was to screen emails with notifications of concern for children's wellbeing and identify notifications needing acute responses – the so-called "red notifications" (project description) and in this way prioritise their assessments of notifications. Recalling their initial idea, the project manager explained that they chose the most vulnerable citizen group (children) for algorithmic experimentation because "this is where we have the greatest potential of being able to help (…) and because, there is a big volume of data with about 7000 notifications per year. This [assessment of notifications] is, of course, a large and difficult task, every day" (Interview, project manager, January 2022). Articulating its value, the Head of Strategy and Governance characterised the algorithmic model as a "super smart person [Kloge-Åge], like a senior employee, who has accumulated knowledge. (…) An artificial person, the caseworkers can consult, as a support" (Interview, Head of Strategy and Governance, January 2022).

From the onset, the project was very aware of the risk of public controversy. As the evaluation report stated:

> Together with the public affairs department, the project has developed a so-called preparedness [beredskab], to answer the many questions we anticipated to emerge. At this point in time, the media had been circulating stories about the use of AI in relation to case work, e.g., from Gladsaxe municipality. The project was therefore very attentive to having a model that supported [email] categorisation, and in no way supported the execution of interventions. (Internal document)

The valuation of this algorithm is at once formulated positively, as an administrative help in handling the large amount of notifications, yet also negatively in terms of what it was *not*, i.e., the Gladsaxe model.

This practice of 'un-ethicising', i.e., positioning themselves as 'ethical' in comparison with others 'unethical' practices (cf. Tønnesen, 2009), continued when RISK became subject to the bias controversy in 2020. Interviewed by the tech magazine Version2, a critical voice in both Gladsaxe's and RISK's public controversies, the Head of Strategy and Governance stated: "We are more oriented towards cleaning for bias than I think they were in RISK. And our project does not make use of the profiling of citizens or predictions [compared to RISK]" (Internal document).

RISK's controversy in 2020 also produced a concern about the risk of automation bias, characterised by the evaluation report as a situation where "human judgements unconsciously lean toward the categorisations generated by the AI, and thus creates an unintentional effect" (internal document). This led them to ensure that the "AI-models' categorisations to have the least possible impact on the caseworkers' decisions" (Interview, Head of Strategy and Governance, October 2022). Thus, instead of visualising the acute-labelled notifications during decision-making, they ran the algorithm as a so-called 'shadow process' and showed the caseworkers the algorithmic classifications during weekly meetings, *after* notifications had been prioritised. In this regard the algorithm was less valued for its ability to provide support during decisions and more as an opportunity to learn and reflect about what is possible to do with algorithms (Interview, project manager and Head of Strategy and Governance, January 2022). This valuation also hinges on it being a 'non-decision model' as the usage of algorithms for decision-support was deemed unethical.

This ambivalent valuation is reflected throughout the evaluation report, our interviews with the project manager and Head of Strategy and Governance. Rather than defining the algorithmic model as something specific, both the project manager and the head of strategy and governance in the municipality emphasise that their main goal is to learn: "what can we do with AI?". The Head of Strategy and Governance continuous:

> This is the cool thing about the project. We get to investigate what is possible and where are the boundaries? (…). The purpose was to (…) feel the boundaries of what is applicable, what is acceptable, what is meaningful and so on. (interview, Head of Strategy and Governance, January, 2022).

This process of adjusting according to "what is acceptable" demonstrates the shifting ethical plateaus to not only be a theoretical concern but a very practical one. It establishes the algorithm as a valuable means to take part in the drawing the boundaries of this new field of innovation in public administration.

With the final evaluation report concluding a rather low accuracy as 0well as a lack of trust from case workers, it was decided to keep the model running in shadow mode, which, apart from enabling organisational learning, had the assets of "retaining knowledge and competences in the IT-department, which, as a consequence, will be better equipped for working with AI, also in other units" as well as "retaining the possibility of eventually further developing the AI model [for future implementation]" (internal document). Thus, in the end, its valuation has been solidly reconfigured as a (vague) future potential, both in terms of the model itself but also in terms of the municipality's AI competencies.

Municipality Y: Algorithmic sorting of emails – the new colleague

Also in 2019, another municipality (Y) prepared for an experiment in the field of child protection. This municipality also used machine learning to develop an algorithm to detect notifications in an email inbox and to analyse their acuteness. In a meeting agenda in 2020, the algorithmic model was valued for its capacity to "[save] employees the time used to search for notifications [in the mailbox]" (internal document). Indeed, the algorithm was described in meeting agendas and power point presentations as a "mail sorting programme". In an interview with the project manager in charge of the project, we were told that they specifically wanted to "avoid the mistakes that others [Gladsaxe and RISK] have made" (interview, Project manager, November 2021) when deciding how to design their model. She explained these mistakes as 1) the merging of data from different welfare departments (Gladsaxe) and 2) using the predictive capacity of algorithms (Gladsaxe and RISK). "Therefore", she added, "[our algorithm] only collects data, which relates directly to the function it has" (interview, Project manager, November 2021), i.e., searching and marking emails with notifications containing words indicating acuteness. And she relates their choice of model to municipality X as she underscores: "We do not attempt to prioritise". Thus, rather than expanding human analytical capacity, the algorithmic model was trained to do the same as the caseworkers, only faster. It is the speed and not the scope which is articulated as valuable. This minimalised model showcases how the Gladsaxe controversy established clear boundaries of what *not* to do. Correspondingly, we here see how the ethical plateaus of AI for child protection shift towards benefitting administrative work rather than the child and its family.

Despite this starting point, the ethicality of the mail-sorting programme was a concern to begin with. They were particularly concerned with the risk of the algorithm making mistakes in its mail sorting. As they wrote in a meeting agenda: "Whereas "ordinary" IT-systems can "take care of themselves", machine learning demands more ongoing maintenance" (meeting agenda, October 2020, citation marks in the original). Correspondingly, they enrolled a team of skilled, administrative caseworkers to test the algorithmic model in what they termed a "hyper care period" of three months with careful attention to its precision. For this purpose, they personified the algorithm as a new colleague, naming it 'Naomi Notifications'. Below, the project manager describes how she introduced this 'new colleague' to the test team:

> As someone who, well, doesn't care if she sleeps at night and who doesn't go to the toilet, doesn't need food, and that sort of thing. (…). We always say that when we [introduce the algorithm] ...But, because it is also...if you have a challenge with a turnaround in employees, then you can say: "we have a technology who absorbs data in the same way as an employee". It [the algorithm] is a way of consolidating knowledge (…). We try to explain to them that it [the algorithm] can become a very experienced employee who remembers well and can work fast. But to begin with, it isn't. It is more like having an intern. (Interview, project manager, November 2021)

Besides of evoking the algorithm as a person in need of care and training, the personification also enacted the algorithm as a future potential rather than a problem-solving tool for the present. In this valuation, the figure of a new, untrained colleague mobilizes the algorithm as unfinished and full of beginners' mistakes. And, as in the case with interns, the future potential is only achieved through the professional involvement of the human employees, critically scrutinising the algorithm's work.

An evaluation report, however, concludes that this strategy failed, because the enrolled case-workers, instead of providing the anticipated feedback on its precision, they were simply "viewing it [the algorithmic model] as a 'search function' – as in a word document – and correspondingly, they think it cannot be better than it already is" (internal document). The evaluation report concludes: "there are no signs of impact on practice in any particular way". Nonetheless, the algorithm continues to run, and the project manager explains to us that she views it more as a curiosity project, which helps them learn, both how the algorithm and the employees react when machine learning algorithms are employed in practice. This shifts the valuation of the algorithm from being a potential time-saving administrative tool to one that can teach them about employees' use of algorithmic models.

Coinciding with this, the European Court of Justice gives their verdict on the so-called Schrems II[5]-case, establishing the use of American cloud services as non-GDPR compliant. Running on Microsoft cloud services, the municipality therefore kept the algorithm from being implemented in other departments. In spring 2022, a new manager took over in the social service department and started to enquire about the costs of running the algorithm. In an email to us by the project manager, she narrated her estimation that 40.000 kroners per year for a tool that does not make a difference is a lot of money (email, June 2022). They decided to stop the algorithm entirely. As a side comment, in an interview the project manager mentioned that the team had stopped using the algorithm before it was paused because it had begun marking the emails incorrectly (interview, project manager, June 2022). Indeed, the algorithm suffered from the lack of care and trainings. Yet, rather than reconfiguring its potential value, the combination of GDPR-compliance and its lack of positive impact resulted in it being devalued as a mere expenditure not even worthy of ethical concerns. Thus, while informed by the controversies of the other two experiments, the termination, and de-valuation, of the algorithmic model was the end result of many different agencies: lack of training, poor evaluation and GDPR-compliance.

Concluding discussion

In this paper, we have described the emergence and ruptures of ethical plateaus in Danish child protection services. Analysing the relationship between valuations of algorithmic models as (ethically) good, public controversy, and, eventually, the processes that lead to their termination or revaluation, we gained insight in changing boundaries of what is ethically possible to do with algorithmic models in Danish child protection services. In doing so the paper contributes to calls for "situation-sensitive approaches" in the study of AI ethics (Hagendorff, 2020: 14) and it contributes to STS discussions about the making and configuration of (AI) ethics (Seaver, 2021; Ziewitz, 2019). Below, we discuss two implications of these findings, in terms of (1) the role of public controversy in configuring ethical plateaus and algorithmic development and (2) the relationship between ethical plateaus and the distribution of agency across algorithm and humans.

Firstly, we learn how national media scrutiny is important in contesting the ethicality of child protection algorithms, mobilising the responsible organisations to publicly account for – and hence enact – their algorithmic model as ethically good. Figure 2 provides an overview of the development in valuations of the four models, including public controversies and other events that led to their re-valuation or termination. Here, we see how the first controversy, focusing on the Gladsaxe model, resulted both in the termination of this project but also led RISK to produce documents about the ethicality of their research project. The Gladsaxe controversy also influenced the very formulation of the algorithmic projects in municipality X and Y, in terms of becoming an example of what was *not* ethically acceptable to do with algorithms. These two projects, thus, from the outset limited the scope of their algorithmic models, valuing them in relation to administrative and time-consuming tasks rather than vulnerable children. Similarly, we see how controversy II and III, problematising RISK's plan to test their algorithmic model on real cases, led RISK to re-value the algorithm to the extent that it would no longer be tested on real cases. The controversy regarding algorithmic bias, here automation bias, further led municipality X to keep the algorithmic model running in a shadow mode, decoupled from case workers' prioritisation practices.

Moreover, municipalities X and Y, not terminated by public controversy, are important indications of what is ethically acceptable at the time of writing. Although developed in parallel in two different municipalities, with similar justifications for initiating their projects and shaped by ideas of what *not* to, their valuations of the algorithmic models ended up quite differently. As it seems, the boundaries of what is possible to do in the ethical plateaus of AI for child protection are narrowed down to a point where the employment of algorithmic models seems to be a future ambition rather than a problem-solving tool for the present. In this regard, the practitioners involved in the two projects were reacting to an ethical obligation to prepare the public administration for an imagined future with AI. Whereas municipality X's model became re-valued as a tool for organisational learning and the retainment of AI competences,

Table 1. Ethical plateaus of AI in Danish child protection services 2017-2023.

	Valuations			
Time	**Gladsaxe model**	**RISK**	**Municipality X**	**Municipality Y**
2017	Valued as helping vulnerable children through earlier interventions	Valued as helping vulnerable children through pre-emption of their maltreatment		
2018	**Controversy I: Algorithmic surveillance**			
		Valued as equally ethical as human decisions to test on real cases		
2019	Controversy 1 cancels algorithmic development		Valued for its potential to save time by prioritising notifications	Valued for its potential to save time by identifying notifications
2020	**Controversy II: Algorithmic bias and risky research**			
		Re-valued for its potential to generate knowledge, hence legitimising testing on real cases	Algorithm reconfigured as shadow process. Re-valued as a learning tool for reflecting about how case workers prioritise notifications.	
2021	**Controversy III: Illegal to test algorithm in practice**			
				Evaluation report: Algorithm is not trained and does not benefit practice
				Re-valued as an opportunity to learn how case workers interact with AI
2022		Re-valued as an algorithm only fit for testing on artificial cases	Evaluation report: Low accuracy. Case workers do not trust algorithm	New manager
			Continues as shadow process. Re-valued as a medium for retaining AI competencies in the municipalities	Terminated. Devalued as unnecessary expenditure
2023				

municipality Y's model was devalued as a useless "search function" and unnecessary expenditure.

What are the implications of these developments for our understanding of controversies role in shaping ethical AI? Following Marres' (2021) call to study public controversy, we agree that public controversies should not simply be analysed as instances of democratic interrogations of technological innovation but also as trial grounds for those developing algorithms. Arguing that public controversies increasingly are becoming strategic resources, "[providing] opportunities for the configuration of new markets", Marres warns against "romantic misunderstandings of how scandals happen"(Marres, 2021: 2). In our study, shaping the innovators' sense of what was ethically acceptable to do with algorithms in child protection, public controversy indeed did much more than render visible relations between science, technology, and society. They inadvertently became a platform for scoping what was possible and acceptable to propose to do with algorithms, the two municipal algorithmic models being rather explicit about this.

At the same time, there was no evidence of public controversy being "purposefully used to organise publics for 'innovation'" (Marres, 2021: 13). On the contrary, our interviews with those subject to public controversy indicate that they experienced it negatively. Indeed, municipality Y was concerned that their very use of the term 'algorithm' would associate them with the other contested algorithmic models, and the members of RISK were concerned of more public critique if they followed the state attorney's assessment of their legality. Thus, public controversy was central in shaping ethical plateaus and fed into processes of adjusting the algorithms to what seemed acceptable, but they never became strategic platforms for the innovators.

The second implication of our analysis regards the powers and agency distributed to algorithms. During the emergence of the ethical plateaus, where algorithms were primarily articulated as a solution, most agency was delegated to them. The Gladsaxe model was granted the proactive role of detecting children at risk through prediction, thus intervening *before* any concern or symptom had been registered. Indeed, in encountering critiques of surveillance during rupture 1, the municipality argued for the ethicality of the algorithm by highlighting its 'black hole' processing of data as more ethical than human processing. Compared to the contemporary focus on transparency and responsibility, this articulation may seem absurd but is indicative of the optimism characterising the emergence of the ethical plateaus in Denmark. RISK's decision-support model delegated less agency to the algorithmic model, articulating it as a support *in* human decision making after humans had detected children at risk. Yet, with more ruptures and critique, their enactments of the algorithmic model changed over time, firstly, emphasising the necessity to *research* algorithmic models in use rather than simply implement them without proper evaluation, and later, they decided not to test their model on real cases, removing any risk of harming real casework. Finally, the two municipalities, which initiated their projects after the first public controversy, decided from the onset to limit the agency of their algorithmic models, keeping them from profiling and predicting citizens and from interfering in decision-processes. Whereas one municipality decided to limit the agency even more by reconfiguring it as a possible background algorithm, which they could consult for purposes of reflection *after* decision-making, the second municipality entirely closed their model.

Thus, we see how changing ethical plateaus and a growing awareness of potential controversiality has the effect that gradually less agency is delegated to the algorithmic models, both across the cases but also within those that have been reconfigured. This speaks to Lee and Helgesson's (2020) observation that valuations of algorithmic processes are entwined with distributions of agency across human and system. Thus, ethical plateaus not only shape what is ethically possibly to do with algorithmic models, they also influence on how much agency is given to algorithms – both with regards to how proactive they are *vis-à-vis* humans but also in terms of the roles they are envisioned to have.

Compared to the established literature on AI ethics, with its focus on (developing and assessing) ethical principles for AI, our approach allowed us to explore the relationship between public controversy's contestation of the ethicality of algorithms and enactments of algorithmic models as good or valuable. This analytical move changes ethics from principles, against which organisational practices can be evaluated, to situated valuation practices. Valuations do not come before (as regulation or involvement) or after (as evaluation), they are part of the very development of algorithms – in this regard the valuation shapes the algorithm to an extent where it can also kill the algorithm by devaluing it as unethical, illegal or as a mere expenditure. This requires us to approach 'the ethical' as a process that is incomplete, uncertain and situated. While our movement towards situated ethics obviously could be criticised for deflating the concept of ethics and for destabilising important efforts to scrutinise algorithms critically and holding them accountable, we propose the contrary: Namely, the endeavour to understand ethics as an emic concept, that is, how organisations mobilise, negotiate and enact the ethically good algorithm. This can teach us important lessons about the lived organisational realities of algorithmic ethics and may potentially make it easier to understand failed attempts at implementing ethical principles and how the figuration of what is ethical is central in the co-constitution of AI and society.

Acknowledgements

Funding from the VELUX FOUNDATIONS (Algorithms, Data & Democracy Jubilee Grant) supported the writing of this article.

References

Andersen T (2020) Gladsaxe-modellen spøger: Nyt AI-projekt skal forudsige mistrivsel hos børn. 29 January. Version2. Available at: https://www.version2.dk/artikel/gladsaxe-modellen-spoeger-nyt-ai-projekt-skal-forudsige-mistrivsel-hos-boern-1089919 (accessed 3 March 2021).

Andersen T (2021a) Jurister: Børneopsporing med kunstig intelligens er ikke forenelig med lovgiving. Available at: https://www.version2.dk/artikel/jurister-boerneopsporing-med-kunstig-intelligens-er-ikke-forenelig-med-lovgiving (accessed 15 January 2023).

Andersen T (2021b) Professor advarer mod kommunal udrulning af børne-AI: »Jeg synes, det er uforsvarligt« | Version2. Available at: https://www.version2.dk/artikel/professor-advarer-mod-kommunal-udrulning-af-boerne-ai-jeg-synes-det-er-uforsvarligt (accessed 15 January 2023).

Andrejevic M (2017) Digital Citizenship and Surveillance| To Pre-Empt A Thief. *International Journal of Communication* 11: 18.

Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11(6): 985–1002. DOI: 10.1177/1461444809336551.

Callon M (1986) Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In: Law J (ed) *Power, Action and Belief: A New Sociology of Knowledge?* London: Routledge, pp. 196–223.

Cuccaro-Alamin S, Foust R, Vaithianathan R and Putnam-Hornstein E (2017) Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review* 79: 291–298. DOI: 10.1016/j.childyouth.2017.06.027.

Danish Government (2018) Et Danmark uden parallelsamfund - ingen ghettoer i 2030. Available at: https://www.regeringen.dk/nyheder/2018/ghettoudspil/ (accessed 31 May 2022).

Dencik L, Redden J, Hintz A and Warne H (2019) The 'golden view': data-driven governance in the scoring society. *Internet Policy Review* 8(2). Available at: https://policyreview.info/articles/analysis/golden-view-data-driven-governance-scoring-society (accessed 4 May 2021).

Douglas-Jones R (2017) Making Room for Ethics. *Science & Technology Studies* 30(3): 13–34.

Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Fortun K and Fortun M (2005) Scientific Imaginaries and Ethical Plateaus in Contemporary U.S. Toxicology. *American Anthropologist* 107(1): 43–54.

Hagendorff T (2020) The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* 30(1): 99–120. DOI: 10.1007/s11023-020-09517-8.

Helgesson C-F and Muniesa F (2013) For What It's Worth: An Introduction to Valuation Studies. *Valuation Studies* 1(1): 1–10. DOI: 10.3384/vs.2001-5992.13111.

Jørgensen RF (2023) Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society* 26(1): 123–138. DOI: 10.1080/1369118X.2021.1934069.

Kjær J (2018) Regeringen vil overvåge alle landets børnefamilier og uddele point, *Politiken* March 3rd. Available at: https://politiken.dk/indland/art6365403/Regeringen-vil-overv%C3%A5ge-alle-landets-b%C3%B8rnefamilier-og-uddele-point (accessed August 11 2023).

Kristensen K (2022) Hvorfor Gladsaxemodellen fejlede - om anvendelse af algoritmer på socialt udsatte børn. *Samfundslederskab i Skandinavien* 37(1): 27–49. DOI: 10.22439/sis.v37i1.6542.

Kulager F (2021) Kan algoritmer se ind i et barns fremtid? I Hjørring og Silkeborg eksperimenterede man på udsatte børn. Available at: https://www.zetland.dk/historie/s8YxAamr-aOZj67pz-e30df (accessed 15 January 2023).

Lee F and Helgesson C-F (2020) Styles of Valuation: Algorithms and Agency in High-throughput Bioscience. *Science, Technology, & Human Values* 45(4): 659–685. DOI: 10.1177/0162243919866898.

Marres N (2021) No issues without media, The changing politics of public controversy in digital societies. In: Swati J and Wasko J (eds) *Media: A Transdisciplinary Inquiry*. Chicago: Intellekt Books, pp. 228-243.

Meilvang ML and Dahler AM (2022) Decision support and algorithmic support: the construction of algorithms and professional discretion in social work. *European Journal of Social Work* 0(0)1–13. DOI: 10.1080/13691457.2022.2063806.

Mittelstadt BD, Allo P, Taddeo M, Wachter S and Floridi L (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society* 3(2): 1–21. DOI: 10.1177/2053951716679679.

Puig de la Bellacasa M (2010) Ethical doings in naturecultures. *Ethics, Place & Environment* 13(2): 151–169. DOI: 10.1080/13668791003778834.

Ratner HF and Elmholdt KT (2023) Algorithmic constructions of risk: Anticipating uncertain futures in child protection services. *Big Data & Society* 10(2): 1–12. DOI: 10.1177/20539517231186120.

Redden J (2018) Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society* 5(2).

Redden J, Dencik L and Warne H (2020) Datafied child welfare services: unpacking politics, economics and power. *Policy Studies* 41(5): 507–526. DOI: 10.1080/01442872.2020.1724928.

Russell J (2015) Predictive analytics and child protection: constraints and opportunities. *Child Abuse & Neglect* 46: 182–189. DOI: 10.1016/j.chiabu.2015.05.022.

Seaver N (2021) Care and Scale: Decorrelative Ethics in Algorithmic Recommendation. *Cultural Anthropology* 36(3): 509–537. DOI: 10.14506/ca36.3.11.

Statistics Denmark (2022) www.statistikbanken.dk/UND1, (accessed November 11th 2022).

Tønnesen C (2009) *Ethicising: Towards an Understanding of Ethics as Material Practice*. PhD Thesis. Oxford: University of Oxford.

United Nations Department of Social and Economic Affairs (2022) UN e-Government Surveys. Available at: https://publicadministration.un.org/en/Research/UN-e-Government-Surveys (accessed 9 August 2023).

Villumsen AM and Søbjerg LM (2023) Informal pathways as a response to limitations in formal categorization of notifications in child and family welfare. *Nordic Social Work Research* 13(2): 176–187. DOI: 10.1080/2156857X.2020.1795705.

VIVE (2022) Fakta om anbringelser. Available at: https://www.vive.dk/da/temaer/anbragte-boern-i-danmark/fakta-om-anbringelser/ (accessed 12 December 2022).

Ziewitz M (2019) Rethinking gaming: The ethical work of optimization in web search engines. *Social Studies of Science* 49(5): 707–731. DOI: 10.1177/0306312719865607.

Notes

1 Recognizing that the term AI has many competing definitions, we here use it to signify algorithmic systems developed with the use of machine learning techniques.

2 Of course, compared to our brief history of changing ethical boundaries, geological plateaus develop over longer periods and may thus appear more stable.

3 We use the term "internal document" for references that we have as part of our research data, but which we cannot make public due to issues of confidentiality.

4 One explanation for the age bias is that schools send more notifications than daycare institutions. This means that the data set of notifications have an overweight of children going to school. This, of course, does not have anything to do with the situation of the child.

5 The ruling implicated that EU customers of US cloud services, such as Microsoft, Amazon and Google, were themselves responsible for assessing the risks of data being accessed by third party countries as well as verifying the data protection laws of the recipient country – a task impossible to achieve for a Danish Municipality (note from the Data Protection Officer of the municipality, 2021).

Pepper[1] as Imposter

Ericka Johnson
Department of Thematic Studies, Linköping University, Sweden/ ericka.johnson@liu.se

Introduction

"An imposter is commonly understood as a person who pretends to be someone else in order to deceive others" (Vogel et al., 2021: 3). This is the starting point of Woolgar and colleagues' (2021) recent work on imposters, in which they explore how thinking with imposters can be a useful analytic for social theory, i.e. a tool or lens through which to observe social-material phenomena. In the book, they trace early sociological use of imposters to articulate (underlying and/or performative) social orders, and how impostering was initially seen as an example of deviation from the normal. In these early uses, examples of impostering could be interpreted for clues to which mechanisms held together the social order. However, their reworking of the term impostering moves the figure of the imposter to 'center stage' and uses it to explore indeterminacy, uncertainty and disorder, the frictions and disruptions that are actually central to social relations (Vogel et al., 2021: 4). Rather than using it to discover underlying normative mechanisms, this new use of impostering keeps the analytical focus on the messy practices of social relations but also encourages analysis of which other actors are collaborating in the impostering practices, and what purposes the imposter is supposed to serve.

For my discussion here, I will use impostering to focus on the messy and collaborative practices of the human-robot relation involving the robot Pepper. This will complement an analysis of the power dynamics of (robotic) care. I find the analytic of the imposter is useful when combined with science and technology studies (STS) and feminist technoscience discussions of care to reveal the complexities of human and non-human actors and discursive concerns engaged in presenting robots like Pepper as solutions to the care needs of older adults in residential eldercare homes.

This combination of impostering and care is particularly relevant for the study of robots and their introduction to the social constellations of care, a place that Pepper is often imagined to inhabit sometime in the near future. But, as I will show, using the figure of the imposter to explore Pepper makes apparent how the robot is both an essential but not unassisted character in the production of (imagined) caring relations. These relations, which are often messy and muddled, involve more users and more desires than one often finds mentioned in robotics research. Using Pepper as an imposter in this analysis articulates the underlying relational (dis)orders of care provision and makes visible how much work is required to choreograph the provision of care. It also starts to unpack the entangled relation between Pepper as a figuration (a character in different imaginaries who is assigned roles in narratives of care and care provision) and Pepper as a research object.

Robots as imposters

Just like the imposters of the older sociological term (magicians, con-artists, forged art-dealers), robots can stir complicated emotions when we in the well-resourced North are asked to seamlessly integrate them into our daily lives. They mow our lawns and vacuum our floors. They park our cars and, in some imaginaries, even drive them. In disembodied form they decide what playlist of music we should listen to. They sit in speakers on our fireplace mantle, listening, and turn the lights on and off while logging our activities and speculating about potential purchases from snippets of overheard conversation. Some people think they are going to be able to fulfil our needs for friendship and intimacy (Strengers and Kennedy, 2020), and some robots already do, as Harrison found in her study of chat-bots on infidelity sites (Harrison, 2019).

For the last few years, I have been working with a team of colleagues on an interdisciplinary robot project, thinking about the ethics of care robots while doing ethnographic and interview-based studies of robots in development. Part of our study has involved collaborative work with roboticists, participating in their studies with the semi-humanoid robot, Pepper. These studies have engaged older adults supposed to follow a series of aerobic movements led by Pepper. Our roboticist colleagues have been interested in how Pepper could produce and read engagement. As social scientists, we worked with them in developing the studies and ethnographically observed as the research was conducted. We also paid attention to other things happening around us. We made notes about the way experiments were run, the smooth parts and the glitches, but we also made notes when the experiments were interrupted because the laboratory received visitors, for example a small group of managers from the local municipality who wanted to discuss the integration of robots into existing care configurations. For us, these disruptions were as much a part of our material as the studies themselves.[1]

As a backdrop to this study, I have also been sensitive to the way Pepper is presented in various popular media and in care discourses as a solution to a bouquet of care needs predicted to appear in the near future. In this paper, I will be reflecting on these imaginaries of Pepper as a figuration. A typical image of Pepper the care robot for older people shows Pepper standing in a room, leading a smaller group of older people in chairs through some arm-waving exercises. *The Times* (Cavendish, 2018) presented one such image, but a simple Google Image search will present many examples of this figuration.

Our larger study is informed by recent, ethnographic work on how robots are being developed, studies which have shown that the practices and imaginaries of engineers and designers are (still) helping to shape the types of robots that are developed to 'serve' us (Fischer et al., 2020; Robertson, 2017; Søraa, 2021). However, here I will also be drawing on work that explores the imaginaries and figurations of the robotic on the edges and outside of the lab (DeFalco, 2020; Rhee, 2018; Sparrow and Sparrow, 2006; Strengers and Kennedy, 2020; Suchman, 2007). There are many different constellations of sites which are currently developing robots and which are engaging different user imaginaries and landscapes: engineering laboratories; social robotics labs; commercial entities entangled with university research structures; municipal and regional innovation platforms with input to and from research groups. These constellations are often interdisciplinary collaborations engaging theories and methods from cog sci, psychology, design, linguistics, STS. Fischer et al. (2020) suggest that the sites of design impact the image-evoking activities used by designers and engineers. We have also seen this with the robotic work we have followed at a robotics lab, where Pepper is entangled in concerns that non-academic actors have (for example, the municipality's team of managers mentioned above). Thus, it is logical to suggest that the work done at these sites is also embedded in cultural discourses, in the hegemonic quadrant of power that frames our imaginaries, helps us imagine what is possible, and therewith impacts the research questions we can pose. Exploring these framings is thus important to understanding how our research takes shape and what robots it can produce.

The imposter as analytic for the robots

In wider social discourses about care robots, Pepper is often presented as a robotic solution to the problem of care for older adults. Occasionally, Pepper, or a similar humanoid-ish robot, is seen with out-stretched arms about to lift an older person or interact with that person one-to-one. But often, the images shown in newspaper articles, YouTube clips, and promotional materials are of Pepper leading aerobics classes for older people, sometimes in robotics labs but usually in care home environments. Pictures of Pepper helping lead exercise groups (usually focused on the upper body) are a stock representation of how Pepper is imagined to be addressing social needs today and tomorrow. It is not a coincidence that this imaginary made it into the robotics lab we collaborated with. Their research – and the research of many roboticists working with Pepper – is impacted by the question of how Pepper could react in a similar situation and is, as Suchman would put it 'infused with its inherited materialities' that afford limited conceptualizations of what the robot can do (Suchman, 2011: 119). For example, one nuance worth mentioning is that Pepper has no movable legs but does have movable arms, and leading an aerobics class for stationary human bodies is within the realm of possibilities for Pepper. Jumping around or shuffle dancing is not. When picking apart the nuances of Pepper as aerobics instructor, one can find the material limitations of the robot becoming entangled in our human encounters, both in promotional images and in the research done with robots.

Understanding 'imposter' in the earlier sociological sense, as a deceitful figure whose deception can reveal an underlying order, would highlight the work Pepper does when assuming some particular aspects of performing 'aerobic instructor', including producing roles for the often wheelchair-bound aged to perform (as wheelchair-bound and as aged) and the terribly enthusiastic and motivational frontstage mask that Pepper assumes when the robot unwaveringly smiles and flairs about its arms, encouraging the bodies in the wheelchairs to do the same. This understanding of Pepper as an imposter would also speak to the underlying distrust of robots (machines) to provide something normally thought to belong to the human realm, care (c.f. DeFalco, 2020).

However, the image of Pepper teaching an exercise class can also be understood with an alternative sense of impostering – the kind that Woolgar and colleagues (2021) want to put forward. Such an analysis recognizes the intentional deception, but highlights that for it to occur, others in the room have to make a series of moves; that pepper-the-imposter is not just a version of Jane Fonda or Richard Simmons minus the pastel-coloured sweatbands and legwarmers. Observing Pepper through the new lens of the imposter could prompt questions about the concerns surrounding and awakened by those robotic arm waving exercises. Who is involved in being intentionally deceived? Why? Pepper as an imposter can provide insight to the "disorganized" social relations and cultural forms from which it is emerging. In the case of Pepper the aerobics instructor, those social relations are imagined to be in need of re-organization in a care home which is probably trying to readjust their care provision to economic efficiency demands in the face of (at least imagined) labour shortages. The Woolgar and colleagues' analytic helps me look for these discursive moves behind the production of Pepper as an aerobics instructor by reminding me to find the actors making the moves necessary for the deception.

Seeing concerns in the care of robotic aerobics

The analytic of impostering reminds us that human/non-human constellations, networks, and/or entanglements are and have long been a part of care (DeFalco, 2020; Puig de la Bellacasa, 2011). We can use the imaginary of Pepper the imposter to probe the dynamic relations they are involved in.

For example, Pepper as aerobics instructor tries to get older adults to follow the arm motions that Pepper initiates. This happened in the experiments we observed, and is also prevalent in the visual and textual representations of Pepper. However, there is more at work than a mimicking of arm motions. Looking beyond the movement, one can also discern a concern that much older adults in care homes are not exercising enough, that they are in imminent danger of just sitting still, staring out into space.

We can also read a concern that 'activating' older adults by hiring staff to lead exercise groups would require too many personnel resources. One can sense worry about the price (salaries? social guilt? employment policies? HR headaches?) of paying someone else to lead the interactions. These concerns become visible as one untangles elements in the knot of societal needs, care provision policies, and robotics research that produce Pepper the aerobics instructor.

One can also sense concern that the older people around us are socially isolated. Entangled with that is a reluctance to socially interact with them, ourselves. Note the enactment of the categories us/them, othering the aged. These different elements of the discourse become clearer when sensed through the figure of an imposter, and can trigger questions that go beyond the roles that an imposter may be making visible and instead ask about *whose* concerns these are, the care providers or the care recipients? It can even prompt the question: is the imagined user of the robot the resident being nudged into participating in aerobics or is the user the children of those residents? Or the municipality managers interested in developing the interaction to help provide care? Who are the people responsible for the residents? Is Pepper addressing the social needs of the care home residents or the guilty consciouses of their children or the limited budgets of care provision institutions? Or all of them, together?

A related question of what social relations are being constituted when Pepper gets the older people to move their arms around in a coordinated way is: why is it important that these aged bodies are all gathered under one roof, comprising a target audience. In many of the images of Pepper circulating in care discourses, the people responding to Pepper are in a group. They are gathered together into a large room, all focused on the robot and doing just as Pepper tells them to do. They are disciplined – either by Pepper or (more likely) by the context that puts them into a collective home and demands they do as they are told. In many such photos, people look like they were wheeled into the room by someone else and positioned in front of Pepper. Again, this speaks to the power dynamics in group care, and the power dynamics of eldercare or the care of other bodies which need assistance, a type of power dynamic which is exactly a node of tension, of not-necessarily-nice-or-benign care (especially when combined with technology) that STS work on caring has helped articulate (Latour, 2004; Lindén and Lydahl, 2021; Martin et al., 2015; Murphy, 2015; Puig de la Bellacasa, 2011). Pepper and the use of robots like Pepper in residential care shows that the dynamics of care are not always benign (c.f. Murphy, 2015) and that the introduction of a non-human into the loop can articulate practices of valuation and value which we otherwise have been wont to ignore (unless we've been working in nursing or theories of nursing care – where these valuation practices have been explored in depth (c.f. Allen, 2013; Hochschild, 1983; James 1992; Tronto, 1993)).

Research on care robots has also pointed out their purported future use as a replacement for human bodies of 'undesirable' or 'uncomfortable' colors/races/nationalities/classes in the care/cared for relationship (Benjamin, 2019; Robertson, 2017; Sparrow, 2020). Countries that imagine the impending demographic crisis of aged adults in need of care, but which do not have sufficient labouring bodies to care for those adults, imagine being able to provide this care with robots rather than through racialized immigrant labour (Robertson, 2017; Sparrow and Sparrow, 2006). It has been suggested that this imaginary is not going to work (Wright, 2023)

Of course, not everyone welcomes a future with Pepper. The dystopian imaginary of care robots as perceived by older adults has also been discussed (Sparrow and Sparrow, 2006) and reflections on the subjectivity ideally assigned to these robots (as a responsive but unobtrusive servant) explored (DeFalco, 2020; Suchman, 2007) – even as the use of caring technologies like voice operated assistants has been widely adapted (for an in-depth discussion, see Strengers and Kennedy, 2020; Sutton, 2020; Søndergaard, 2019; Søndergaard and Hansen, 2018). But the 'utopian' dream of Pepper helping untangle that knot of needs and concerns about how to care for the older people in our lives seems still to be finding its way into our robotic imaginaries and our robotic laboratories.

Robotic constellations need work

If we understand an imposter as someone who is pretending to be someone else with the intent to deceive, this definition carries with it a sense of betrayal and deception. While that is a common theme in many science fiction works about robots (see *Origin* by Dan Brown, *Machines Like Me* by Ian McEwan, *Klara and the Sun* by Kazuo Ishiguro), and seems to be the goal with the life-like humanoid replicas produced by Hiroshi Ishiguro, this is hardly something that Pepper strives for. The smooth white plastic is vaguely shaped in a human form, and there is an unmoving face with eyes and mouth that gives a hint of humanness, but Pepper's design is in no way meant to deceive a user to think it is a human. Pepper is not an imposter like a magician or a con-artist. Pepper is a part of an impostering event, with other actors, wills and desires all messed up with Pepper's materiality.

This is why I find it useful to thinking of Pepper through the analytic of the imposter and at the same time engage the analytic together with the STS discussion of care and its discontents. Doing so produces a more actionable lens to view and articulate the other actors involved in producing the impostering event.

For example, we are asked to think of Pepper as the leader of the exercise moment. But often in such photographs, one can also see a human instructor live or on a TV behind Pepper doing the same exercise, complementing Pepper's instructions with practices of interfacing (Lipp and Maasen, 2022). One can see chairs and wheelchairs that Pepper's followers are sitting in and the way those body/chair hybrids allow some movements but not others. The room is also important, even if it becomes the background, with its closable doors that let the body/chair hybrids in and out at (someone's) will (Johnson/Latour, 1988), and the lights and heat or air conditioning that keep the human bodies comfortable. There is considerable material worlding going on to make the environment which produces Pepper as an aerobics instructor. And, of course, this aerobics session is a session – with a beginning, middle and end, a dramaturgical arch that supposes that Pepper's presence in this narrative will transform the social order that existed in the room before they entered. One could even suspect the whole event was carefully staged.

However, when understanding imposters as "engines of indeterminacy, uncertainty and disorder," (Vogel et al., 2021: 4) and thinking of Pepper the aerobics instructor this way, my analysis snags and slows down on the term indeterminacy, not deception. Pepper is definitely a robot, yes, but as an aerobics instructor, Pepper is also imagined to be a motivational speaker, role model, cheerleader and generally pleasant persona helping create a sense of enthusiastic movement in the collection of previously still bodies in front of the robot. Pepper articulates an indeterminacy – or a complexity – in what we imagine that aerobics instructor is actually doing. A mild sense of disorder is produced from Pepper's actions, with occasional smiles (only on the human faces – Pepper's smile is permanent, unchanging) and the waving of arms. This robot aerobics instructor is creating a stir… albeit a relatively slow-moving stir. They are revealing an instability. Stabilizing this disruption into a legible example of group exercise requires work; staging props, and captions to tell the viewer what they are seeing.

Now, to be fair, recognizing the work that is necessary to make anything even remotely resemble a stable social interaction with a robot is not new. Human Robot Interaction (HRI) researchers use an established experimental technique called 'wizard of oz' in which the researcher speaks for or remotely manipulates the robot (often behind a one-way mirror or curtain) to trick the human participant into thinking the robot is actually interacting with them. This is done because many robot interactions are theoretically interesting but practically difficult, if not impossible, given the current state of technology (Baxter et al., 2009; Rea et al., 2017; Riek, 2012). Suchman (2007, 2011) uses rich ethnographic observations of her interactions with MERTZ to show with STS terms how much a robot interaction depends on the human participant or, more to the point, participants, in constellations of users, programmers, inventors, study designers, etc. In more recent work, Treusch uses this human participation in cobot research to articulate the complexities of practice (Treusch et al., 2020). STS gives us tools for thinking through this, for paying attention to the affordances of the material entanglements that engage bodies (of imposters and others) into contingent constellations that allow an activity to occur, ideas also taken up in design work, not least in the conversations about technological dis-affordances (Costanza-Chock, 2020).

In images of Pepper leading exercise classes and in the research we observed at the robotics lab, it becomes clear that other material artefacts play an important role in naming and shaming 'imposters' or in allowing them to 'pass,' and in staging them as subject objects (Suchman, 2011). This is something STS as covered extensively, with the relational turn and recognition of the way human and technological knowledge-objects are produced as intra-active phenomena (Suchman, 2007, 2011; Barad, 2007). One can assume that the group of older care home residents in this picture were told that Pepper was a robot who would be leading them through their exercises. One can also wonder at the technical limitations which led to this imaginary – an aerobics instructor – in a discourse about robots which is dominated by robots assisting the human labor force to increase efficiency and optimization. An aerobics instructor-robot in a care home seems almost, to use Treusch's term, an example of 'useful uselessness' (Treusch, 2020). Yet, there is Pepper, waving its robotic arms.

By thinking through Pepper the aerobics instructor as an imposter, I gain insights into the disorganized social relations that are framing the robotics research engaging Pepper. But using this analytic can do more than produce insight. It can be an analytical lens that shows how Pepper is imagined to reconfigure the work of producing exercise for older people and why. This draws attention to the constellations of desires and wills, of reasons and resources, which are imagining robots in this way.

Doing so will allow us as researchers to challenge how the actors positioning Pepper and similar robots frame and inform robotics research, pointing out the political of what is often thought to be neutral, scientific robotics development. Finally, using this concept we can both draw from the theoretical discussions of relational work in STS, and also engage with our own affective responses to the emotional aspects of robotic care that are so present but so difficult to articulate as researchers. Impostering provides a poignant term to, in collaboration with insights from STS and feminist technoscience analysis of care, help articulate the ambivalence I have in my research; the unsettled feelings I have about the introduction of robots and the wary, almost unwilling attraction I note as I watch and imagine robots as care providers. By directing my attention to the messiness of care and its complex concerns, impostering combined with care can produce a vocabulary of resistance in the face of 'utopian' technological solutions to eldercare needs.

Acknowledgements

This work was supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation.

References

Allen D (2013) Introduction – The sociology of Care Work (Special Issue). *Sociology of Health & Illness* 35(6): 928-938.

Arnelid M, Harrison K and Johnson E (2022) What Does it Mean to Measure a Smile? Assigning Numerical Values to Emotions. *Valuation Studies* 9(1): 79-107.

Barad K (2007) *Meeting the Universe Halfway*. Durham: Duke University press.

Baxter P, Kennedy J, Senft E, Lemaignan S and Belpaeme T (2009) From Characterising three years of HRI to methodology and reporting recommendations. *ACM/IEEE International Conference on Human-Robot Interaction*. Christchurch, New Zealand pp. 391-398, doi: 10.1109/HRI.2016.7451777.

Benjamin R (2019) *Race After Technology*. London: Polity.

Cavendish C (2018) Robot carers for the elderly are now a reality in Japan but do we want them here? *The Times*. Available at: https://www.thetimes.co.uk/article/robot-carers-for-the-elderly-are-now-a-reality-in-japan-but-do-we-want-them-here-mw8zpw0zd (accessed 3.7.2023).

Costanza-Chock S (2020) *Design Justice*. Cambridge: MIT Press.

DeFalco A (2020) Towards a theory of Posthuman Care. Real Humans and Caring Robots. *Body and Society* 26(3): 31-60.

Fischer B, Östlund B and Peine A (2020) Of Robots and humans: Creating user representations in practice. *Social Studies of Science* 50(2): 221-244.

Garcia I (2021) A New AI Lexicon: Artificial Identity Cataracts. AI Now Institute. https://medium.com/a-new-ai-lexicon/a-new-ai-lexicon-artificial-identity-cataracts-29f73c32e6bb

Gleisner J and Johnson E (2021) Caring for affective subjects produced in intimate healthcare examinations. *Health* 27(3): 303-322.

Harrison K (2019) Automated and/or authentic intimacy: What can we learn about contemporary intimacy from the case of Ashley Madison's bots? *First Monday* 24(10).

Harrison K and Johnson E (2023) Affective corners as a problematic for design interactions. *Transactions on Human Robotics Interactions*. https://doi.org/10.1145/3596452.

Harrison K, Somasundaram K and Loutfi A (forthcoming) The Imperfectly Relatable Robot: An Interdisciplinary Approach to Failures in Human-Robot Relations. In Johnson E (ed) *What that Robot Made Me Feel*. Boston: MIT Press.

Hochschild AR (1983) *The Managed Heart. Commercialization of Human Feeling*. Berkeley: University of CA Press.

James N (1992) Care = organization + physical labour + emotional labour. *Sociology of Health & Illness* 14(4): 488-509.

Johnson J [Latour B] (1988) Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer. *Social Problems* 35(3): 298–310.

Latour B (2004) Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry* 30(2): 225-248.

Lindén L and Lydahl D (2021) Editorial (introduction to the special issue on care). *Nordic Journal of Science and Technology Studies* 9(1): 3-12.

Lipp B and Maasen S (2022) Techno-bio-politics. On Interfacing Life with and Through Technology. *Nano-Ethics* 16: 133-150.

Martin A, Myers N and Viseu A (2015) The politics of Care in technoscience. *Social Studies of Science* 45(5): 625-641.

Murphy M (2015) Unsettling care: Troubling transnational itineraries of care in feminist health practices. *Social studies of Science* 45(5): 717-737.

Puig de la Bellacasa M (2011) Matters of care in technoscience: Assembling neglected things. *Social Studies of Science* 41(1): 85-106.

Rea D, Geiskkovitch D and Young J (2017) Wizard of Awwws: Exploring the Psychological Impact on the Researchers in Social HRI Experiments. Proceedings of the Companion of the 2017 *ACI/IEEE International Conference on Hurman-Robot Interaction*. March 2017, pp. 21-29.

Rhee J (2018) *The Robot Imaginary.* St. Paul: The University of Minnesota Press.

Riek L (2012) Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1(1): 119-136.

Robertson J (2017) *Robo sapiens japanicus. Robots, Gender, Family and the Japanese Nation*. Berkely: University of CA Press.

Sutton SJ (2020) Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20)*. ACM 11, 1–8. DOI:https://doi.org/10.1145/3405755.3406123.

Sparrow R (2020) Robotics has a race problem. *Science, Technology & Human Values* 45(3): 538–560.

Sparrow R and Sparrow L (2006) In the hands of machines? The future of aged care. *Minds and Machines* 16: 141–161.

Strengers Y and Kennedy J (2020) *The Smart Wife.* Cambridge: MIT Press.

Suchman L (2007) *Human-machine reconfigurations: plans and situated actions*. Cambridge: Cambridge University Press.

Suchman L (2011) Subject Objects. *Feminist Theory* 12(12): 119-145.

Søndergaard DM (2019) On Humanoids, Avatars and the Rest of Us. In Hasse C and Søndergaard DM (eds) *Designing Robots, Designing Humans* Routledge: London, pp. 21-39.

Søndergaard, ML and Hansen LK (2018) Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction. *DIS* 2018. June 9-13.

Søraa R, Nyvoll P, Tøndel G, Fosch-Villaronga E and Serrano J (2021) The social dimension of domesticating technology: Interactions between older adults, caregivers, and robots in the home. *Technological Forecasting and Social Change* 167: 1-13.

Treusch P, Berger A and Rosner D (2020) Useful Uselessness? Teaching Robots to Knit with Humans. DIS '20, July 6-10, 2020 Eindhoven, Netherlands.

Tronto JC (1993) *Moral Boundaries: A political argument for an ethic of care.* New York: Routledge.

Vogel E, Moats D, Woolgar S and Helgesson CF (2021) Thinking with Imposters: The Imposter as Analytic. In: Woolgar S, Vogel E, Moats D and Helgesson CF (eds) *The Imposter as Social Theory.* Bristol: Bristol University Press. pp1-30.

Winkle K, McMillan D, Arnelid M et al.(2023) Feminist Human-robot Interaction: Disentangling Power, Principles and Practice for Better, More Ethical HRI. *HRI'23 ACM* ISBN 978-1-4503-9964-7/23/03

Woolgar S, Vogel E, Moats D and Helgesson CF (eds) (2021) *The Imposter as Social Theory. Thinking with Gatecrashers, cheats and charlatans*. Bristol: Bristol University Press.

Wright J (2023) *Robots Won't Save Japan. An Ethnography of Eldercare Automation*. Cornell: Cornell University Press.

Notes

1 Pepper is a white, plastic, ¾-sized, almost humanoid robot. In addition to writing with the roboticists (Harrison et al., forthcoming), our analysis has also focused on issues like the valuation of care and emotions (Gleisner and Johnson, 2021; Arnelid et al., 2022), the responses expected and provoked by human-robot intra-actions (Harrison and Johnson 2023), the integration of social optics through intersectional categories in the design and analysis of robots (Garcia, 2021), and feminist methods for HRI (Winkle et al., 2023).

Dutreuil Sébastien (2024) Gaïa, terre vivante. Paris: La découverte. 512 pages. ISBN 978-2-35925-140-1

Jérémy Rollin
jeremy.rollin@yahoo.com

Science being embedded in paradigms that are imposed through struggles is a relatively old idea (Kuhn, 1962). These struggles can be made visible by controversies, a classic entry point for science studies (Collins, 1985). In the book *Gaïa, terre vivante* Sébastien Dutreuil is thus following a well-known tradition, to which he has yet added an original touch. By starting from a controversy, he shows that it is not (only) the result of a debate on the administration of evidence, but a philosophical and political difference. In doing so, he adopts a Latourian perspective, which leads him to extend his analysis beyond the scientific field (Latour, 2004).

The subject of Dutreuil's book is the 'Gaïa hypothesis', introduced by James Lovelock in 1972, with the hypothesis referring to the possibility of considering life as a whole that consists of all the interactions between organisms and their global environment. Three issues interest Dutreuil here. Firstly, the controversy between Lovelock and biologists. Secondly, the gap between a scientific publication, with a modest coverage in the Earth and environmental sciences, and its appropriation by a much wider public. Thirdly, the link between politics and philosophy that Gaïa provides. In the first three parts of the book, Dutreuil successively analyses the scientific proposals on which Gaïa is based and the careers of the researchers involved, mainly Lovelock and, to a lesser extent, Lynn Margulis (their correspondence was also the subject of a previous book (Clarke and Dutreuil, 2022)). In the fourth part, Dutreuil situates Gaïa in the history of Earth and environmental sciences, and in the fifth part, highlights Gaïa's contribution to a new philosophy of science.

For Lovelock, Gaïa relies on the observations that life influences geology, the physical conditions for life are constrained, and external perturbations can lead to conditions unsuitable for life. For Dutreuil, these observations lead to different analyses between Lovelock and biologists, especially evolutionists. In the evolutionists' perspective, natural selection through reproduction allows living organisms to adapt to an environment that is itself changing, including through the involuntary action of living organisms. Yet for Lovelock, the stabilising effect of living organisms on the environment is primary. This debate is discussed in the book through examples and counter-examples used by the various parties, with the case of altruistic behaviour being particularly controversial. In the Gaïa approach, altruism enables the environmental changes necessary to sustain life, whereas for evolutionists altruism is not a competitive advantage in the process of natural selection. Finally, Dutreuil shows that the protagonists of the issue do not share a common definition of life, which can be understood as an individual or global concept; in biology, it is defined by criteria applied to individuals (reproduction, selection), whereas Lovelock's concept is more fluid and general. This can be explained by the fact that Gaïa was addressed less to researchers in biology than to those in the Earth sciences. For Dutreuil, part of the misunderstanding is also due to the use of the term 'hypothesis': Gaïa is not a hypothesis that can be tested empirically, but a new ontology, which explains why life is approached in a broad and not really fixed sense. As a result of thus not only different scientific communities disagreeing, the debate being ontological rather than an empirical, Dutreuil considers that the idea of controversy (shared in the public debate) is not appropriate.

One of the strong points of the book is that it reinscribes Gaïa in the history of the Earth sciences. Dutreuil shows how empirical geology was gradually supplanted by geophysical and geochemical models in the 1940s and 1950s, before the biological was gradually introduced into the models in a way that was contemporary with Gaïa, as biogeochemistry in the 1970s proposed a role for living beings in the cycles of matter. This latter approach was supported by institutions such as the International Geosphere-Biosphere Programme, which encouraged interdisciplinary research from an Earth system science perspective. In this way, Gaïa does not have a monopoly on the relationship between the biological and the environmental, nor on the perspective that combines complexity and anti-reductionism. Dutreuil succeeds in showing that Gaïa is not an isolated and eccentric project, while at the same time highlighting its specificity, which explains its ambiguous treatment within the scientific community. For Dutreuil, it is not simply a research programme, but a compositionist and vitalist philosophy of nature.

In Gaïa being approached from this perspective, life is an entity in itself, without functional integrations but with internal differentiations. These differentiations are the organisms (p. 277). Dutreuil calls this ontology 'vitalist' in a specific sense. Life on Earth is the central entity, with an influence that extends far beyond the material boundaries of the cells: for the author, the atmosphere is a vital extension of organisms (p. 322). In the same time, he adopts Latour's (2017) notion of 'compositionism', that is, connectivity without holism, where the entity is composed piece by piece, loops after loops. Life is thus an association of properties and processes rather than a presumed big whole. For Dutreuil, these characteristics of Gaïa explain its spread beyond scientific circles, but also its reluctance to be named in scientific publications. The author shows us that Gaïa's approach to the world is the result of a mixture of science, philosophy and politics (in terms of Gaïa's cultural and militant appropriation and Lovelock's public positions). This is as much part of its richness as it is part of its academic contestation – frontal in the case of controversy, more discreet in the case of invisibilisation, i.e. the use of Gaïan concepts without quoting them explicitly.

Despite its value, two critical comments can be made about the book. First, the reader may question Dutreuil's position in this issue. On the one hand, Dutreuil has begun his career in the field of Earth sciences (Dutreuil et al., 2009). On the other hand, he has worked with Lenton, one of Lovelock's two doctoral students (Lenton et al., 2020), as well as Latour, who has played a role in the contemporary popularisation of the Gaïa theme (Latour, 2017; Latour and Porter, 2017). All this background undoubtedly contributes to his analysis of Gaïa, and a reader coming from sociology may then resent the limited explicit self-analysis. Second, following a Latourian tropism, the analysis of the book follows Gaïa more in its epistemological and ontological implications rather than the researchers' social positions. Apart from the figure of Lovelock, researchers are seen primarily in terms of their scientific output and less in terms of their power relations, their social properties, and their day-to-day interactions. It is however possible to argue that the profession of researcher is also characterised by knowledge, skills, experience and a symbolic economy characterised by hierarchies, between disciplines, between networks and between peers (Bourdieu, 1997). Their professional autonomy is supported by the display of peer control and limited to a given field (research). As a result, Gaïa's interdisciplinarity and politicisation can put the researchers' habitus under pressure, but this subject is rarely addressed in this book.

References

Bourdieu P (1997) *Les usages sociaux de la science: pour une sociologie clinique du champ scientifique une conférence-débat*. Sciences en questions. Versailles: Éditions Quae.

Clarke B and Dutreuil S (2022) *Writing Gaia: The Scientific Correspondence of James Lovelock and Lynn Margulis*. Cambridge ; New York: Cambridge University Press.

Collins HM (1985) *Changing Order: Replication and Induction in Scientific Practice*. London ; Sage Publications.

Dutreuil S, Bopp L and Tagliabue A (2009) Impact of enhanced vertical mixing on marine biogeochemistry: lessons for geo-engineering and natural variability. *Biogeosciences* 6(5): 901–912.

Kuhn TS (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latour B (1991) The Impact of Science Studies on Political Philosophy. *Science, Technology, & Human Values* 16(1): 3–19.

Latour B (2004) *Politics of Nature: How to Bring the Sciences into Democracy*. Cambridge: Harvard University Press.

Latour B (2017) Why Gaia is not a God of Totality. *Theory, Culture and Society* 34(2–3): 61–82.

Latour B and Porter C (2017) *Facing Gaia: Eight Lectures on the New Climatic Regime*. Cambridge: Polity.

Lenton TM, Dutreuil S and Latour B (2020) Life on Earth is hard to spot. *The Anthropocene Review* 7(3): 248–272.

Oreskes Naomi (2021) Science on a Mission: How Military Funding Shaped What We Do and Don't Know about the Ocean. Chicago and London: University of Chicago Press. 738 pages. ISBN: 978-0-226-73238-1

Adam Tamas Tuboly
tubolyadamtamas@gmail.com

In *Science on a Mission*, Naomi Oreskes asks the seemingly simple question of how military funding has shaped what we do and don't know about the oceans and introduces us to the world of oceanography. In the first third of the twentieth century, one very rarely got on huge ships to cross new seas, struggle with ice monsters and nightmares, and discover new continents. In the United States, most have paddled out of California ports in small boats to take water samples, and collected plants and fish for later lab-tests. Few undertook this either, as there were only a few major institutes that dealt with the world of the oceans, and retirement jobs did not exist in these at first; salaries were low, and the outlook was, to put it mildly, bleak.

During World War II, as the Germans became successful at sea and underwater, oceanography became more and more interesting to the infamous Navy of the United States. Biologists, physicists, and mathematicians who previously had an unquestionable pedigree of spending most of their time in small boats and within four walls were now approached by military officers. The Navy promised money and bigger boats, and everyone saw that the military industry had inexhaustible resources. A promise made is a debt unpaid – and the army kept its word. As Oreskes explains (especially in chapters 2-3), leading institutions in American oceanography, such as Scripps and Woods Hole, all enjoyed the unbreakable flow of money, though internal resistance did raise its head from time to time.

The Navy promised a variety of goods: stable jobs, more staff and time for research because of the better conditions, advanced technology, and social esteem; all in all, wider access to the oceans. Demand and supply had come together in the free market: the Navy wanted knowledge and information, and the oceanographers needed money and resources for their research. Oreskes grabs the readers' hand on the first page, and then guides them through the intricate world of science and the military industry. Each chapter is based on a chronological case study from the 1930s to the 1970s. In these few decades, American oceanography evolved from the underfunded, local research of some determined scientists into an international, state-funded science enterprise with a well-developed infrastructure. Scientists have enriched our knowledge with a wealth of new results and information, and a significant portion of this has been tied to the use of technological tools and accesses that were essential to the Navy.

When leading oceanographers look back at the history of their discipline after many decades, they tend to deny, or at least undermine, the Navy's role and effect on their research. Oreskes points out, however, that although scientists were free to research many problems, they always had to meet the expectations of the military first. This is why, for example, a novel deep-sea research submarine called Alvin served military purposes in the 1960s for many years, and it took a long time and several attempts to be able to spend military money on basic scientific research (chapter 6). Oreskes cites several similar examples of how the military industry has limited science along the "need-to-know" principle (p. 106 ff.): many results have been encrypted and not made available to colleagues, so numerous research projects have been wasted or delayed for decades.

Occasionally, the simple disinterest of the Navy made certain topics disappear altogether, and thus changed the course of research. Such was the case with climate change. While some oceanographers recognized the role of the oceans in climate change as early as the mid-1950s and 1960s, the Navy was not particularly interested, and the money continued to be spent on researching and exploring radars, submarines, and deep-sea communication channels. By the time the oceanographers were able to breathe a little fresh air and detach themselves from the Navy, they had only managed to join the climate research community somewhat lately. Something similar happened with regard to fishes: at first the Navy was not interested, and when the fauna of the seas had shown its importance for climate, oceanographers were late once again. As Oreskes presents the story overall, it seems that without the money of the Navy, oceanography would have plunged into a crisis; it got the money in the end, but it was not a free lunch.

Science is always funded, and oceanography is no exception. Earlier it was often paid for by inherited money, later from public offerings or by the private sector, then the Navy arrived as a military actor. The question is not, says Oreskes, who will fund the science, but how it will be funded, what they expect in return, and how honestly one can talk about these issues. And to talk about it, we need to see our subject more clearly, that is, as science embedded in social processes. After World War II, for example, many were happy to join oceanography precisely *because* they conceived it as a way of contributing to the Cold War; they could take up the fight against communism, protect their homeland, and thus they were more interested in the operation of radars in a submarine than in going out in small boats for seaweed. During this period, the funding of research by the Navy was not identified as a problematic external influence. For this reason, it has not even occurred to many that there are other issues or different approaches than those proposed by the Navy in the decades following the World War: researchers had already been *socialized* in this. This is a very exciting take on the internal/external influences debate by Oreskes; according to the debate – that goes back most prominently to Robert K. Merton, Imre Lakatos, and many sociologists of science from the 1970s and 1980s – there are factors in the development of a field/theory/discipline that are internal to its cognitive aims and goals, and there are issues and factors that are external to its cognitive business. The former often refers to the logic of justification, while external issues include politics, morality, and societal concerns that are deemed irrelevant (thus external) to the truth-seeking business of science. According to Oreskes, however, "military concerns were naturalized, so the extrinsically motivated became the intrinsically interesting" (p. 502), and thus the external-internal take is getting demolished – a conclusion often anticipated or well-supported by STS scholars.

Of course, we can ask the counterfactual question (as Oreskes points out): what would have happened if the Navy were to leave oceanography alone, if oceanography could have followed its own path, free from the military industry after the 1930s? But this counterfactual question already assumes that science has an *essence*, a necessary course that has been overshadowed by the Navy. Oreskes denies the existence of such an essence; science is what it is, by its own cultivating practice over time. This is what oceanography was, it had to cook from these ingredients. Finally, someone has presented the menu in detail, and if you have the patience and time to eat yourself through the starter and the various main courses, you earn the dessert of finding out that although money may be dirty sometimes (with a lot of insight on *how* dirty is it), we can still talk about it honestly.

I highly recommend this book to all those who are interested in the philosophy and history of a special science. Our perspective and appreciation of science will be highly widened by entering the field of oceans. Oreskes made a huge service to the profession by going over all the details and events. This is one of the strengths, and one of the weaknesses of the book as well. It is unbearably long, five hundred pages with an extra two hundred pages of notes. It is literally too heavy to pick up and start reading every day. However, especially as oceanography is not the typical choice for philosophers and historians of science as an interest, this book deserves our time and engagement.

Science & Technology Studies

Volume 37, Issue 3, 2024

Articles

Discussion paper

Book reviews