Exploring experiments

Guest editors
Andy Pickering
Professor emeritus, University of Exeter, UK

Finn Olesen
Associate professor, Department of Digital Design & Information Studies, Aarhus University

Peter Danholt
Associate professor, Department of Digital Design & Information

DASTS is the primary academic association for STS in Denmark. Its purpose is to develop the quality and breadth of STS research within Denmark, while generating and developing national and international collaboration.
Contents

Author bios

Introduction

Experimenting with experiments
A. Pickering, F. Olesen, P. Danholt

Collection

Virtual Educational Laboratories: Instructive or explorative learning?
S. Lisborg

Experiments in Artificial Sociality
Curious robots, relational configurations, and dances of agency
F. Vejlin

Recasting ethical dilemmas in participatory research as a collective matter of ‘response-ability’
T. Friis

Experimenting on the Enactment of Predictive AI: The Quest for a Future Proactive Healthcare Sector
A. Henriksen, F. Olesen

“We Can Be Pioneers”
– Exploring experimental knowledge sharing in an online peer-support forum for non-offending pedophiles
M. Pedersen
Author Bios

Andrew Pickering is professor emeritus of sociology and philosophy at the University of Exeter. A leading figure in STS, his books include Constructing Quarks, The Mangle of Practice and The Cybernetic Brain: Sketches of Another Future as well as the edited volumes, Science as Practice and Culture and The Mangle in Practice. He is currently completing a book on nonmodern relations with the environment, tentatively entitled Acting With Nature: Doing Without Science.

Anne Henriksen is a PhD fellow at Aarhus University as a part of the Science-Technology-Society Centre. Her research focuses on the development and use of algorithmic systems for decision-making in relation to citizens and government resources and on the ethical and political steering of AI development. Her work has been presented at several international conferences and published in Information, Communication & Society and the Proceedings of the AAI/ACM Conference on AI, Ethics and Society.

Frederik Vejlin is a PhD Fellow in the Department of Anthropology at Aarhus University, working at the intersection of sociocultural anthropology and STS. His dissertation research explores the experimental development and design of social robots in Japanese robotics laboratories, with a particular focus on how culture, science, sociality, and technology blend together in the creation of artificially social entities.

Finn Olesen is Associate Professor, PhD, at Information Studies, Aarhus University. His research is grounded in philosophy of technology and STS studies, and he has made several ethnographic studies of sociotechnical work practices in health care. He has published articles and papers in Danish and international books and journals about sociotechnical practices in health care, posthumanity, and philosophy of technology.

Mikkel Rask Pedersen is an Industrial PhD fellow at Science and Technology Studies, School of Communication and Culture, at the University of Aarhus, working in a joined partnership with the NGO: Save The Children, Denmark. His research focuses on legal, online peer-support forums for people sexually attracted to minors as means to combat the risk of committing child sexual abuse.

Peter Danholt (PhD) is associate professor at Information Studies, Aarhus University. His research is on the role of IT and data in healthcare, organizations, design and daily life. His research draws on STS, feminism, post-structuralism and actor-network theory. He is vice president of the Danish Association of STS (DASTS). He is chief-editor of STS Encounters.

Sanne Lisborg is a PhD fellow at Culture and Learning, Aalborg University (Copenhagen) and Digitalisation in the school, University College Copenhagen (KP). Her current research focuses on digitalisation and learning technologies in compulsory school and teacher education, with a special focus on virtual laboratories and simulations (the topic of her PhD project).

Tine Friis is a PhD fellow at Medical Museion, Department of Public Health, and Novo Nordisk Foundation Center for Basic Metabolic Research (CBMR) at the University of Copenhagen. Her disciplinary background is in the field of Psychology. In her research, she explores personal experiences of how the human gut and psyche entangle and influence the ways in which we understand each other, our bodies, and ourselves—in light of knowledge about gut, brain and psyche. In exploring this, she connects her research to the exhibitions at Medical Museion, investigating methodological potentials in doing so.
Experimenting with experiments - an introduction

Andy Pickering
Professor emeritus, University of Exeter, UK

Finn Olesen
Associate professor, Department of Digital Design & Information Studies, Aarhus University

Peter Danholt
Associate professor, Department of Digital Design & Information Studies, Aarhus University

DASTS is the primary academic association for STS in Denmark. Its purpose is to develop the quality and breadth of STS research within Denmark, while generating and developing national and international collaboration.
Vicky Hearne recounts how she once heard experienced experimenters advising young scientists never to work with cats. I would note, in passing, that it is also strongly discouraged, in laboratories, to work with parrots, not only because they never do anything that is asked of them but because they take advantage of their free time by destroying, with remarkable care, all of the equipment. (Despret et al., 2016: 89)

Contrary to what many think, laboratories are not controlled by brilliant scientists. They are not sites where everything goes according to the white-coated scientist’s plan. Cats and parrots ‘disrupt’ the experiment and literally ruin the laboratory! They are places where scientists are not fully in control of what they produce knowledge about, but are thrown into the messy and risky practice of trying to produce knowledge. A site of material and non-human agency, where the risk is not only misbehaving animals but also that perhaps nothing happens at all. Or even worse, things actually go exactly as anticipated by the scientist, because if this is the case, it might be because the setting has been overly determined and constrained by the scientist, so nothing, but what the scientist imagined could happen. Then nothing new is learned or discovered.

This special issue is dedicated to the exploration of experiments and experimentation. It follows a PhD. course entitled “Exploring and performing experiments” that we organized at Department of Digital Design and Information Studies in spring 2019. The course was attended by 12 PhD fellows, and during the course we and the participants decided to produce a special issue based on the participants’ PhD research projects. The literature for the course included a variety of texts and research articles focusing on experiments mainly from the field of Science and Technology Studies (STS). The readings included the work of Ian Hacking, Andy Pickering, Bruno Latour, Steven Shapin and Simon Schaffer, Isabelle Stengers, Shirley Strum and Brian Eno among others. In the call for papers for this issue authors were asked to draw on the literature in the field of STS in order to explore the role of experiments and experimentation in their own projects, and to consider their articles as vehicles for bringing insights from STS to their own fields. The spirit of this special issue is thus one of ‘STS pollination’ by bringing STS to other fields, rather than necessarily being contributions to STS itself. Hopefully it will generate novel insights and contributions and perhaps cross-pollination.

At the most basic level we might say that experiment is a way of producing knowledge about the world. A way of posing questions and getting answers. But, that said, experiments are not only a way of knowing the world, but also a way of doing in, and acting upon the world in specific ways and, as suggested by Latour in one of his seminal papers on Louis Pasteur and his development of the anthrax vaccine, a way of rearranging society and the world (Latour, 1983).

Experiments and experimentation are an essential part of science and technology development. Experiments can vary greatly from being a detailed, circumscribed practice with a well-defined objective to an open-ended exploration. Indeed, it can be argued, with reference to both Martin Heidegger’s notion of being ‘thrown into the world’ and Alfred North Whitehead’s process philosophy, that existence is experimental through and through, since existence entails ever changing circumstances and unexpected emergence.

In the philosophy of science, experiment is traditionally considered a method by which to test hypotheses about a given object in a controlled, secluded environment. Since Francis Bacon, the experimental method has been the example par excellence of reliable empirical science: Nature must be interpreted through the senses and aided by experiments ‘fit and apposite’ (Goode et al., 1989; Harré, 1981; Latour; 1992; Shapin, 2017). During the scientific revolution laboratory experiment came to be a particular event in which a phenomenon could be secluded, manipulated and observed, a site of “purification” allowing for detailed scrutiny and description that paved the way for a specific scientific perception of reality. Over the years experiments and the laboratory have come to play a central role in the reproducibility
and circulation of knowledge. Given identical instrumental set-ups and procedures, experiments can be repeated and generate identical findings, thus corroborating or validating facts of nature. Laboratory experiments has on that basis gained a reputation in common thought as science in its hardest form.

Experiment as a hypothetico-deductive approach imposes a binary frame – it may either verify or falsify. But experiments can also be sites of discovery where we come to understand an object in a new way and thus expand our knowledge of it. Ideas about hard science as that which lives up to the standards of controlled and reproducible experiments also today play a significant role in public and scientific debates around what qualifies as fac- and trustworthy science. Such discussions are actualized with the advent of public and political debates on climate changes, vaccines, various medical and health issues, immigration etc. and those debates are fueled by how information is now being circulated and mediated via internet technologies and social media. The science wars are thus still raging, it seems, and an emphasis on ‘solid scientific facts’ seems to be what is both in demand and debatable at every point.

In this issue, we do not propose to solve any of those ‘big questions’ or decide on what qualifies as good experiments or research. Rather, the departure is more modest. We allow ourselves to be uncertain. We propose that what an experiment is, or should be, is an open-ended question and an ongoing concern that must be invented and explored in every research project and related to basic concerns such as: what is the object of inquiry? what are its contours, boundaries and how is it demarcated? And by what means and questions is it able to make a separate account that sets it apart from the researcher? As the Belgian philosopher of science Isabelle Stengers would put it: research is a risky and interested practice that works to provide its object the ability to give an account of itself that is not determined by the researcher (Stengers, 2000, 2010).

Although the classical laboratory experiment is considered by many to be science par excellence, empirical studies of laboratory experiments and settings conducted in the field of STS have added importantly to our understanding. Since the 1970’s a performative turn has emerged, away from abstract philosophical and theoretical characteristics of experiment towards practical or pragmatic understandings. Scientific experiments can be grasped through a number of themes: e.g. instrumentation, experiments in written arguments, representations of phenomena, experimentalists versus theorists. Practice-philosophical and STS studies have shown that experiments are sites of hard work, contingency and messiness (Latour and Woolgar 1986; Hacking 1983; Pickering 1995; Knorr Cetina 1999). They are sites in which ‘worlds are raised’, i.e. the knowledge produced through experiments has consequences beyond the confines of the lab. The knowledge-producing capacities of experiments have multiple consequences for society and for everyday human and non-human existence, but experiments are also themselves products of political, social, economic and cultural factors. Furthermore, experiments are events in which ‘dialectics of resistances and accommodation’ occur (Pickering 1995). Experiments are thus incidents where things might explode and mishaps happen, objects resist and escape scrutiny, or simply be available as docile objects that lend themselves to immediate interrogation and ‘discovery’. Following this, experiments can also be sites of violence, in which objects of all sorts (things, humans, animals, plants etc.) are molested, amputated and reduced ‘in the name of science’. A violence and objectification that is not only a reduction of the objects of study, but also consequential for the knowledge produced, perhaps leading to ignorance rather than insight (Despret, 2006; Kleinman & Suryanarayanan, 2012; Stengers, 1997; Strum & Fedigan, 2000).

For numerous reasons, experiment is no longer confined to the laboratory but seems to pervade the world (Blok, 2020; Latour, 1987; Pickering, 2016; Pickering & Guzik, 2008). The experimental organization has become an organizational form. Experiments are part of design practices: design processes that emphasize mutual engagement between human actors and their materials are inherently experimental. One could argue that experiment is pervasive today because of a societal concern with agility and continuous adaptation coupled with a concern
for making well informed ‘evidence-based’ decisions and to avoid haphazard actions. And with climate changes the planetary system of Earth seems to have become a gigantic laboratory with millions of scientists, companies, lay persons and politicians debating what to make of the ‘data’ and how to act.

Experimental practices are embodied situated practices where humans and our conceptual and intellectual ‘tools’ are entangled in complex, dynamic assemblages. Also, we might argue that when we are doing research, in the field or the lab, we are enmeshed in a practice of testing and experimenting with things, phenomena, instruments, methods, concepts, existing knowledge, interpretations and their various combinations in order to produce knowledge and do stuff with what we come to know—account for it, publish it, defend and negotiate concepts, empirical material and interpretations, get grants, jobs and credentials. Experimentation is the continuous unavoidable practice of being in a world that is never stable and continues to escape our control.

Not anyone can conduct a specific experiment and replicate established facts, and even when skilled experimenters go about their business, things still go wrong: equipment and instruments fail, things explode or nothing happens, or cats refuse to do what they are supposed to and parrots – with remarkable care – destroy everything around them. When things do not go as expected, what do experimenters do? Discard some of the incidents as anomalies and mishaps? Or meticulously account for every ‘nonsensical’ result, every mishap, happening and unhappening as rightfully adhering to the experimental context as a good and observant scientist? Hopefully, the latter, but all those happenings rarely make a good publishable article, so some things have to be excluded, but which parts and on what grounds? And when incidents and results get excluded where does that leave us if we are firm believers in laboratory experiment as hard science and scientists as value-free neutral observers? Experiment involves the risk of making the wrong or bad seclusions and thus conducting, not science par excellence, but just poor research...

This issue is comprised of five articles that in various ways address or experiment with experiment in relation to specific fields or studies. Sanne Lisborg is concerned with the role of virtual laboratories in school teaching. Virtual labs are computer simulations of ‘real experiments’ which are used to teach pupils about, for instance, physics or biology. Lisborg analyses and explores how we may conceptualize the virtual laboratory by drawing on STS and pragmatism.

In his article Frederik Vejlin presents his ethnographic field work in a social robotics lab in Japan. In the lab, the roboticists experiment with how to make robots social, that is, how to design and program robots in order for them to interact with and be experienced as social by human beings. Vejlin draws on recent social anthropology and shows that how and what ‘social’ means is extremely difficult to define and accordingly to reproduce. However, the ambition and assumptions about the social as a specific human quality, are central to work in and around the social roboticists’ lab.

Tine Friis article explores the relation between the gut and the psyche. Her article presents and discusses how patients diagnosed with gut-related diseases together explore the relations between their gut and their psyche through writing third-person stories and group exchanges. She understands ‘memory work’ as itself an experimental exploration of the self.

Anne Henriksen and Finn Olesen present a study of a company experimenting to develop a system based on AI and Machine Learning to be used in relation to predicting somatic patients at risk. Henriksen and Olesen draw on Latour’s concept of partially existing objects and show how the system comes into being and changes in relation to organizational and design decisions made along the way. Andy Pickering’s notion of ‘social mangling’ is used to emphasize this point.

Mikkel Rask Pedersen’s article is based on his research on online peer-support forums for pedophiles who seek to live non-offending lives. He draws on, among others, Donna Haraway’s ‘situated knowledge’, Marilyn Strathern’s work on relations, hybrids and networks and Pickering’s concept of ‘experimenting in the wild’. And he shows
how we can understand the peer-support forum as an arena in which pedophiles experiment with becoming different subjects, and how these practices can be valuable for child sexual abuse prevention.

Acknowledgements
The authors would like to thank the PhD school at Arts, Aarhus University, for economical support. Kasper Ostrowski for editorial support and design. All of the participants attending the course and the three anonymous reviewers for thorough readings and commentaries of the articles.

References


Despret, V., Buchanan, B., & Latour, B. (2016). What would animals say if we asked the right questions? University of Minnesota Press.


Virtual Educational Laboratories: Instructive or explorative learning?

Sanne Lisborg
PhD fellow, University College Copenhagen and Aalborg University

DASTS is the primary academic association for STS in Denmark. Its purpose is to develop the quality and breadth of STS research within Denmark, while generating and developing national and international collaboration.
Abstract

This paper explores how the field of Science and Technology Studies (STS) can inform and help conceptualise a relatively new form of laboratory work in education: virtual laboratories. To date, STS have not addressed laboratory work in education. This paper focuses on the virtual educational laboratory by synthesising arguments from the STS literature on laboratory work and proposes research questions that can guide future ethnographic research on how virtual laboratories are applied and constructed locally in the classroom. I argue that the virtual laboratory, like the physical one, must be understood in a broader cultural, social and material context. Moreover, the virtual laboratory has both constraints and affordances, tied to the medium through which it is materialised. I conclude that the virtual laboratory can be understood as a hybrid between explorative and instructive learning.

Introduction

Science and technology studies (STS) have been occupied for a long time with exploring how scientific facts and knowledge are produced in laboratories and how these affects and transforms society (e.g., Knorr-Cetina, 1992; Latour & Woolgar, 1979). For example, laboratory studies have been instrumental in demonstrating that cultural, social, and material factors are substantial for stabilising and producing scientific facts. Scientific knowledge is a product of a complex network of relations and artefacts, not natural ‘givens’ discovered by science (Latour, 1978; Pickering, 1995). This insight makes it relevant to study the production of knowledge concerning the laboratory. But STS have not explored the role of the laboratory in educational settings, which this paper addresses, and, more specifically, the use of virtual laboratories in education. By synthesising arguments from the STS literature on scientific laboratories, I suggest that STS can provide nuance and broaden the understanding of virtual laboratories in education.

Over the past years, a new kind of laboratory has emerged in education: virtual laboratories. Virtual laboratories are computer simulations in which pupils can explore and interactively engage with subject areas related to STEM in a laboratory environment. These interactive simulations enable the pupils to actively change or define different parameters and observe the consequences of their actions. Virtual laboratories are primarily being used in universities and upper secondary schools (Lewis, 2014; Achuthan, 2018) but are used in lower secondary schools as well (Implement, 2018). This paper concentrates on the educational setting of lower secondary schools in Denmark, which is the empirical focus of my research. I have performed field work at three schools, where I have done video-based observations, screencast the pupils while they work with virtual laboratories and interviewed pupils and teachers.

There has been a political push to implement virtual laboratories in Denmark as a part of science teaching in lower and upper secondary...
Virtual Educational Laboratories

S. Lisborg: Virtual Educational Laboratories

schools. The 'Action Plan for Technology in Teaching', published in 2018 by the Ministry of Education, states, 'Virtual simulations can increase pupil's motivation and learning in science. They can supplement the traditional (physical) experiments' (Undervisningsministeriet, 2018, p. 19). This goal is closely connected to a STEM agenda to increase the pupil's motivation and learning in science teaching. The Ministry of Children and Education has launched different initiatives to stimulate virtual laboratories in schools. One initiative involves free access to a selection of virtual laboratories from the Danish company Labster from May 2019 to June 2020. Furthermore, a development project with ten schools was launched in 2019 to experiment with and develop different didactic and technical approaches to integrate virtual laboratories in science teaching (Børne- og Undervisningsministeriet, 2019).

As stated above, there are different political expectations and aims regarding the use of virtual laboratories in education. STS scholar Estrid Sørensen (2009) argues that research should not focus on what we would like learning technologies to do, rather on how technologies form and change teaching practice in often unexpected manners. New learning technologies must be understood in the situated, social and material contexts in which they are embedded and should be studied ethnographically. The research field of virtual educational laboratories is primarily quantitatively oriented, e.g. it uses pre- and post-test or control groups. These studies focus on different matters, such as the impact that virtual laboratories have on the learning outcome or motivation, for example, in comparison to traditional teaching (Dyrberg et al., 2017; Makransky et al., 2016; Smetana and Bell, 2012; Vogel et al., 2006). This paper makes the case that there is a need for an ethnographically and STS-inspired approach to studying virtual educational laboratories to understand which situated educational practices come about.

In the first part of the paper, I propose distinguishing between four kinds of laboratories to conceptualise the understanding of laboratory work and highlight my contribution to STS. First, I argue that there is a significant difference between the laboratories used for scientific purposes versus those used for educational purposes. While the first is concerned with the production of knowledge, the second is concerned with replicating knowledge for learning purposes. When considering virtual laboratories in education, how knowledge is replicated becomes important. Inspired by learning theory, I suggest a distinction between a more instructive approach to learning and a more explorative one. The distinction is used to discuss and conceptualise which kind of learning is enacted with a virtual laboratory. Second, I distinguish between virtual and physical laboratories. While STS have traditionally focused on the physical laboratory, I pay attention to the virtual laboratory.

In the second part of the paper, I draw on and synthesise perspectives from the STS literature on laboratory work and experimentation and examine how these insights can propose research questions for further ethnographical research. First, I argue that the virtual laboratory, like the physical one, is a part of a broader socio-material network, such as the materiality of the classroom, interactions among pupils and teachers, teaching styles, etc. It is crucial to shed light on these networks to understand which learning situations are constructed with the virtual laboratory. Second, based on STS studies on computer models and simulations, I argue that one must pay attention to the role of the medium that virtual laboratories work through, i.e. the computer, and to which affordances and constraints the medium brings about. Third, the STS literature has demonstrated that mangling, mishaps, contingency and tacit knowledge are essential elements of experimental and scientific practice. All the hands-on mangling practice is left out in the virtual laboratory, and the knowledge is conceived utilising a simulated laboratory. It is vital to examine how the above-mentioned aspects of the physical laboratory work are affected when the experimental practice is moved to a virtual laboratory. Towards the end of the paper, I argue that the virtual laboratory can be conceptualised as a hybrid between explorative and instructive learning.
Four ways of distinguishing between laboratories

To nuance and specify the understanding of laboratory work, I make some distinctions to relate my contribution to the field of STS and laboratory studies. The first distinction is between the laboratories used for scientific purposes and those used for educational purposes. The fields of STS and laboratory studies have been engaged with scientific laboratory work and scientific knowledge production (Sismondo, 2010). However, they have not paid attention to laboratory work in an educational setting. A central distinction between these two kinds of laboratories is that scientific laboratory work is concerned with producing new knowledge. In contrast, laboratories in an educational setting are designed to reproduce established knowledge. This distinction is inspired by Estrid Sørensen, who argues that there is an essential difference between Bruno Latour’s understanding of scientific and educational knowledge. Latour argues that knowledge is produced based on circulating references, such as diagrams, maps and samples, where this circulation establishes it as factual (Latour, 1987). Here, scientific knowledge is new because it is not yet established and must circulate to accomplish the quality of being knowledge (Latour, 1999). Hence, scientific knowledge is new not only to the scientist but also to the science community and society.

In school, knowledge is already established as knowledge and is only new to the learner. The references that make up the knowledge in school are already accepted as valid, and learning is achieved when pupils connect to these references (Sørensen, 2009). The knowledge produced in and around laboratory work in an educational setting is connected to certain representations of knowledge and understandings of learning. In the educational laboratory, the purpose is to introduce the learner to established knowledge within teaching plans, curricula and other learning activities. This goal makes it relevant to investigate which learning situation the virtual laboratory creates and is embedded in to understand which kind of learning practice is enacted. It is important to note that the distinction between the two kinds of laboratories is not as strict as stated above. For example, some scientific laboratories are used for educational purposes, and, sometimes, established knowledge is reproduced in the scientific laboratory. Similarly, in educational laboratories, especially in higher education, new knowledge can be discovered and contributed to the scientific community. But the distinction is useful for defining and conceptualising the characteristics of knowledge production in education and science, respectively.

The second distinction is between virtual laboratories and physical laboratories. The STS literature has been occupied, for the most part, with work conducted in a physical laboratory. Some scholars have addressed the entry of ‘dry’ 2 or ‘virtual’ laboratories in science and how this new laboratory setting affects the production of knowledge (Merz, 2006). However, the field of STS has not paid attention to the virtual laboratory in an educational context. The distinction between the two laboratories is not clear-cut. There are different ‘virtual’ or technological elements in most physical laboratories, including simulations, apparatuses and computer visualisations. The virtual laboratory is physically embedded in a learning situation and the interpersonal dynamics between pupils and teachers. But the central point is that the STS literature has focused the most on the knowledge production embedded in a physical laboratory and has paid less attention to the virtual laboratory. In other words, four central distinctions or quadrants can be constructed to conceptualise the different laboratories referred to in this paper, which are illustrated in the diagram below.

---

2 Dry labs are laboratories in which computers are used for data collection, modeling and simulation, in contrast to wet labs, in which various liquids and chemicals are tested and analysed.
S. Lisborg: Virtual Educational Laboratories

While the field of STS has been occupied with the right side of the diagram, i.e. physical and virtual laboratories in science, the left side of the diagram, i.e. the virtual and physical laboratories in education, have been ignored. This paper aims to shed light on the upper left quadrant, virtual educational laboratories.

The virtual laboratory in education

Virtual laboratories are computer simulations, which, as the name suggests, take place in a virtual context and are designed to give the user the experience of performing experiments in a laboratory setting (Jones, 2018). Virtual laboratories are a part of a broader field of computer simulations in science education. These simulations are defined as computational models of a physical phenomenon, where the user can change different parameters to observe the consequences of their actions (Clark et al., 2009). Simulations differ from static visualisation (e.g., an illustration in a textbook) because they are interactive (Plass et al., 2009). There are different nuances in the definitions in the research literature on interactive simulations, but a common thread is that it is not enough for pupils to observe a simulation; they must engage with it to achieve a better learning outcome (Stoney & Wild, 1998). The user must be able to choose or define actions in the simulation and observe the created sequence (Vogel et al., 2006).

However, how interactive the virtual laboratories are varies a great deal. Some simulations give the user concrete instructions, while others are more openly designed and call for a freer inquiry. Some simulations even allow the user to access the software system and remodel the simulation with an easy-to-use programming language (Honey & Hilton, 2011). In other words, the element of interactivity is broadly interpreted in practice. Furthermore, some are simple 2D visualisations, while others are 3D game-like simulations. In certain virtual laboratories, the emphasis on immersive and game-based elements is prioritised to create engagement and motivation (Jones, 2018). To illustrate the diversity in the field, I focus on two virtual laboratories used in science education: Physics Education Technology (PhET) and Labster. These are also the objects of my ethnographic study of science teaching with virtual laboratories in lower secondary school.

The University of Colorado Boulder developed PhET. It offers a large online library of virtual interactive simulations that are freely available for science education. The simulations are designed so that pupils and teachers can use them with minimal prior training (Honey & Hilton, 2011). Each simulation is designed to target a core science concept, such as wave physics. The pupils can create their own waves, watch how they interact in various ways, and learn about the basics of wave physics. The amount of freedom pupils have varies within the simulations. While some are quite open and allow the user to explore more freely, others

Diagram 1: Laboratories and the field of STS

Note. The diagram is visualised as a continuum where the arrows on the x-axis and y-axis indicate that a laboratory can be more or less on one side of the diagram. Examples of STS scholars focused on physical and virtual scientific laboratories have been inserted in the diagram.

While the field of STS has been occupied with the right side of the diagram, i.e. physical and virtual laboratories in science, the left side of the diagram, i.e. the virtual and physical laboratories in education, have been ignored. This paper aims to shed light on the upper left quadrant, virtual educational laboratories.

2 STS scholars, such as Fenwick and Edwards (2010), have studied education and learning through an Actor–Network Theory (ANT) lens. Further, Sørensen (2009) has focused on the digital technologies, such as the online 3D virtual environment and weblogs, in education from a socio-material perspective. But these scholars have not addressed virtual educational laboratories.
do not encourage much free exploration. The simulations have simple graphics and no specific focus on gamification elements (Jones, 2018).

Labster was developed by a Danish company by the same name. It was primarily developed for upper secondary schools and universities but is also used in lower secondary schools. Like PhET, a large selection of virtual laboratories is accessible online, but schools need to buy a license to get access. There is a strong focus in Labster simulations on immersive or game-based elements, such as a storyline that the simulation is built around and a mission that the pupils need to fulfil to make it more engaging. The user can ‘walk around’ the laboratory, and when a procedure involves a physical act, they must be mimicked by the user, such as picking up a pipette, using it, putting it back, etc. Unlike PhET, there is a strong focus in Labster on making the user feel as though they are in a real laboratory (Jones, 2018).

The Labster simulations also include multiple-choice questions, and the pupil can read the associated theory to answer the questions. There is a scoring system in which the pupil gets a final score depending on how well he or she answers the questions. Labster simulations are quite story-boarded or instructive: the user is often guided through the technical steps for carrying out laboratory procedures by a voice-over. There are also experiments where the user can change the variables and experiment more freely, but the learning is still more guided than in PhET, which often allows for a high degree of freedom. In contrast, Labster simulations have a strong focus on creating an authentic laboratory environment, game-based elements and context around the experimental practice, which is not the focus of PhET.

Laboratory studies have not been occupied with learning since it is not the focus of the knowledge production conducted in scientific laboratories. But the notion of learning becomes a key focus when studying virtual educational laboratories, where the technology is intended to enhance and support learning. In order to conceptualise the virtual educational laboratory, the diagram below summarises four learning modes that can be said to characterise the virtual laboratory. I have inserted examples of the virtual educational laboratories used in upper secondary schools in Denmark. The diagram is visualised as a continuum where the arrows on the x-axis and y-axis indicate that the different virtual laboratories can be placed more or less on one
Explorative and instructive learning

To develop my understanding of explorative learning, I draw on perspectives from constructivist learning theory. Philosopher of education John Dewey espouses the idea of learning as an affair of interacting with the world, as captured by Dewey's famous 'learning by doing' principle. By doing things, Dewey (1916) argues, we experience the consequences of our actions. One cannot know the world without being an actor in it, and it is through acting in the world, one obtains new experiences. In the Deweyan perspective, experience is understood as bodily and sensory experiences of the world in which one must interact directly with objects in practice (Brinkmann & Tanggaard, 2012). Because learning is essentially experience-based, teaching must focus on the pupil's active investigation and experimentation. Dewey divides experience-based learning into four stages: 1) sensing a problem, by being in doubt or frustrated, and starting to think about how to change it; 2) investigating the problem and formulating solutions; 3) reaching a conclusion based on the investigation; 4) testing and experimenting with the new knowledge. Learning, in this sense, is about active construction and participation instead of a passive transfer of knowledge (Beck et al., 2014). This focus on experiencing and doing is also the core concept of the central didactics used in science education today, such as 'inquiry-based science'. In inquiry-based science, the pupil explores and develops explanations for the phenomenon under investigation and evaluates their understandings by experimenting and observing (Gillies & Nichols, 2014).

Another source of inspiration comes from the mathematician and computer scientist Seymour Papert, who, in line with Dewey, puts active doing at the centre of learning. Papert focuses on the role of technology, especially the computer, as an important artefact for creation and participation. He sees computer technology as an effective and useful tool for constructing meaningful products (the 'doing' element) because computer technology has opened up new possibilities for creation (Papert, 1980), such as experimentation in a...
virtual laboratory. Explorative learning is now understood as a learning mode that focuses on the experience-based, active and free exploration and investigation of a scientific phenomenon. For example, this is the case when pupils in a virtual laboratory can experiment freely, observe the consequences of their actions and test hypotheses. In this regard, the focus is on open-ended experimentation, and the ‘doing’ element is highlighted here.

The term instructive learning is inspired by the behaviourist theories of learning. In this context, learning is regarded as a causal connection between stimuli and response, where the learner’s actions will be corrected or rewarded according to the correct response (Selwyn, 2011). The psychologist B.F. Skinner developed what he called ‘teaching machines’, which should engage the pupils and give them immediate responses. For Skinner (1958), the learning process is divided into several small steps, and the pupil reviews the feedback received after each step. This behaviouristic approach has played an important role in developing computer-assisted instructions (Selwyn, 2011). In relation to virtual laboratories, this mode of learning becomes particularly visible in step-by-step guidance, where the pupils are guided through the different steps in the experiment without much individual freedom. This kind of experimenting is more controlled, and there are ‘rules’ for what is desired or possible to do. Here, the focus is more on the transmission or acquisition of certain laboratory skills or procedures.

The divide between explorative and instructive learning is an analytic distinction used to illustrate a spectrum of learning modes that can help discuss which kind of learning virtual laboratories bring about. However, it is important to note that the learning process is far more complex than expressed by this divide. Different socio-material factors affect the learning situation, which will be elaborated upon in the paper. In the next section, I address what we know about laboratory work from STS and how these findings can help us conceptualise and nuance the understanding of virtual laboratories in education.

How is knowledge produced in a laboratory?

Early laboratory studies and the sociology of scientific knowledge (SSK) ask the central question: how are scientific facts made, and how are they made stable? The answer to this question is that the production of knowledge is essentially carried out socially and culturally. Scientific facts are not natural givens discovered by science but, rather, products of social dynamics. As Sismondo (2010) writes, ‘(...) decisions about claims are negotiated, a matter for different actors to decide through micro-sociological or political interactions’ (p. 107). This point means that conversations and negotiations between scientists and other actors decide what is being established and communicated as scientific knowledge. What is stated as scientific claims cannot be predicted in advance, and the laboratory work must be studied through a cultural and social lens. In this sense, laboratory work and the production of scientific knowledge are no different from other interactions in our everyday life: they are just as messy and complex as any other social interaction (Sismondo, 2010).

In the second wave of laboratory studies represented by ANT, the material perspective of laboratory work is introduced. As Latour and Woolgar (1979) demonstrate in Laboratory Life, the laboratory becomes an important source for the construction of facts since the whole material setup in the laboratory enables the scientist to manipulate and arrange natural objects in certain ways. It is the social and cultural factors that are interesting in the construction of science and the equipment and technical instruments used in the laboratory, i.e. the non-human actors. As per this perspective, scientific claims are negotiations between different actors: materials, techniques, technologies, scientists, and others (Knorr-Cetina, 1992). The laboratory is an association machine that establishes connections between different entities.

Latour and Woolgar (1979) illustrate how natural objects and phenomena are translated into figures, numbers, graphs, etc., via inscription devices, which are the different instruments and apparatuses
used to produce data about the object and phenomena in question. Furthermore, these inscriptions are translated into scientific claims. Laboratory studies show a discrepancy between what the textbooks say makes science valid and how science works in practice. As Merz (2006) writes, ‘In this perspective, knowledge production is closely associated with the laboratory as a site of locally embedded practice’ (p. 157). These studies make it scientifically relevant to shed light on how the order is produced out of the seemingly messy network of things and actions (Latour and Woolgar, 1979; Pickering, 1995). Which learning situations unfold around virtual laboratories should therefore be empirically questioned. The distinction between entities cannot be made a priori (Latour, 1987) but must be made based on the connections between human and non-human actors constructed locally in the classroom. How this plays out in a context with computer simulations will be further elaborated below.

The rise of dry or virtual laboratories

In recent years, a new kind of laboratory has emerged in the production of scientific knowledge: ‘dry labs’ and ‘virtual laboratories’. The field of STS has raised the questions ‘How do computer simulations influence the production of scientific knowledge?’ and ‘Does the computer simulation bring about a new way of doing science?’ (Knuuttila et al., 2006; Merz, 2006). One response to these questions is provided by Merz (2006), who investigates the use of computer simulators in particle physics. She argues that scientific practice in a virtual laboratory is just as messy and complex as work in physical laboratories. The simulation activities are embedded in a wider material and social network and are entangled with ‘other endeavours in experimental physics, theoretical physics, or other fields of practice; the relations, rankings, and hierarchies of different data types; and so forth’ (Merz, 2006, p. 166).

Thus, instead of understanding the virtual laboratory as a delineated practice, it must be understood as being intertwined in a broader network of things, understandings and actions. The learning situations that will be created cannot be predicted by reading the teacher guidelines or lesson plans. The use of virtual educational laboratories must be studied as being embedded in a broader sociocultural and material context. When studying these laboratories, we need to ask how the sociality and local cultures in the classroom, such as teaching styles and interactions among pupils and teachers, affect and form the use of virtual laboratories. For example, the framework and guidelines provided by the teacher play a central role in a concrete learning situation. In my pilot study, the teachers provided different guidelines for how pupils should interact with the virtual laboratory. These included providing questions related to the simulation, letting the pupils work in pairs and participating in follow-up discussions in class. In this way, the teacher counteracts some of the more individualised and instructive elements when working with virtual laboratories in education. In other words, technology makes some understandings of learning easier or more difficult than others. However, it is essential to investigate the situated empirical context around the use of virtual laboratories. Moreover, a material perspective must be adapted to investigate how learning situations are constructed in a network of both human and non-human actors. A question to ask from a material perspective is how the technology brings about a certain ordering of things that affects the learning situation (the role of the medium), which will be discussed in the next paragraph.

Affordances and constraints of the virtual laboratory

STS scholars Knuuttila and Voutilanien (2003) have examined the use of computer models in science. They argue that the models must be understood as material artefacts and not just as an abstraction or concept. They define the models as ‘epistemic artefacts’: that is, ‘as intentionally constructed things that are materialised in some medium
S. Lisborg: Virtual Educational Laboratories

and used in our epistemic endeavours in a multitude of ways’ (Knuuttila, 2005, p. 1266). It is especially important to consider the media through which models are materialised and pay attention to their affordances and constraints, which are often tied to the medium through which they work. As materialised things, models are not open to all possible uses and understandings.

Furthermore, a model is not representative in and of itself but will be interpreted and unfolded in multiple ways when implied in a specific context. This perspective allows actors to hold different conceptions of the same artefact and serve multiple purposes (Knuuttila, 2005). Like computer models, virtual laboratories are tied to the computer and representation through computerised algorithms (the medium). Consequently, virtual laboratories will always be constrained by how they are programmed and designed. The virtual laboratory is configured and programmed according to existing knowledge or theory associated with the field. There are rules or limitations for actions in a simulated world, and exploration done in a virtual laboratory will never be as free as that in a physical laboratory, where the range of possible outcomes are endless. Therefore, one could argue that the virtual laboratory work cannot be fully explorative: it will always be instructive or controlled to some extent due to the medium through which it is materialised.

Virtual laboratories can be compared with flight simulators to illustrate this perspective. Simulations in aviation have been used for a long time since the novice pilot needs a safe environment to learn how to fly without putting himself or others at risk (Aebersold, 2016). The flight simulator is not designed to accumulate new knowledge but to imitate the real experience of piloting an aeroplane. Here, the pilot can learn from his mistakes without real danger and repeat the simulation until he is good enough to fly a real aeroplane. But it is when the pilot flies in a real aeroplane, he will discover new knowledge that is not a part of existing theory or perception of the world.

As with pilot training, the use of virtual laboratories provides the opportunity for pupils to interact with processes that otherwise would be too dangerous or impossible in the physical laboratory. Virtual laboratories in education are often used as a supplement or a rehearsal for physical experiments and a safe training environment in which the pupils can learn from their mistakes and refine their skills (Honey & Hilton, 2011). But the medium implies an epistemological problem. On the one hand, the virtual laboratory gives the pupils a sense of hands-on experience by simulating the experience of being in a laboratory, mimicking the working procedures, interacting with materials and apparatuses, etc.

On the other hand, the practice in the virtual laboratory is essentially not a hands-on experience, but a simulation of a reality wherein the pupil’s actions are mediated by a computer. In my pilot study, the pupils experienced or sensed that the physical interaction with the materials was essential to their learning; as one pupil put it, ‘It is easier to incorporate when it is physical (...). I think that it matters that you are actually doing it and not just clicking with this one [the mouse], and then it falls down in this one (...), but it is something I do myself’ (Interview Anna, 26.08.2020). The direct interaction with objects in practice is essential to experienced-based learning, as described by Dewey, which can be said to be problematised with virtual laboratories. So, the medium (the computer) possesses both essential affordances, such as a safe training environment, and constraints and limitations to possible actions and discoveries. In the next section, I turn to the STS literature on the epistemology of experimental practice and how these perspectives can inform and conceptualise our understanding of virtual experiments in education.

The (unnatural) experimental practice

There is a long tradition in STS of dealing with the relationship between experiments and laboratories (Knorr-Cetina, 1992; Sismondo, 2010). Shapin and Schaffer (1985) unfold the concept and role of experimental
practice in modern science. They investigate the controversy between chemist Robert Boyle and philosopher Thomas Hobbes over Boyle’s experiment with an air pump in the seventeenth century. Boyle’s experimental practices can be regarded as the beginning of modern science, where phenomena are taken out of their natural environment and isolated in a laboratory context: ‘proper natural philosophical knowledge should be generated through experiment and the foundations of such knowledge were to be constituted by experimentally produced matters of fact’ (Shapin & Shaffer, 1985, p. 22). By performing experiments, it becomes possible to replicate and validate the results. Hobbes was very sceptical about this practice and denied that experiments done in an artificial setting, such as in a laboratory, could lead to scientific facts. Despite Hobbes’s opposition, Boyle’s experimental paradigm eventually became the beginning of modern laboratory experimental practice. Experiments came to be seen as a phenomenon where natural objects could be isolated, manipulated and observed.

Knorr-Cetina (1992) unfolds this relationship between the experiment and the laboratory. STS has shown that the experimental practice in the laboratory ‘rests upon the malleability of natural objects (…)’. In fact, laboratories rarely work with objects as they occur in nature’ (p. 116). In the laboratory, scientists work with for instance visualisations of objects, isolated components and extractions – in other words, purified versions. Knorr-Cetina identifies three central features of natural objects that are changed when working in the laboratory. First, laboratory science does not need to work with the object as it is. Second, the scientist does not need to work with the object where it is (i.e., in nature) but can bring it home to the laboratory context. Third, the scientist does not need to attend an event when it happens; they can speed up a natural process or make it happen more frequently (Knorr-Cetina, 1992). The epistemological advantage of the experiment is that it does not need to accommodate objects within a natural order and the scientific phenomenon can be investigated in a controlled and manipulated laboratory setting.

Enhanced purification and manipulation?

With these characteristics of the experiment in mind, it is time to discuss experimental practice in the school. In a school setting, one could argue that the field studies related to observing, for example, insects or fauna in nature, are as close as we get to Hobbes’s understanding of investigating scientific phenomena in their natural environment. But most scientific practice in schools is performed as experiments in laboratory settings, where the pupils conduct small experiments in groups (Hodson, 1990). Here, the scope of the experimental practice is to (re)produce scientific facts or established knowledge in a controlled environment or, as Knorr-Cetina (1992) puts it, using a purified and malleable version of the natural object.

In the virtual laboratory, one could argue that natural objects become even more purified since the objects are not physical but virtual representations. When the experimental practice occurs in a virtual and programmed setting, the reality can be adapted to specific learning outcomes. Virtual laboratories can be designed to simplify learning by removing confusing details and focusing on core elements (De Jong et al., 2013). As illustrated earlier, as a computer-programmed object, the virtual laboratory enables a learning setting that is safe and designed for specific learning outcomes (affordances). However, it also limits the range of actions and possible outcomes (constraints).

Furthermore, the virtual laboratory offers efficiency since processes can be sped up without authentic delays, and the pupils can experience photosynthesis with one click of the mouse. In other words, restrictions of time and space are abolished in the virtual laboratory. As we can see, it becomes empirically relevant to investigate how the features of the virtual laboratory affect the learning situation in multiple ways. In the next section, I unfold the concept of tacit and embodied knowledge and discuss how this kind of knowledge might be affected by experimentation in a virtual setting.
Tinkering and tacit knowledge

Even though experiments are set in a manipulated and ‘designed’ environment, laboratory work is not as smooth as one might expect. Hacking (1983) argues that ‘laboratory work is not merely about representation but about invention’ (Sismondo, 2010, p. 108). The scientist is actively engaged in manipulating the object because materials do not behave as one expects, apparatuses do not work, etc. Thus, laboratory work is indeed about tinkering (Knorr-Cetina, 1981) or bricolage (Latour & Woolgar, 1979) in order to make disobedient materials act how we want them to act (Sismondo, 2010). Here, the embodied and tacit knowledge becomes important to create successful scientific procedures (Knorr-Cetina, 1992). Collins (1974) argues that not all knowledge travels easily. In his study on the transfer of knowledge about building a TEA laser, Collins shows that no scientist could build the laser by only following a manual: the passing on of skills from an experienced informant is essential to building a functioning laser. He argues that some knowledge is essentially tacit. The transfer of knowledge requires a socialisation process (or is enculturated) in opposition to ‘algorithmic’ knowledge that can be formalised and standardised and, thereby, travels more easily (Sismondo, 2010).

In physical laboratories, the pupils interact with real materials, apparatuses and chemicals, and the tinkering component is essential to the outcome of the experiment. For example, it can be crucial if the pupil adds a little too much acid, heats something for too long, etc. Here, pupils encounter real errors, failures and unexpected results since they cannot just redo it, as is the case in the virtual laboratory (Tho & Yeung, 2018). The pupils also articulated this lack of real errors and mishaps in the pilot study. They claim that they are more focused when doing experiments in a physical laboratory than in a virtual one since their actions are irreversible and have real consequences. As one of the pupils said, ‘(...) In here [in the virtual laboratory] you know you cannot do it wrong, but in there [the physical laboratory] you need to focus in another way, because you do not want to spill sulphuric acid all over’ (Interview Astrid and Simone, 20.02.20). This situation makes it interesting to further investigate how the lack of real errors and irreversible consequences affect the learning situation.

Moreover, in physical experiments, pupils are faced with unexpected events, such as measurement errors, whereas ‘in virtual laboratories, students are not distracted by aberrations in the equipment or unanticipated consequences’ (De Jong et al., 2013, p. 306). The tinkering element is essential for acquiring a practical knowledge of how materials and objects behave in unexpected ways. But even though pupils in the virtual laboratory interact not with physical materials but virtual representations, one could assume that there is just as much ‘tinkering’ going on in virtual laboratory as in the physical one; it is just another kind of tinkering. When pupils are working with virtual laboratories, they interact with the medium, such as the computer, the mouse and so on. Furthermore, they navigate through the laboratory mediated by the medium. In this sense, the pupils interact with the virtual laboratory in a double sense – both in the laboratory and with the laboratory (i.e. the technology that allows it to function). When investigating educational practices involving virtual laboratories, it becomes interesting to explore how embodied, or tacit knowledge becomes important for creating successful learning situations. Since knowledge cannot be transferred easily, what kind of tools, guidelines, skills and social interactions are needed to allow knowledge to travel into the classroom and beyond when working with virtual laboratories?

Discussion and concluding remarks

This paper offers an investigation into the virtual educational laboratory based on perspectives from the STS literature on laboratory work and my ethnographic pilot study. In contrast to virtual scientific laboratories, I have argued that virtual educational laboratories are concerned with established knowledge instead of producing new knowledge. In education, the focus is on learning already established knowledge
with specific learning goals. Therefore, the focus shifts from a scientific production of knowledge to a learning perspective. Moreover, I point to a continuum between explorative and instructive learning, which is inspired by the constructivist and behaviourist learning theories. This distinction allows for a discussion of which learning modes are constructed and made possible with virtual laboratories. Even though virtual laboratories are shaped by different design principles and understandings of learning, as illustrated using Labster and PhET, there are some shared characteristics of the virtual laboratory.

Based on insights from STS studies, the paper makes the case that virtual laboratories, like the physical ones, must be understood as being intertwined with a broader network of social and material relations. We have learned from laboratory studies that actual knowledge production is always far messier and more complex than what the textbooks tell us. This insight also applies to the study of virtual laboratories in education, where learning must be understood in the situated, social and material contexts in which the learning practice is embedded, such as lesson plans, curricula, teaching styles, the materiality of the classroom and so on.

Moreover, STS scholars Knuuttila and Voutilanien (2003) argue that models must be understood as intentionally constructed things that can be used and understood in multiple ways. It is especially important to pay attention to the medium through which they are materialised and which affordances and constraints are tied to the medium. When studying virtual laboratories, we need to investigate how the medium through which it is materialised, i.e. the computer, implies some constraints and affordances. On the one hand, the computer can expand the scope of actions in science teaching. It allows pupils to conduct experiments and take actions that would not be possible in real life because they are too dangerous, expensive or just impossible to do so. This point is often stated as one of the great advantages of virtual laboratories (Honey & Hilton, 2011). For example, pupils can take a sample from Mars and go back and analyse it in the virtual laboratory, or they can get close to micro-biological processes that are not possible to observe in real life.

Moreover, the virtual laboratory offers efficiency since natural processes can be sped up, and the restrictions of time and space are abolished. In this sense, the virtual laboratory enables the exploration of problems and concepts in a way that would not have been possible in the physical laboratory. It also becomes more effective to do experiments in the class since the virtual laboratory does not require the pupils to find the materials, set up the experiment, etc. Furthermore, the virtual laboratory makes it easier to work with specific and delineated learning objectives because it is possible in a virtual environment to isolate a part of a natural phenomenon or work with specific elements of an experiment. In this context, learning can be targeted, and some more complex matters can be simplified to be made more understandable for pupils.

But there are also some constraints related to virtual laboratories that need to be addressed. When pupils are conducting experiments in a virtual setting, they encounter more purified versions of the objects. They do not experience the contingency, mishaps and errors – the tinkering element – that is essential to get a practical understanding of how materials, objects and apparatuses behave in a physical laboratory, as STS studies have demonstrated. In the virtual laboratory, pupils do not experience the lack of smoothness of the physical experiment, i.e. when the apparatus suddenly does not work, or a tube is too small and so on. One could argue that the learning with virtual laboratories become too smooth: the pupils do not experience the valuable feeling of frustration and the process of understanding why things are not working as expected.

Moreover, pupils do not interact physically with the materials and miss the tactile and physical elements of doing experiments. The hands-on interaction with objects and the experience of frustration and unexpected discoveries are essential to acquiring new knowledge within experience-based learning. In the two versions of virtual laboratories described in the paper, Labster and PhET, the problem and the learning outcome is formulated in advance. In Labster, the simulations
are often built around a mission, so the element of problematisation and inquiry is incorporated into the simulation. But the inquiry is not based on the pupils formulating a hypothesis from their own musings and interactions with the world, which again is at the core of experienced-based learning. Other more open-ended simulations, which allow the pupils to access the software system and re-design or code the simulation, can be more explorative and supportive of creation and participation, as Papert advocates. However, the simulation is still tied to the restrictions of how it is programmed and is a representation of reality.

The virtual laboratory can be understood as a hybrid between explorative and instructive learning. In the virtual laboratory, some of the central elements of the experimental physical practice are removed, as stated above. The pupils interact with representations or visualisations of objects and do not learn the craft or practical skills of experimenting in a physical laboratory. The pupils experience tinkering in the virtual laboratories and interact both with (the medium) and in the laboratory. But it is another form of tinkering than the hands-on interaction with the objects in the physical laboratory. Experience-based learning is simulated, and the practice of experimenting is translated into a virtual setting, which is another form of learning practice.

On the one hand, one could argue that the virtual laboratory cannot be a fully explorative practice due to the medium; on the other hand, virtual laboratories can be both designed and used in more explorative manners. Further, compared to other learning materials, such as textbooks, static illustrations or videos, the virtual laboratories offer a more interactive and explorative approach to learning. Moreover, the virtual laboratory allows for exploring phenomena and discovering processes that only are made possible because of the medium, e.g. working with and exploring cells or atoms. In this sense, the virtual laboratory can be a hybrid between instructive and explorative learning.

To give a more nuanced and comprehensive answer regarding which learning situations come about with virtual laboratories, they must be understood as being embedded in a wider cultural, social and material context. Even though technology is intentionally constructed, the situated learning practices with virtual laboratories cannot be predicted. Which network revolves around the use of virtual laboratories is fundamentally an empirical question. In other words, educational virtual laboratory practices must be studied ethnographically in the local cultural context in which they are used. Therefore, it is essential to study how learning is constructed around virtual laboratories and how it plays out in the classrooms. How does the sociality in a classroom affect the use of the virtual laboratory? Is it so strong that it leads to using the virtual laboratory in other ways than the originally intended use of the technology? Or does the virtual laboratory have such a strong agency that it forces a specific ordering in the classroom? And how can tacit and embodied knowledge travel in learning situations with virtual laboratories? Questions like these are important to address in more in-depth ethnographical studies, where the sociality and materiality around virtual laboratories are explored and unfolded.

Acknowledgements
I wish to thank Anders Kristian Munk, the anonymous reviewers and the editors for their constructive and useful feedback on this paper.
References


Experiments in Artificial Sociality
Curious robots, relational configurations, and dances of agency

Frederik Vejlin
PhD fellow, Department of Anthropology, Aarhus University

DASTS is the primary academic association for STS in Denmark. Its purpose is to develop the quality and breadth of STS research within Denmark, while generating and developing national and international collaboration.
Abstract

In this article, I explore how experiments with social robots enact and reconfigure more-than-human forms of sociality. I combine recent anthropological discussions of nonhuman sociality with Andy Pickering’s work on dances of agency (1993, 1995) and John Law’s method assemblages (2004) to show how human-robot interaction experiments enact open-ended and decentred configurations of entangling relations between humans and robots. I propose the concept of artificial sociality to capture both the ongoing enactments and multiple results of such experimental reconfigurations. Using these conceptual tools, I unpack the “curious robot experiment” from my ethnographic fieldwork in a Japanese robotics laboratory and compare the kinds of sociality produced in the two experimental conditions. I argue that the curious robot exemplifies what Pickering calls technologies of engagement (2018) by manifesting a form of artificial sociality that augments the unpredictability of dances of agency enacted in (re)configurations of entangling relations.

Introduction

In Robot Futures, roboticist Illah Nourbakhsh describes how modern robotics has “invented a new species, part material and part digital, that will eventually have superhuman qualities in both worlds at once, and the question that remains is, how will we share our world with these new creatures, and how will this new ecology change who we are and how we act” (2013, p. xv). For the field of social robotics, such questions seem particularly relevant (see Seibt, 2016). Social robots are machines designed to have a form of social intelligence that will enable them to interact with, relate to and understand us humans in humanlike ways (Dautenhahn, 1998). Accordingly, humans ideally treat a social robot “as if it were a person, and ultimately a friend” (Breazeal, 2002, p. xi). In this sense, social robots do not seem particularly superhuman. But this does not render Nourbakhsh’s questions less pressing since it still appears critical to ask how we will share our worlds with social robots and how they might change who we are and how we act.

A growing number of scholars in the humanities and social sciences are already grappling with these questions (e.g., Alač, 2016; Robertson, 2017; Šabanović, 2014), often with an unmistakably critical edge (e.g., Jones, 2017; Richardson, 2018). For example, psychologist Sherry Turkle describes our present as the “robotic moment” and argues that we are gradually replacing the intimacy of authentic human relations with the shallow and unfulfilling illusions of connection offered by social robots and similar technologies (Turkle, 2011, pp. 129-133). If we do not recognise how these deceptive technologies endanger the future of human sociality, we risk eroding the essence of humanity (Turkle, 2011, pp. 17-20). In the following, I endeavour to provide a slightly less bleak alternative to Turkle’s depressing predictions. Since 2017, I have been doing ethnographic research in robot laboratories in Japan, where I study how roboticists think about, develop and experiment with various kinds of social robots. I also investigate how they use robots to explore what it means to be human (Ishiguro & Nishio, 2007; Ishiguro, 2020).
In this article, I combine ethnographic material with discussions of nonhuman sociality and experimentation in anthropology and STS to explore the design of robots that technologically simulate aspects of human sociality. Further, I ask how such robots, via interaction experiments, are involved in reconfiguring the ‘thing’, human sociality, they were initially intended to reproduce. This raises issues that transcend social robots by inviting us to rethink what it means to be human and what, if anything, distinguishes us from other entities, both natural and artificial (Moore, 2012). Such issues are also central to Turkle’s critique of social robots, which echoes a humanist tradition that vigorously upholds a strict dualism between humans and nonhumans and maintains that humans are intrinsically exceptional (see also Jones, 2017; Richardson, 2018).

By design, social robotics disturbs such dualistic distinctions, and this partly explains the apprehension from some parts of the humanities and social sciences. But human exceptionalism is not only challenged by roboticists and their creations. A diverse group of scholars in anthropology, STS, posthumanist philosophy and related disciplines have also grown increasingly dissatisfied with the dualisms of traditional humanism. In response, they have developed alternative ways of recognising and exploring how humans and nonhumans are entangled in intimate and complex webs of relations (e.g., Grusin, 2015; Latour, 2005; Kohn, 2013). My work contributes to this project by showing how experiments in and with social robotics might expand how we think about sociality beyond the human.

Here, I do this by examining a human-robot interaction (HRI) experiment from my fieldwork at the Hiroshi Ishiguro Laboratories (the HIL) in 2017. The experiment was designed to evaluate whether a humanoid robot equipped with an algorithm that simulates ‘curiosity’ would enable “more humanlike, interesting human-robot interactions” (Doering et al., 2019b, p. 20). Based on a discussion of this experiment, I suggest that researchers in social robotics and their robotic creations might be experimentally enacting new forms of artificial sociality that emerge in the experiment’s two conditions. I argue that the curious robot exemplifies what Pickering calls technologies of engagement (2018) by augmenting the unpredictability enacted in relational and experimental dances of agency. Finally, I connect my argument to the work of other scholars with similar interests in more-than-human socialities and human-robot relations.

**Social Robotics**

In Designing Sociable Robots (2002), roboticist Cynthia Breazeal defines social robots as technological systems designed to simulate the cognitive and communicative capacities that enable humans to engage in social interactions and establish intensive, intimate and durable relations with others. Building such machines is a lofty goal. As Breazeal explains:

> [o]ur sociability touches upon the most human of qualities: personality, identity, emotions, empathy, loyalty, friendship, and more. If we are ever to understand human intelligence, human nature, and human identity, we

---

1 Hofstede and Liu (2020) and Rezaev et al. (2018) also discuss artificial sociality but in ways that differ substantially from my use of the concept.
F. Vejlin: Experiments in Artificial Sociality

cannot ignore our sociality (Breazeal, 2002, p. 239).

Big stuff! But researchers in social robotics face an even more significant challenge. In-depth scientific knowledge of the mechanisms that make humanlike sociality possible is required to begin engineering technologies that can even approximate such abilities (Breazeal, 2002; Alač, 2016). However, as several roboticists argue (Breazeal, 2002; Asada et al., 2002; Kuniyoshi, 2019; Nagai, 2019), our current grasp of sociality does not allow us to reverse-engineer its necessary components and simulate them technologically.

Nevertheless, the process of creating such technologies, integrating them in robotic systems and evaluating them in experiments is expected to yield an “uncanny advantage” for producing the in-depth knowledge of sociality we are currently lacking (MacDorman & Ishiguro, 2006a). A brief and partial review of research in social and cognitive robotics shows how robots of varying complexity have been used to explore: the workings of human sociality (Breazeal, 2002; Dautenhahn, 2007), the development of social cognition (Asada et al., 2002), the mechanisms of empathy (Asada, 2015), the role of embodiment for intelligence (Kuniyoshi, 2019; Pfeifer & Bongard, 2006), and even artificial consciousness (Tani, 2017). In this sense, social robotics is a peculiar scientific discipline that marries the development of social machines with the creation of experimental approaches that use these machines to investigate human sociality.

For example, Karl MacDorman and Hiroshi Ishiguro argue that an android, a robot that is “indistinguishable from humans in its external appearance and behaviour” (MacDorman & Ishiguro, 2006b, p. 289), will elicit the same conscious and subconscious reactions as a human subject would in interaction experiments (MacDorman & Ishiguro, 2006a, p. 298). In such experiments, android subjects might be preferable to humans since they can be programmed to behave consistently across different conditions. Therefore, androids, and potentially other robots, might provide novel occasions for exploring “what it means to be human” since they “offer insights into human behaviour that may be unobtainable by other methods” (MacDorman & Ishiguro, 2006a, pp. 301-302). They do this partly by providing “the unique opportunity to study human affect, cognition, and behaviour when confronted with social agents other than humans” (Bartneck et al., 2020, p. 7).

But this brief introduction leaves the notion of sociality largely unexamined. In the next section, I work through perspectives on more-than-human socialities in anthropology and STS to show the tensions that emerge when compared to the view from social robotics.

Sociality Beyond the Human

It is somewhat ironic that anthropology, the eponymous study of humankind, has recently seen a surge of interest in exploring how human worlds and lives are deeply intertwined with the nonhumans with whom we co-exist (Cerulo, 2009; Tsing, 2013). Although earlier anthropologists recognised how nonhuman beings participate in social life (e.g., Leenhardt, 1947 [1979]; Hallowell, 1960 [2002]), it is only within the last decades that we have become genuinely comfortable with not-quite-human forms of sociality (Long & Moore, 2012; Remme & Sillander, 2017). Indeed, as Anna Tsing asks, “How could it have ever occurred to anyone that living things other than humans are not social?” (2013, p. 27). Accordingly, Tsing (2013, 2016), along with many like-minded thinkers (e.g., Bubandt, 2020; Kohn, 2013; Lien & Pálsson, 2019; Remme & Sillander 2017), have persuasively advanced the claim that human social life always has and always will exist and evolve in intimate and complex entanglements with nonhuman beings.

Although notions of more-than-human sociality are gaining traction within anthropology, they differ from how sociality is understood in social robotics, despite agreeing that genuine sociality is not reserved for humans alone. Compared with contemporary anthropological usage,
the notion of sociality mobilised in social robotics seems incomplete. In social robotics, human sociality is conceived as something we possess as skills and capacities for cognition, communication and interaction (Dautenhahn, 1998). As noted above, roboticists have yet to figure out precisely what these skills are and how to recreate them technologically. But this issue is understood as a scientific puzzle they can solve most productively via technological innovation and experimental evaluation. Once they have determined these mechanisms and have built robots equipped with technologies that can sufficiently simulate them in interactions with humans, they are likely to possess humanlike sociality (Breazeal, 2002, p. 235). Put differently, in social robotics, sociality is a question of what you can do rather than who or what you are.

However, from an anthropological perspective, this seems to ignore a crucial part of sociality, namely the entangling webs of social relations that produce and are produced by our evolved, species-specific capacities for social cognition, communication and interaction (Ingold, 2000, pp. 4-5). As understood in anthropology, the concept of sociality points not only to the possession of such capacities but also to "the relational matrix which constitutes the life of persons" (Strathern, 1996, p. 55), wherein persons are conceived as "simultaneously containing the potential for relationships and always embedded in a matrix of relations with others" (Strathern, 1996, p. 55). In this view, we can simultaneously acknowledge that humans possess something that affords them the potential for relationships – the attributes and abilities that roboticists attempt to simulate – while insisting that such abilities grow "in entangling relations with significant others" (Tsing, 2013, p. 27).

Such a notion of relational sociality does not discriminate as to which entities can be relationally entangled. Partly resonant with the view from social robotics, relational sociality is less concerned with intrinsic qualities and assumed abilities. Instead, it attends to how diverse entities do relations together and the knotty entanglements such relations grow from and extend. As Alfred Gell notes, "it does not matter, in ascribing 'social agent' status, what a thing (or person) 'is' in itself; what matters is where it stands in a network of social relations" (1998, p. 123). Thus, following Gell and others of a similar persuasion, I understand sociality as an "ongoing relational process which can take a variety of forms" (Remme & Sillander, 2017, p. 9), through which both humans and nonhumans collectively create and inhabit webs of relations that are "intrinsically plastic and malleable, expanding and contracting, including and excluding, continuously" (Remme & Sillander, 2017, p. 20).

This version of relational sociality does not deny that some entities relate differently and that this is partly a result of differences in their respective abilities. But it maintains that such abilities are relational products to be investigated, empirically and conceptually, as they affect ongoing relational entanglements (Šabanović & Chang, 2016, p. 540). Superficially, it seems that social robotics leaves this relational dimension unexplored and ignores how sociality extends beyond the abilities of individuals. But consider Yuji Sone's summary of Hiroshi Ishiguro's take on the relational production of humanity:

The notion of the human should be defined in terms of one's ability to form relationships, that is to say exteriorised encounters, with other humans, and, further, that these relationships are based upon mechanistic exchanges built of specific gestures and behaviours that can be replicated (Sone, 2017, p. 100 summarising Ishiguro, 2012, p. 49, original references removed).

Despite Ishiguro's somewhat behaviouristic tone, he seems to argue that what we understand as characteristically human, such as sociality, is an emergent effect of our relations with others, rather than being derived from intrinsic qualities (Otsuki, 2015, p. 158). But I do not think Ishiguro takes it far enough. Thus, my relational version of sociality emphasises how relations with other beings are partially responsible for producing the abilities that humans (and nonhumans) use to establish and maintain these relations, while also insisting that...
such entangling relations enact and are enacted by multiple kinds of overlapping socialities.

In short, my argument is that when researchers in social robotics attempt to simulate sociality in robotic systems, the very form of sociality itself is dynamically transformed as new relational configurations emerge in interactions between humans and robots. Concomitantly, I propose that the simulations of sociality embodied by social robots will probably not result in perfect replications of existing relational configurations, regardless of how well they reproduce the abilities humans are said to possess. This becomes clearer when considering how social robots are involved in HRI experiments. Luckily, STS scholars have developed tools for showing how experiments in social robotics already rely on relational configurations of humans and nonhumans.

Exploring and Performing Experiments

In the late 1970s, a small revolution rocked the social studies of science as a growing number of researchers started doing fieldwork in laboratories and began investigating what Bruno Latour has aptly described as science in action (Sismondo, 2010, pp. 106-107; e.g., Knorr-Cetina, 1981; Latour & Woolgar, 1986[1979]; Latour, 1987). As Karin Knorr-Cetina explains, such laboratory studies “furnished the optics for viewing the process of knowledge production as “constructive” rather than descriptive; in other words, for viewing it as constitutive of the reality knowledge was said to ‘represent’” (1995, p. 141). Where earlier work in the sociology of science investigated the social construction of scientific knowledge (Sismondo, 2010, p. 54), the laboratory ethnographers instigated a more radical project. Roughly, they studied how collectives of human and nonhuman agencies brought together in scientific laboratories construct reality (Law, 2004, pp. 31-32).

This move became most apparent when STS scholars interrogated experimental practices. They argued that experiments do not only produce descriptions of natural phenomena but also construct the phenomena in question (Law, 2004, p. 45). To paraphrase Ian Hacking, experiments never just represent reality but always intervene in it (1983), insofar as establishing proper experimental conditions requires “control of the variables studied, of the technologies applied, of the experimental design” (Roepstorff & Frith, 2012, p. 103). As a result, scientific experiments looked less like rigorous applications of methodological principles and more like “a complicated practice, a bricolage tinkering with the possible elements to make things work” (Roepstorff & Frith, 2012, p. 103). Importantly, they also revealed experiments to be more-than-human, “for as scientists well know it is only through an organized and coordinated effort, using multiple machines and other things as mediators, that different entities become able to reliably ‘express themselves’” (Jensen, 2010, p. 7).

Consequently, the realisation that the sciences produce knowledge by constructing the phenomena they only claim to describe emerged alongside a heightened sensitivity to the participation of nonhuman actors in experimental practices. Callon and Latour aptly capture this sensitivity via the principle of generalised symmetry (1992). As Casper Bruun Jensen explains:

Generalized symmetry can be viewed as a methodical insurance policy against taking for granted any preconceived notion of who has the power to act. It thus multiplies the potentially relevant actors and forces attention on their differences and relations. The aspiration is to thereby facilitate a more nuanced analysis of how humans and things (broadly construed) together create, stabilize and change worlds (2010, p. 5).

I hinted at this symmetrical attitude above by suggesting that we should recognise nonhumans as active participants in producing the entangling relations that shape socialities, and the same holds for whom or what we identify as participants in experiments (Pickering, 2021).
When doing experiments, scientists inevitably rely on a whole “realm of instruments, devices, machines, and substances that act, perform, and do things in the material world” (Pickering, 1993, p. 563). To describe how experiments produce configurations of relations between scientists and their nonhuman collaborators, Andy Pickering proposes the image of the dance of agency, wherein “material and human agencies are mutually and emergently productive of one another” (Pickering, 1993, p. 576).

As Pickering stresses, such dances of agency are performative, open-ended and decentred. They are performative since “performances are what agents do, whether human or nonhuman”, and experiments involve diverse agents doing things together (Pickering, 2010, p. 195). They are open-ended because we never know where the dance will lead; their results are never given. Ideally, experiments traverse the unknown to produce moments of surprisal that reciprocally transform the humans and nonhumans involved (Pickering, 2018, p. 3). Also, by recognising how nonhumans crucially contribute to experiments, we relinquish dualistic control and decentre the human subject (Pickering, 2018, pp. 4-5). In short, dances of agency are “zones of intersection where the nonhuman world enters constitutively into the becoming of the human world and vice versa. They cannot be accounted for by focusing either on the human or the nonhuman alone” (Pickering, 2010, p. 195).

Thinking of experiments as dances of agency seems far from the rigid methodological prescriptivism usually attributed to the scientific method (Law, 2004, p. 40). John Law provides a clear, albeit slightly caricatured, description of this ‘standard view’:

If you want to understand reality properly then you need to follow the methodological rules. Reality imposes those rules on us. If we fail to follow them then we will end up with substandard knowledge, knowledge that is distorted or does not represent what it purportedly describes (2004, p. 5). Law clarifies that he is not advocating for the casual dismissal of ‘standard’ methods. But, he maintains, when we conceive of methods in this prescriptive register and use them to study things that are “complex, diffuse, and messy,” we tend to make an even bigger mess of things (Law, 2004, p. 2). In an example of redescription (see Lebner, 2017), he invites us to understand methods as method assemblages, as processes “of bundling, of assembling, or better of recursive self-assembling in which the elements put together are not fixed in shape, do not belong to a larger pre-given list but are constructed at least in part as they are entangled together” (Law, 2004, p. 42).

In thinking with method assemblages, we abandon the standard view’s anthropocentric quest for experimental control and the associated assumption that reality (in the singular) pre-exists and determines our use of specific methods (Law, 2004, p. 9). Additionally, the relational configurations that constitute method assemblages exhibit the mutual emergence of Pickering’s dances of agency by the way in which they “are constructed at least in part as they are entangled together” (Law, 2004, p. 42). But, Law continues, method assemblages do not involve constructivism in the traditional sense, seeing as this often implies singular, fixed and relatively stable representations and associated objects (Law, 2004, pp. 55-56; Mol, 2002, pp. 41-42). Instead, drawing on Annemarie Mol (2002), he prefers the notion of enactment since “to talk of enactment is to attend to the continuing practice of crafting” (Law, 2004, p. 56). In contrast to the construction of singularity, enactment results in the production of multiplicity as “the permanent possibility of alternative configurations” (Mol, 2002, p. 164). To think with enactment means attending to processes of configuration and reconfiguration without definitive trajectories and stable outcomes that produce complex, multiple and sometimes overlapping entanglements of heterogeneous elements (Law, 2004, p. 42).

In sum, I understand experiments as method assemblages that enact multiple configurations and reconstructions of relations between humans and nonhumans through open-ended and decentred dances of agency. With this in mind, social robotics and its experiments do not
have the appearance envisioned originally by roboticists. That is, when roboticists build social robots and use them in experiments, they are, like other social sciences, creating "an extension of – and a reflexive moment in – the continuing elaboration and enactment of social life" (Law & Urry, 2011, p. 392). But the ways in which they enact and transform social realities diverge from the traditional social sciences. On the one hand, social robotics is unapologetically constructivist in that it actively creates new social beings, while on the other hand maintaining that these artificial creations enable them to produce more rigorously scientific representations of existing forms of sociality.

But in light of the above redescription of experiments, a different perspective is possible. Extending my argument from the previous section, I propose that when researchers in social robotics attempt to simulate sociality in robots and use these robots as experimental tools for an artificial science of sociality, they will probably not end up with unaltered reproductions or straightforward descriptions of an existing and singular form of sociality. Instead, robots and roboticists might experimentally enact and reconfigure multiple forms of sociality via performative, open-ended and decentred dances of agency.

At this stage, an empirical example seems appropriate. In the following sections, I analyse an HRI experiment designed to test which of two algorithms, the appropriateness learner and the curiosity learner, would produce the most interesting interactions (Doering et al., 2019b). I then discuss how the concept of artificial sociality, understood as experimental enactments of sociality via relational and decentred dances of agency, might help outline the differences between the two interactions.

**Curious Robots**

I experienced the curious robot experiments during my 2017 fieldwork at the HIL, a robotics laboratory in the Advanced Telecommunications Research Institute International (ATR) in Japan. The lab’s eponymous director, Professor Hiroshi Ishiguro, is internationally (in)famous for his work on the Geminoids, a series of androids made to appear indistinguishable from existing persons. Together with colleagues from ATR, Ishiguro also developed the humanoid robot Robovie, who we will meet below, in the late 1990s (see Kanda et al., 2002). The present experiment was the latest in a series of studies on designing robotic behaviours using machine learning algorithms trained on human interaction data, a strategy the HIL researchers call *data-driven HRI* (Liu et al., 2016):

> By directly capturing behaviour elements, such as utterances, social situations, and transition rules from a large number of real, in situ human-human interactions, it may be possible to easily and automatically collect a set of behaviors and interaction logics [sic] that can be used in a robot (Liu et al., 2016, p. 988).

They started working with data-driven HRI to solve the problem of programming robotic behaviours that can adapt to the unpredictability of social life outside the laboratory (Liu et al., 2016, p. 988). So they turned to machine learning. As Adrian Mackenzie explains, “the techniques of machine learning nearly all pivot around ways of transforming, constructing or imposing some kind of shape on the data and using that shape to discover, decide, classify, rank, cluster, recommend, label or predict what is happening or what will happen” (Mackenzie, 2015, p. 432).³ For this experiment, the learning algorithm was an artificial neural network trained on interaction data recorded from scripted human interactions staged and recorded at the laboratory.⁴ The interaction data was abstracted from the recordings using techniques from their previous work (e.g., Liu et al., 2016). The data, now operational, was subsequently used to train two multilayer perceptron neural networks.
For this experiment, the researchers deliberately scripted the interactions used as the training data to exhibit the behaviour they wanted the robot to learn and reproduce (Doering et al., 2019b, pp. 4-6). But the technique ideally scales beyond such scripted interactions. Thanks to current advances in audio-visual and tactile sensor technology, speech recognition and the increasing ubiquity of tracking systems in public spaces, the HIL researchers imagine that “data-driven interaction design based on real-world interactions could soon become a realistic possibility” (Liu et al., 2016, p. 988). For example, “deploying sensor networks in a chain of retail stores could provide hundreds of thousands of example interactions in a manner of months, which could be used to train a robot to perform the role of a shop clerk” (Liu et al., 2016, p. 988).7 Therefore, it seems fitting that they designed the experiment as a camera shop scenario, with Robovie acting as the shop clerk and the human participants playing the customer.

The experiment compared two conditions where they equipped Robovie with different algorithms programmed to generate learned behaviours according to distinct logics. In the first condition, Robovie’s behaviour was generated by a neural network called the appropriateness learner, while the second condition layered the curiosity learner on top of the appropriateness learner. The appropriateness learner generates actions based on their perceived ‘social appropriateness’. After training, the algorithm should enable the robot to “follow the social rules observable in the human-human data” (Doering et al., 2019b, p. 6). When the appropriate robot is confronted with humans performing recognisable actions, the learner selects the top five most appropriate reactions based on the actions pulled from the training data (Doering et al., 2019b, p. 8, 11). The robot then performs the action that most directly replicates what a human would do in the same situation.

The curiosity learner fundamentally reconfigures the logic of replication that animates the appropriateness learner. It does so by exploring and potentially expanding the interactive limitations imposed by training data through the active pursuit of surprisal. The algorithm was inspired by recent research on intrinsic motivation and curiosity, defined by Pierre-Yves Oudeyer and Linda B. Smith as “an epistemic motivational mechanism that pushes an organism to explore activities for the primary sake of gaining information” (2016, p. 2). Moreover, Oudeyer and Kaplan show that curiosity mechanisms can endow robots with “general motivations that push them to explore, manipulate or probe their environment, fostering curiosity and engagement in playful and new activities” (2009, p. 1).

Experimental Interactions

The curiosity experiment was divided into two parts, with each consisting of brief interactions with Robovie followed by a questionnaire and an interview (Doering et al., 2019b, p. 15). As usual in laboratory experiments, the researchers meticulously attempted to manage the human subjects for the sake of comparability. But also to minimize potential disruptions caused by Robovie’s occasional incompetence. They told me to treat Robovie as a knowledgeable shopkeeper, despite its frequent failings, and to ignore its appearance and mechanical voice. Additionally, I should not ask Robovie to repeat itself or go ‘off script’ by asking questions that were too complex or unrelated to the scenario. However, I was encouraged to ask the same questions multiple times or rephrase them slightly to see how it dealt with different and ambiguous questions.

In the first interaction, Robovie promptly welcomed me as I entered the designated area.8 I walked around the room and started fidgeting with a camera. After a few minutes, Robovie approached me, “Is there anything I can help you with?” As per the instructions, I started asking various questions about the camera, “How much does it cost?” “How

---

7 A thorough consideration of the alarming issues that such sensor networks will entail, and what this means for artificial sociality, is beyond the scope of this article.

8 The following descriptions are based on fieldnotes written after or during the experiment. Some of what Robovie said might not be fully accurate since it was recorded from memory.
many megapixels does it have?” “What is the shutter speed?” When I asked straightforward questions about camera specifications, it provided clear and concise answers. When I upped the ante a bit and asked questions that were a bit trickier, for instance, “What kind of manual settings does the Sony DX500 L220 have?” Robovie’s responses would occasionally be completely irrelevant, e.g., talking about the price when asked about the sensor size. When I moved around the room, Robovie would follow me faithfully while listing the specs of nearby cameras. If passivity, predictability and anxiously shadowing the customer are considered appropriate behaviour for a salesclerk in a camera store, Robovie passed the exam.

In comparison, interacting with the curious robot felt substantially different. Here, Robovie was much less predictable, passive and ‘socially appropriate’. Instead, it moved around the room haphazardly and randomly asked me quite detailed questions. Once, it asked me the same question seven times in a row while ignoring my increasingly futile replies. This forced the researchers to intervene and restart the robot. Later, as I was moving towards the Canon display, Robovie decided to show me the Sony camera instead, and from its position near the Sony display, it proudly exclaimed, “This one is a fully professional camera. This professional top-end camera comes in black; it works with nary a little noise! It takes fantastic photos, but it has very complex settings.” As I was about to leave, Robovie moved to each camera, repeating, “Sorry, this one comes only in black” again and again.

Whereas the appropriate robot was reactive, predictable and frankly dull, interacting with the curious robot was interactive, messy and surprisingly entertaining. I will admit that having a human shopkeeper act as the curious robot did would be disconcerting. Even so, when equipped with the curiosity learner, Robovie’s behaviour was much more exciting and enjoyable. Even lively. It is crucial to note that the curiosity learner works, in principle, by adapting the robot’s behaviour to the indeterminacy of ongoing and open-ended interactions. It sustains this unpredictability by choosing actions that produce surprising reactions. Therefore, it might be tempting to attribute the production of unpredictable behaviour to the curiosity learner alone. However, as I will show below, the curiosity learner was not exclusively responsible for enacting the surprising sociality I experienced in the experiment.

To understand why, we need a brief look under the algorithm’s hood to see how it works. Recall that the curiosity learner is created by stacking an additional neural network on top of the appropriateness learner. The appropriateness learner is programmed to classify and choose behaviours based on a principle of replication – if it successfully replicates what a human would do in a similar situation, then the action is considered socially appropriate (Doering et al., 2019b, pp. 7-8; see also Doering et al., 2019a). But this makes it challenging to deal with ambiguous or uninterested customers. In the case of ambiguity, the robot does not explore alternative actions unless they co-occur with the current human action in the training data (Doering et al., 2019b, p. 20). In the case of uninterested customers who refuse to diversify their interactions, Robovie will produce highly repetitive behaviour, e.g. asking the same question multiple times, because the same robot action will always be the most appropriate when the customer’s behaviour stays consistent (Doering et al., 2019b, p. 20).

In contrast, the curiosity learner is animated by what I see as a logic of experimentation, with direct replication being sacrificed for the sake of performing actions with unpredictable results. As one of the researchers told me, the curiosity learner outputs a curiosity score for each available robot action, which is a numerical representation of how confident the algorithm is in predicting how a human will react to the robot. He also told me that “with things like curiosity, the robot can explore a lot of different directions. They might learn things that aren’t necessarily useful for one goal, but maybe they are useful in some other way.” Further, as Doering and colleagues explain, “the ‘curious’ robot was able to adapt its behaviors to some individual customer difference (e.g., interested versus uninterested customers) rather than always using the same default behaviors it learned from the off-line training” (2019b, p. 20). Thus, rather than ignoring a customer’s lack of interest by resorting to repetitive questions, the curious robot will adapt its
behaviour to the current interaction state and leave the customer alone. In this case, actions that might appear useless or inappropriate, such as ignoring the customer, are surprisingly the most curious because they set the stage for further surprisal down the line.

But how does the training data limit these different behavioural logics? In the above, I seem to have ignored the obvious fact that despite their differences, the two algorithms are still trained on the same data. As the HIL researchers explain, “[t]he ‘curious’ robot can only exhibit behaviours that are perceived as curious if such behaviors occurred in the human-human dataset, from which the robot learns” (Doering et al, 2019b, p. 20). In short, if an action does not occur in the data, then the robot cannot do it. However, by attending to the experiment’s open-ended and decentred dimensions, such obvious limitations are redrawn. Even within the boundaries of experimental interactions, the robot requires ongoing relational entanglements to perform both appropriate and curious behaviour.

During my fieldwork, I observed a version of the experiment in which the human participant enthusiastically explained how he liked to ‘break’ robots by purposely testing their interactive limits. In the first condition, he was a stereotypically ‘uninterested’ customer and refused to treat Robovie as a competent salesclerk. Despite this, Robovie would always have an appropriate action up its sleeve, like repeating the same actions until the customer responded. Robovie seemed to be focused more on ‘appropriately’ replicating learned behaviour than on following the flow of interaction. As the HIL researchers note, “[g]iven the same situation, the non-curious robot would simply continue to respond with the same ‘default’ behavior regardless of whether the customer was interested or uninterested, potentially resulting in less ideal interaction than if it had adapted to the individual’s needs” (Doering et al., 2019b, p. 20). But something different happened in the second interaction despite the participant behaving with similar indifference. As they report:

When an uninterested customer continued to ignore the curious robot for some time, the robot would often go back to the service counter, saying “I will be at the service counter if you need any more help.” While this was not the most proactive, salesmanlike behavior, it had the highest curiosity value for that particular situation, due to the fact that the robot had “lost curiosity” [...] about previous actions [...] since those actions did not elicit any unanticipated customer responses (Doering et al., 2019b, p. 20).

To put it bluntly, curious robots do not dance with boring humans (Pickering, 2018, p. 7). The curious robot feeds on the indeterminacy that emerges in open-ended interactions, and it prefers being alone to interacting with someone who does not scratch its curiosity itch. But when humans were willing to dance, “the robot would usually continue answering the customer’s questions and would not leave the customer alone” (Doering et al., 2019b, p. 20).

It might be said that Robovie is still limited by its training insofar as it can only perform actions that are already present in the data (Doering et al., 2019b, p. 20). When the other participant was completing the post-experiment evaluation, he provocatively said: “I smell scripting.” When one experimenter asked him to elaborate, he told her that the “behaviour seemed scripted rather than generated, or at least I hope it was.” In some ways, he was not wrong. In the experimental interaction, Robovie cannot ask about the weather or what you had for breakfast because such actions are not present in the training data. But the algorithms do not replicate any single scripted interaction. Doering et al. explain that:

It is possible that a robot trained on a dataset without curious behaviors can still learn about the humans it interacts with. This is because, at a fundamental level, the mechanism that drives the robot’s behaviors will
always result in robot actions that lead to uncertain human responses, such that the robot can learn more about the human (2019b, p. 20).

Surprising Choreographies
Thinking with Pickering’s dances of agency, I would say that the algorithms find patterns, or choreographies, in the scripted interaction data and, in the case of the appropriate robot, try to replicate the dances these choreographies model. The curiosity learner identifies the same choreographies but reconfigures them by executing actions that disturb the learned patterns. In breaking with scripted choreographies, the robot’s attempts at maximising surprisal reverberate through the entangling relations that enact and are enacted by the experimental dance of agency. When successful, humans will adapt to the curious robot by accommodating its thirst for uncertainty. Therefore, the actual limits are found not in the training data, but instead in the lack of surprisal that emerges when interactions do not yield novel engagements, like when the curious robot meets boring humans.

In these cases, inaction becomes the most curious action because it primes future interaction for increased surprisal should the customer decide to join the dance. From this perspective, the curious robot exemplifies what Pickering describes as technologies of engagement, machines and technologies that overtly embrace and encourage “wild and open-ended dances of human and nonhuman agency in which the nonhuman can always surprise us” (Pickering, 2018, p. 3). He poses this in contrast to technologies of disengagement, “free-standing machines” designed to restrict open-ended dances by limiting the scope of nonhuman agency through designed passivity (Pickering, 2018, p. 4). If the curious robot is a technology of engagement, then the appropriate robot is a technology of disengagement, a relational configuration of agencies that neither decentre the human nor afford surprisal.

As technologies of engagement and disengagement, the differences between the algorithms emerge through how they participate in dances of agency. The point is not to say that the curious robot is properly social while the appropriate robot is not. But only the curious robot embraces the spirit of open-endedness by reconfiguring the experimental dance of agency. In successfully replicating configurations of patterned interaction, the appropriate robot is “cut off from any performative contact with the world” (Pickering, 2018, p. 2). The appropriate robot is tamed by its algorithmic configuration that prevents it from reconfiguring the dance of agency by upholding an asymmetrical dualism that renders it passive and reactive. It only surprises when it “fail[s], and the standard reaction to that is annoyance, not amusement” (Pickering, 2018, p. 4). In contrast, the curious robot reconfigures the interaction by improvising new choreographies in decremented dances of agency that produce multiple and surprising relational entanglements. While the appropriate robot replicates existing configurations and always repeats the same old moves, the curious robot actively encourages indeterminate, open-ended and decremented reconfigurations that transforms the relational choreography by embracing what Pickering calls the “open-ended and exploratory sense of experiment: experimentation as brute finding out. Try and see; what happens if...?” (Pickering, 2016, p. 2).

Nonetheless, there is an additional dimension to the experiment that I have only touched upon superficially and which, initially, seems to suggest a different analysis. Recall how the experimenters encouraged the participants to ignore Robovie’s appearance and voice, avoid specific questions, and play along with the scenario by treating Robovie as competent and knowledgeable. These attempts at pre-configuring both human and robot behaviour to fit with the experimental condition might seem to indicate a return of dualistic human-centred control and as such, the curious would appear to be just another example of the ‘standard’ experiments parodied by John Law above. However, without denying that the experimenters tried to control their participants and streamline the experimental interactions, I contend that the curious robot experiment also reveals the limits of human control.

That is, the two versions of the experiment described above not
F. Vejlin: Experiments in Artificial Sociality

only show how Robovie’s algorithmic configurations contributed to enacting different forms of artificial sociality, but also how experiments, as method assemblages involving decentred, performative and open-ended dances of agency, cannot be understood by only looking at the human part of the equation. Although the experimenters tried to make our actions fit the experimental design, they were not entirely successful. Obviously, the other participant seemingly refused the experimenters’ requests. Even so, the experiment ‘worked’ without his compliance. Moreover, even though I tried to follow their instructions, I occasionally forgot in the heat of the moment. For example, I once made the mistake of asking Robovie to compare two cameras, “Is the Sony better than the Canon?” Robovie responded with the Sony camera’s megapixels and shutter speed.

These examples of intentional and unintentional resistance might have made it slightly harder to stage the right conditions for providing comparable results and producing proper knowledge. The experimenters might have cursed our blunders or reluctance to follow directions. They might even have left out our interactions from the report in Doering et al. (2019b). But the question is whether their attempts at controlling the interaction and our occasional failure to comply changed the enactment of artificial sociality? For the appropriateness learner, it did not seem to make much difference. Even without behavioural guidelines, Robovie would probably continue to replicate supposedly appropriate actions undeterred. In contrast, the curious Robovie actively changed its behaviour by choosing disengagement and inaction when the somewhat apathetic participant did not behave ‘ideally’. It behaved as ‘designed’ even when the participant did not, since the curious robot prefers unpredictable interactions. In other words:

While the simple Appropriateness Learner can learn the repetitive behaviors, the Curiosity Learner can discover which behaviors are likely to lead to individual variation in customer responses. By guiding the interaction toward these behaviors, the curious robot creates opportunities for interactions to develop in diverse ways, opening up paths in the dialog that have the potential to branch out according to an individual’s interests or needs (Doering et al., 2019b, p. 15).

I am trying to make a predictable point. Although the curiosity experiment, like most experiments, involves attempts at controlling the behaviour of the human and nonhuman entities involved, such attempts do not determine how the experiment proceeds in practice. The experimenters’ initial effort at steering the participants’ behaviour to line up with the experimental design undoubtedly influences the dance of agency to some extent. Nevertheless, experimental dances of agency “cannot be accounted for by focusing either on the human or the nonhuman alone” (Pickering, 2010, p. 195; Suchman, 2012).

The experiments required the active participation of humans and robots, as well as a technological infrastructure of sensors, algorithms, speech synthesisers, actuators, and other tools. When we recognise the complex entanglements of human and nonhuman entities involved, it becomes clear that neither the curiosity learner nor the experimenters were solely responsible for enacting artificial sociality in the experiments above. Accordingly, Robovie’s behaviour cannot be fully explained by how the humans involved – participants and experimenters – potentially acted differently across experiments. Instead, I have looked to the zones of intersections and relational reconfigurations produced in the experimental dance of agency. Here, I claim, artificial sociality emerges.

Conclusion

Summing up, I argue that the curious robot provides an example of the experimental phenomenon I propose to call artificial sociality. By technologically replicating human sociality in social robots and developing a new science of sociality, roboticists and robots are experimentally enacting dances of agency that configure and reconfigure entangling
relations. This perspective generally seems consistent with existing anthropological theories of more-than-human sociality. However, artificial sociality still looks slightly different in some respects. As the name implies, artificial sociality is deliberately designed to simulate human sociality, and it is precisely the seemingly reductive and deceptive artificiality of social robotics that critics deplore (e.g., Jones, 2017; Richardson, 2018; Turkle, 2011).

I should like to make two brief comments at this point. Firstly, by ‘defending’ artificial sociality, I do not render humans or other living beings somehow less social. Instead, I am proposing that sociality is not a zero-sum game and that differences in how humans and nonhumans contribute to the enactment of sociality are contingent effects that result from how they are mutually entangled in relational configurations. Secondly, following Law and Urry (2002), I suggested that as soon as we merely claim to describe sociality, we inevitably contribute to its enactment and reconfiguration and that social robotics and the social sciences are equally complicit in intentionally and unintentionally doing so. As I have discussed above, social robots participate in enacting new relational entanglements that experimentally reconfigure the kinds of sociality they are meant to reproduce. It is this experimental dynamic, embodied primarily in the curious robot’s designed augmentation of enacted unpredictability, which makes my artificial version slightly unlike existing perspectives on nonhuman socialities.

Still, I am not alone in examining the experimental reconfigurations that social robots potentially produce. Several scholars in anthropology, STS, philosophy and related disciplines are doing important work on social robots. For example, Jennifer Robertson has convincingly shown how Japanese roboticists and politicians imagine reactionary understandings of ethnicity, nationality, gender and kinship in their visions for human-robot coexistence (2007, 2010, 2014, 2017). Staying with Japan, Casper Bruun Jensen and Anders Blok reveal how Japanese techno-animism reconfigures robotic entities’ ontological nature and challenges modern Western intuition that the spiritual and religious is fundamentally separate from, and opposed to, science and technology (2013, p. 87). In a laboratory context resembling mine, Aslı Kemiksiz explores how Japanese roboticists’ bricolage of scientific styles leads them to understand their robots as mirrormike and partial depictions of organic life and humanlike intelligence, since perfect replication “seem[es] too elusive to pursue” (2019, p. 78). In direct dialogue with my present concerns, Selma Šabanović and Wang Lin Chang analyse interactions with the robot PARO and argue that “rather than being a static characteristic of particular people or artifacts, an actor’s sociality is continuously in the process of enactment” (2016, p. 540). These pioneering scholars, and many others⁹, have been vital in developing my nascent version of artificial sociality.

Finally, I do not intend this article to be an uncritical apology for social robotics. As others have argued (e.g., Suchman, 2007), social robotics often involves reductive visions of human sociality and a lack of appreciation for the contingent complexity and unpredictability of social life outside the laboratory. A more comprehensive account of artificial sociality would carefully evaluate such critical questions, while considering the broader sociocultural and economic implications that social robots likely engender (see Seibt, 2016). However, in the present article, I have bracketed such issues to focus on what we can learn from the experimental practices of social robotics by bringing anthropological discussions of nonhuman sociality to bear on entities who, by design, simulate human forms of sociality. In doing so, I hope to set the stage for further experiments with sociality in anthropology, STS, social robotics and beyond.

References


Bubandt, N. (2019). Coral theory, or what corals can teach anthropology about multispecies sociality [Koralteori, eller hvad koraller kan lære antropologien om multispecies sociality]. Norsk antropolgisk tidsskrift, 30(03-04), 186-204. [in Danish]


Ishiguro, H. (2020). How Human is Human? The View from Robotics Research. JPIC.


F. Vejlin: Experiments in Artificial Sociality


Recasting ethical dilemmas in participatory research as a collective matter of ‘response-ability’

Tine Friis
PhD fellow, Medical Museion, Department of Public Health, and Novo Nordisk Foundation Center for Basic Metabolic Research (CBMR), University of Copenhagen

DASTS is the primary academic association for STS in Denmark. Its purpose is to develop the quality and breadth of STS research within Denmark, while generating and developing national and international collaboration.
This article investigates ethical dilemmas in a research practice of collective memory-work that experiments with articulating personal experiences of how the human gut and psyche connect. Memory-work is a group-based participatory research method, in which participants and researcher write down a personal memory in the third person, read the memory aloud and analyze it collectively. Based on memory-work with a group of women living with autoimmune diseases, I analyze how ethical dilemmas arose around new self-realizations in our articulations of ‘gut’ and ‘psyche’, and how the dilemmas brought up issues about the ‘therapeutics’ and ‘matters of care’ of memory-work. I discuss how the dilemmas were configured through objectifying our personal experiences by writing them down, creating unfamiliar care positions in relation to each other and our past and present selves. In dialogue with Donna Haraway’s concept of response-ability and Steinar Kvale and Svend Brinkmann and Steinar Kvale’s notion of ethical fields of uncertainty, I argue that ethical dilemmas demand ongoing attention, rather than a solution. The analysis shows that ethical dilemmas, when discussed among research participants and researcher, can become useful tools for developing capabilities to attune and respond to ethically important moments in experimental practices. Thus, this article empirically recasts ethical dilemmas as a collective concern, and seeks to contribute to the fields of STS and Psychology by discussing researcher and participant positions, our responses and capabilities to act when facing ethical dilemmas, and the place ethics and response-ability have in participatory research practices.

Jeanette, a woman in her early sixties, spoke these words at a collective memory-work session after having read aloud a personal memory and discussed it with myself and the five other female participants. In this moment, we evaluated what it was like to use the method for the first time. Collective memory-work is a participatory research method, in which a group of participants and a researcher write down a personal memory in the third person, read aloud each memory and together analyze these memories. This textual objectification of participant memories means that the subject and object of memory-work are ‘the same person’ (Haug et al., 1999). Jeanette’s statement invokes an important challenge to this method: Being both subject and object of a research practice can create realizations that touch and move us in unexpected ways. When we analyzed our personal experiences from the third person perspective, our relations to each other and ourselves were externalized, enabling us to ‘look’ at our experiences from the outside and to care about our past and present selves from unfamiliar positions.

Jeanette’s realization mirrors the emancipatory aim of memory-work

---

1 Jeanette used the Danish expression “det sidder da lige der”, in which she uses the Danish verb ‘to sit’. It refers in a figurative sense to appearance, to be expressed in a certain way or to feel something in a particular way. In this context, it seems like Jeanette is feeling something she has not been aware of before. This feeling appears and is now possible to place by verbally pointing to it, making it more tangible for herself and our conversation, although we in principle do not know what “there” refers to. We can only guess: it might appear in the topic, in her body or in a third place.

2 Quotations from the empirical material used in this article are author’s translations. The material was transcribed in Danish and subsequently translated from Danish to English.

3 Jeanette, Lisbeth and Beatrice are pseudonyms used in the article to ensure the participants’ anonymity.
as a feminist practice that experiments with articulating, in words, our bodies and selves to enable memory-work participants to free themselves from self-images that constrain and cause suffering (Haug et al., 1999). Her realization also illustrates the emergence of ethical dilemmas around new self-realizations, and called forth my own discomfort about how to respond appropriately to her statement given my background as a psychologist and novice researcher: I felt my position became ambiguous. While my immediate reaction was to draw on a classical way of conceptualizing ethical dilemmas as the sole responsibility of the researcher, I argue in this article that the conversation that followed Jeanette’s statement recast the dilemma as a collective issue for cultivating a ‘response-able’ research practice. However, as hierarchies of power relations will always be present in research practices, ethical dilemmas will not concern participants and researcher equally, or in the same way.

In this article, I bring together elements from the fields of Science and Technology Studies (STS) and Psychology: I use feminist science studies scholar Donna Haraway’s (2016) concept of response-ability—denoting capability to respond and care in collective practices of knowledge and doing—in dialogue with qualitative research and psychology scholars Svend Brinkmann and Steinar Kvale’s (2008) notion of ethical fields of uncertainty—emphasizing ethically important moments as demanding ongoing attention and reflection. This pairing helps me articulate and understand what happened in the moment Jeanette stated that she might need to see a psychologist. It leads to the argument that ethical dilemmas can be useful tools for developing sensibilities towards others’ motives and actions, thus cultivating capabilities to attune to ethically important moments as these emerge in participatory research practices. Ethical dilemmas can offer us fields of uncertainty in which response-ability can be cultivated both among participants and researcher(s). In arguing so, I also seek to contribute to the discussion of how ethical dilemmas are conceptualized within STS and Psychology. I invite consideration of researcher and participant positions, our responses and capabilities to respond in ethically uncertain situations, and what place ethics and response-ability have in participatory research practices.

I begin by outlining collective memory-work as a participatory research method, and my use of this method in the setting of a medical museum to experiment with how we can talk about the connection between the human gut and psyche. I then discuss some of the ethical dilemmas of memory-work, and how they relate to procedural and relational concepts of research ethics in qualitative research practices. Next, I revisit Jeanette’s statement and analyze the conversation that followed to discuss the implications of Jeanette’s realization. The analysis traces ethical dilemmas about the ‘therapeutics’ and ‘matters of care’ of memory-work as they crystallize around the use of the third person narrator. I examine how these dilemmas can be recast as a collective concern, using the uncertainty of the dilemmas to cultivate response-ability and evaluate the impact of the research, while supporting trusting relations between participants and with the researcher.

**Memory-work on ‘gut’ and ‘psyche’ as ‘experimentation with articulation’**

Frigga Haug and colleagues (1999) developed collective memory-work in the 1980s as a feminist project that sought to theorize and emancipate women’s socialization processes within a Marxist framework. Memory-work enables investigations of personal experiences and their relations to established concepts and social categories by articulating these personal experiences—through written memories and collective analysis—and challenging what otherwise is taken for granted. Haug and colleagues used a procedure of displacement (Haug et al., 1999, p. 55)—in which they as a collective of women investigated the subject of the ‘sexual’ through topics such as hair, legs and body—to show that women’s socialization is a process of sexualization of their bodies. They wrote short texts on these topics, depicting specific actions and
events they experienced while growing up. The texts were written in the third person, as this stylistic form created a distance between the author’s self and the depicted self. Haug and colleagues (1999) argue that this distancing, in translating personal experience into the life of a third person, makes it possible to engage with the depicted actions and emotions from new positions, enabling the memory-workers to step out of conventional figure(s) of ‘woman’ built around the suppressing of norms and ideals that, in turn, restrict women’s lives and self-understandings (Haug et al., 1999, pp. 45–46). Having written texts, Haug and colleagues then collectively analyzed them; deconstructing and rearticulating the absences, contradictions and coherences in the texts.

It is this practice that makes memory-work a highly participatory method, as it blurs the boundaries between studied object and subject, letting the memory-workers’ personal experiences become object of their own theorization (Haug et al., 1999). In memory-work, deconstruction contributes to reconstructing the chosen topic, criticizing a restrictive societal ideology and formulating a new emancipatory ideology, creating awareness about suppressing societal structures and by doing so enabling women to challenge and change them. The ‘deconstructionist’ aspect in memory-work is therefore not just about breaking apart concepts and meanings, but serves as a pertinent step in Haug and colleagues’ affirmative critique (cf. Raffnsøe, 2017) of the socialization of female sexualization.

Inspired by the memory-work method and its later developments in primarily Marxist or poststructuralist feministic research4, I began two memory-work groups at Medical Museion, a research department and medical museum at the University of Copenhagen, Denmark, where my research is located. My study used memory-work to investigate personal experiences of how the human gut and psyche connect5, specifically how we might experiment with articulating this connection. I chose this topic of investigation in light of ongoing biomedical research elucidating the link between mental health and the microbes living in the human gut. Microbiome scientists argue for a bidirectional understanding of this relation, suggesting that gut microbes not only are key in regulating our physical health, cognitive abilities and emotions, but that our mental health also influences the constitution of gut microbes (Cryan & Dinan, 2012).

The possible implications of this recent biomedical research show themselves in ordinary lives outside the laboratory—for example, when people experiment with ‘do-it-yourself’ fecal transplantations that can be prepared on their kitchen counter (Rodgers, 2020). The travelling of biomedical research findings from the laboratory and into ordinary lives reflects a tendency of scientific developments to become interwoven with cultural discourses around the body and self, as happened previously in, for example, genetic research (Hoeyer, 2016).

The recent focus on microbes augments a longer history of cultural and scientific interest in the relations between the digestive system and emotional well-being (Whiteley & Bencard, 2020), and a growing mainstream appreciation of microbes in and on our bodies (Sangodeyi, 2014). Stephan Helmreich (2015, p. 65) describes an emergent ‘microbiomania’ in contemporary popular culture, representing the human body as a superorganism of microbial communities, although microbiome research is still in early stages. We could, for example, embrace the potentials of microbes by eating fermented foods such as sauerkraut with the intention of feeding the microbes in our gut, aiming to improve our cognitive abilities. Thus, scientific and cultural discourses around microbes and the human body not only describe the relation between gut microbes and mental health; they become

---

4 See, for example, Davies et al. (2001), Gillies et al. (2005), Stephenson (2005) and Onyx and Small (2001).

5 My research is part of the project Microbes on the Mind, which explores the connection between gut, brain and mental health across science, culture/media and personal experience. The project is led by principal investigator and Associate Professor Louise Whiteley and located in the research environment at Medical Museion and the Novo Nordisk Foundation Center for Basic Metabolic Research (CBMR), University of Copenhagen. It is primarily funded by the VELUX Foundation.
prescriptive in everyday and scientific practices by reformatting and socializing what should, could and might be (Paxson & Helmreich, 2014).

Surprisingly little has been published on microbiome research into how the connection between the human gut and mental health affects ordinary lives, potentially influencing people’s perceptions of their bodies, selves and health. The cultural and scientific relevance of such investigations is exemplified in studies of how non-expert research participants culturally, emotionally and bodily relate to microbes (Greenhough et al., 2018) and in investigations of patient attitudes toward treatments such as fecal transplantation (Kahn et al., 2012). Collaborative research publications emphasize this relevance, suggesting agendas for future interdisciplin ary research on the intersection between human and microbial life (cf. Greenhough et al., 2020; Rees et al., 2018).

With the aim of contributing to investigations of everyday perspectives on human-microbial relations and the implications for notions of self, body and health, my memory-work practice helped participants articulate personal perceptions of ‘gut-psyche connections’. Focusing on ‘gut’ and ‘psyche’ was a way to displace (Haug et al., 1999) the problem of ‘microbes’: it was meant to shift the attention away from the physiology linking gut microbes and mental health towards everyday practices in which people’s understandings of themselves and their bodies take form and are negotiated in social interactions. This displacement was also motivated by the notion of a ‘gut-brain-axis’ (see for example MacQueen et al., 2017), commonly used to denote the connection between gut and mental health. This notion and the language use around it easily seem to equate the ‘brain’ with the concept of the ‘self’. Nicolas Rose and Joelle Abi-Rached (2013) discuss similar formulations in the field of Neurobiology. Despite the immediate equation, they suggest that—rather than reducing their meanings—such formulations are informed by concepts of the self from disciplines such as Psychology. When my study investigated how concepts of the self, health and body might be influenced by microbiome research, nuances in language use mattered for how my participants and I could describe personal experiences of our bodies and selves. Therefore, I chose to engage with the ‘psyche’ as this notion seemed to enable my participants to use a vocabulary that both could include and challenge equations between ‘brain’ and ‘self’.

Below, I conceptualize memory-work on gut-psyche connections as a practice of ‘experimentation with articulation’ as we cannot experience the connection between our gut and psyche phenomenologically in the same way as, for example, pain or goose bumps, or even our own hair and legs, as Haug and colleagues explored. So how can we remember and describe our experiences of this connection?

Despite the microbiomania in popular culture, microbiome research—still at early stages—has not become fully integrated into cultural understandings of the self and everyday speech. Our digestive system and mental health are also subject to taboo and personal histories of suffering (Hearn et al., 2020; Martin, 2009). Even though we sometimes talk about how they connect, we mostly do so through metaphors or images that unfold vague descriptions and maintain a socially appropriate distance to the topic, for example, by using romanticized expressions such as ‘butterflies in the stomach’ or unnuanced war metaphors (Brives, 2020) denoting microorganisms as ‘our enemies.’

Experimenting with articulating in words our gut-psyche connections might seem like an attempt to speak more accurately about what is going on physiologically in our bodies. However, this article steps away from such an attempt by drawing on Bruno Latour’s (2004) work on articulation. In this sense, articulation is not about the ability to speak authoritatively and definitively about a certain truth—either in biomedical language or in the vernacular. It is about "learning to be affected by hitherto unregisterable differences through the mediation

---

6 The foundational questions of Psychology as scientific discipline relate to ‘what the psyche is’ and ‘how we can know what the psyche is’. It is outside the scope of this article to unfold the history and use of the theoretical concept of the ‘psyche.’ In my practice of memory-work, the term ‘psyche’ was used to invoke the connection to Psychology, but in particular as a broad, collective notion to refer generally to our mental life [e.g., feelings and self-understanding] and social being in society. In this way, the memory-work participants were able to interpret the ‘psyche’ according to what they felt meaningful. 
of an artificially created set-up" (Latour, 2004, p. 209; Latour’s italics). Latour uses the expression “to be affected” in a double sense: it refers to being emotionally touched and moved, and to be enabled to act. Both senses involve humans and/or nonhumans that learn to be affected by each other, drawing attention to the various kinds of materiality that can enable this process of learning.

Memory-work as experimentation with articulation involved not only my participants and myself, but also our written memories, the setting of our practice and the collectivity that emerged when we did memory-work. As gut-psyche connections were articulated in our practice, we experimented with how to understand these connections, in order to mobilize our bodies and selves under conditions that did not immediately reinforce the well-known metaphors and images described above. This necessitated an open-ended form of experimentation that allowed us to become surprised and explore relations that we might not have been aware of (or thought possible) prior to doing memory-work. I will show what this implies in the development of the analysis presented later in the article.

**Memory-work in a medical museum**

I developed memory-work on gut-psyche connections in the context of Medical Museion’s exhibition *Mind the Gut*. The exhibition examines the relationship between gut and psyche by showing glimpses of how we have attempted to understand this relationship in medicine, culture and personal experience (Bencard & Whiteley, 2018; Whiteley et al., 2017). Engaging with the exhibition in my practice of memory-work offered multi-modal ways of approaching the topic of gut-psyche connections, such as aiding the writing-down of memories by triggering memories, but also by using the exhibits and installations as concrete places to ‘meet’ and discuss the often intangible nature of gut-psyche connections.

My participants signed up to a memory-work group through an open call on Medical Museion’s webpage. To reach potential participants, the call was distributed further across social media, on posters, and in collaboration with the Copenhagen section of the Danish Colitis-Crohn Patient Association. As the exhibition *Mind the Gut* revealed a broad public interest in the human gut and psyche, my target group was anyone interested in exploring the connection between their own gut and psyche, and willing to share this in a group. During sign-up, each participant marked their preferred dates and briefly described themselves and their motivations for participating. I used this information to select and group the participants based on their preferred timeslot and how much experience they seemed to have reflecting on their gut and psyche. The purpose of this selection was to support the group dialogues by increasing the likelihood of shared experiences and touchpoints.

In this article, I analyze empirical material from a memory-work group with four female participants, aged 22 – 62. They are all diagnosed with autoimmune diseases, three in relation to their digestive organs and one in relation to her metabolic rate. Before delving into the empirical analysis, I unfold the notion of memory central to memory-work as experimentation and how this draws out the ethical dilemmas I seek to conceptualize.

**Engaging with memories: Ethical dilemmas and response-ability**

Memory-work is informed by a concept of memory that defines memories as reconstructions of past experiences. The act of remembering is seen as an agentive interpretational process in which the memory-worker reconstructs past experiences in light of the present. Memories and experiences are not understood as something people ‘have’, but as indications of how they in social interactions “constitute themselves and are constituted as experiencing subjects” (Davies et
This conceptual frame implies that memory-work can be seen as a socio-material practice of subjectivity that brings “our changing sense of who we are and who we were, coherently into view of one another” (Keightley, 2010, p. 57) while engaging with the textuality of the written memories.

This concept of memory contrasts with mainstream understandings of memories as ‘hard disks’ or ‘containers of past experiences’; understandings that have colored the majority of memory research within psychology (Middleton & Brown, 2005). Mainstream understandings focus on the individual person and see remembering as a neurological process enabled by the neural networks in the brain. They highlight remembering as processes of retrieving more or less correct, factual information and conceive remembering in terms of preservation and loss (Keightley, 2010).

Whereas mainstream approaches to memory point to the biological and psychic fragility of remembering, the concept used in memory-work implies that new self-realizations can emerge, exactly because the coherence and contradictions of each memory, and of the self-depicted within it, are open to questioning and experimentation. Haug and colleagues argue that “in making conscious the material out of which we have made ourselves, we are however not only undermining our own stability; at the same time, we are creating conditions for a more resilient fabric of our lives” (Haug et al., 1999, p. 48). This is a central aspect of memory-work, and the reason why memory-work is described as a feminist method with interventionist and emancipatory aims.

New self-realizations may emerge when, in memory-work, we experiment with articulating past experiences. Such realizations may touch and move us. Some might even experience this as ‘therapeutic.’ However, the therapeutic aspects that might come about from participating in research are crucially different from undergoing psychotherapy. Judith Kaufmann and colleagues (2003) compare memory-work with psychotherapy, and argue that psychotherapy often aims at relieving personal suffering by helping the client live better within social norms and to cope with difficult life circumstances such as illness. Despite their different theoretical underpinnings, psychotherapies generally accept that social norms and demands are a given and not subject to change in psychotherapy (Kaufman et al., 2003). This is an important difference between psychotherapy and memory-work, since memory-work seeks to investigate and challenge social norms and demands. Memories are analyzed to understand how we, as individuals, appropriate social structures, and to use this newly developed understanding to start changing restrictive social structures.

Another noteworthy difference between memory-work and psychotherapy lies in the roles of the researcher and therapist respectively. Psychotherapy conducted by formally trained psychiatrists and psychologists relies on their expert knowledge to relieve suffering, for example, through psychoeducation or the therapist’s training in successful therapeutic strategies. The psychotherapeutic conversation is also relationally structured around this position of therapist as expert and client as ‘service user.’ Although the researcher needs expertise in facilitating memory-work and can begin learning its craft through research guides such as the one developed by Haug (1999), the researcher also writes down her own personal memories and participates just like the other memory-workers. This is an attempt to flatten out the asymmetrical power relation between researcher and participants in memory-work. It allows the memory-workers to experience the research process as a ‘joint venture’ in which they can make a difference, contributing with their experiences. The dual position of the researcher as facilitator and as research subject and object like the other memory-workers, demands an ongoing negotiation of the

---

8 Unfolding the extensive theoretical foundations of memory-work is outside the scope of this article. For a thorough discussion of the theoretical background, see Haug and colleagues (1999).

9 This article uses the notion of ‘psychotherapy’ to refer to evidence-based treatments, in which educated healthcare professionals such as psychologists and psychiatrists aim at relieving mental health issues in conversation with clients of all ages (American Psychological Association, n.d.).
researcher’s role (Given, 2008). It also shows that a complete removal of such power relations is not possible, and I would argue that it would not be a desirable outcome in memory-work or psychotherapy. Rather than relieving suffering (as in psychotherapy), the researcher’s position as facilitator in memory-work is important for ensuring a research practice in which the participants feel comfortable—for example, by knowing that the researcher is leading the discussion and taking responsibility for ensuring confidentiality, time-keeping and positive tone of communication in case of disagreements.

While new self-realizations may affect us while doing memory-work, the ways in which we might be affected cannot be predicted in advance. This is a general ethical dilemma of qualitative, participatory research such as memory-work. Marilys Guillemin and Lynn Gillam (2004) distinguish between ‘procedural ethics’ and ‘ethics in practice’ as two central ways of approaching ethics in qualitative research. Procedural ethics, such as approval from ethics committees, are institutional structures set in place to tackle ethical dilemmas prior to conducting a study, prepare for what might go wrong, and assure the accountability of the research team and institution. Historical examples of medical experiments that gravely mistreated their research subjects have shown the necessity of ethical principles for responsible conduct of research and for institutional adherence to such principles (Hilppö et al., 2019). Since its implementation, however, this formal regime of research ethics has been met with criticism from qualitative researchers who find it a poor fit for their methods where reflections on the researcher’s role, responsibilities and impacts are an integral and evolving part of the empirical work (Ellis, 2007; Swartz, 2011). Here ‘ethics in practice’ is a better fit, because this instead highlights dilemmas that emerge while researching, and invites us to think ethics into the evolution of practice, as in the case at the center of this article.

Ethical dilemmas in practice show themselves in ethically important moments, which are “difficult, often subtle, and usually unpredictable situations” (Guillemin & Gillam, 2004, p. 262). They raise questions about how to respond appropriately in the moment, and where the options for ‘responding ethically’ are not obvious. The notion of ‘dilemma’ employed in this article implies attention to conflicts and ambivalences, and an openness to giving opposing tendencies their proper weight. This notion of dilemma differs from an understanding of dilemmas as dichotomies (Guillemin & Gillam, 2004), as a dichotomy cannot adequately represent the possible array of conflicting tendencies that may emerge in practice.

I draw on Brinkmann and Kvale’s notion of ethical fields of uncertainty to conceptualize ethical dilemmas as such a field: “Rather than seeing these fields as entailing questions that can be settled once and for all in advance of the research project, we conceptualize them as fields of uncertainty, i.e. problem areas that should continually be addressed and reflected upon” (Brinkmann & Kvale, 2008, p. 265). It is in the ongoing attention to the uncertain nature of the dilemma that ethical dilemmas become useful research tools in participatory research such as memory-work, rather than a problem to be solved. They allow the potential for ethically important moments to emerge within the group in a way that acknowledges the uncertainty and contingency of the moment and draws attention to what is happening in the moment, rather than to an external definition of the ‘ethical parameters’ of the situation.

The researcher’s capabilities for ethical conduct are pertinent for both procedural ethics and ethics in practice. While the researcher’s ethical role has been thoroughly discussed in qualitative research literature (Swartz, 2011), this article demonstrates that ethical dilemmas in practice can also become useful tools for developing the ethics of the memory-work collective in dialogue between memory-workers. In the following, I use Haraway’s (2016) concept of response-ability, which articulates an ethical practice of collectively caring and responding. This enables me to nuance the conceptual frame of ethical dilemmas as fields of uncertainty: I draw attention to how we navigate the dilemmas emerging during research as means to collectively cultivate caring responses in participatory research practices.

Being ‘response-able’ differs from the similar-looking expression
‘being responsible.’ Response-ability denotes the capability to respond, react and acknowledge human and nonhuman others rather than having or taking an obligation upon oneself; for example, to keep a promise and be accountable for keeping it. Response-ability can be ‘cultivated’ and ‘shaped’ (Haraway, 2016) by taking part in collective practices of knowing and doing; these practices can develop sensibilities towards other’s motives and actions, calling forth and enabling responses attuned to the specific situation.

I bring this concept together with the notion of ethical dilemmas being fields of uncertainty, thereby emphasizing the relational nature of ethical dilemmas. These emerge when we implicate ourselves in each other’s lives, therefore concerning ‘us’ as a collective researching together. We recall that ethical dilemmas do not concern participants and researcher equally as there are always hierarchies of power present in research practices. This might, however, imply that the uncertainty of ethical dilemmas can be (or should be?) discussed in participatory research practices such as memory-work, exactly because it invites the memory-workers to voice the emotional implications of being both research subject and object as they experiment with articulating experiences of their bodies and selves.

Below, I revisit Jeanette’s statement quoted in the opening of the article to discuss how the conversation that followed her statement exemplifies our way of collectively navigating the ethical dilemmas that emerged from our memory-work practice. The dilemmas discussed relate to the ‘therapeutics’ and ‘matters of care’ of memory-work.

The therapeutics and matters of care in memory-work

It was the second time I had met with my participants to investigate gut-psyche connections. We all wrote a piece of text on the topic...

One time my gut and psyche talked to each other... Each text described a personal memory of a specific action or event in which we experienced our gut and psyche somehow ‘talking’ to each other. Prior to our memory-work session, I had encouraged the participants to write from the perspective of a third person narrator as recommended by Haug (1999). I described the ‘third person’ as an aid to remembering the details of the memory, and as a way to distance ourselves from our experiencing ‘I’ in the memories, thus enabling new positions for interpreting the memories. However, if the participants experienced the third person as restrictive for their writing, I recommended changing the narrator or writing the memory as, for example, a letter or list of bullet points. We took turns reading aloud one of our texts and then analyzed together how it depicted gut and psyche. I asked my participants to avoid interpreting each other’s experiences based on normative or categorical ‘common sense’. The purpose of doing so was to avoid being dismissive by generalizing interpretations such as “she was always anxious as a child so of course she reacts emotionally as an adult.”

Jeanette volunteered to read her text aloud. It was almost one page long and guided us through glimpses of her life. At eight years old, she was hospitalized for inguinal hernia in an adult ward, and was in a bed next to an elderly woman with an ostomy. Because of Jeanette’s hospitalization, she missed the celebratory opening of a new shopping mall. She described her mother’s stomach issues and change of diet, and her family’s good-natured jokes about the mother’s vegetarian “micro-macro” food. As an adult, pneumonia changed Jeanette’s otherwise healthy life after a dose of antibiotics “unsurprisingly” upset her digestive system. A year later, blood from her intestines revealed an autoimmune and chronic inflammatory bowel disease. From experiencing an “indomitable sense of strength and interconnectedness with her body,” Jeanette described the diagnosis as a turning point creating

10 There were four of us in total, including myself, as one participant canceled her participation that day.
a “fundamental separation between her body and mind.”

After a short pause revisiting Jeanette’s text and thinking our own thoughts, we analyzed possible meanings laid out in Jeanette’s text and the resonances it generated among us. For approximately 15 minutes, Jeanette just listened to our discussion before participating in the collective analysis. The content of this analysis will not be discussed in this article. Instead, I direct attention towards our conversation following the reading and analysis of Jeanette’s text, and discuss the ethical dilemmas that emerged in our memory-work practice.

At the end of the day, we evaluated how we experienced doing memory-work together. Jeanette told us how “saying things you don’t usually say out loud” overwhelmed her at times, making her uncomfortable. She elaborated, stating the words quoted at the beginning of this article, which I now revisit in full to explore the implications for our memory-work practice:

I think I need some help from a psychologist. I’ve been seeing a psychologist, but it was for anxiety. I’ve never talked about this. But I can surely feel, it is right there. Because there is clearly something being triggered that I haven’t noticed for many, many years. That’s what was uncomfortable, you know, but there isn’t anything else to do except give yourself a break. I think everyone should know that. Otherwise, we can’t be here. I’m not sure everyone would react like me.

With the words “I’m not sure everyone would react like me,” Jeanette expressed reservation, as if excusing herself. Yet she wanted us to know how uncomfortable she felt, so we could “be here,” thus creating a form of double-bind situation. Her statement called for action while refuting the need for a response. Thus, no reaction to her statement seemed appropriate. It was in this tense field of uncertainty that ethical dilemmas started emerging in our memory-work practice.

We might approach Jeanette’s statement as a discursive expression that places the liability of care on the invoked figure of a psychologist: if a conversation suddenly becomes too personal and overwhelming, you need to see a psychologist. Being educated as a psychologist and now working as a PhD student, I was at first surprised by Jeanette’s statement. I wondered what created this new self-realization and if we somehow were entering a psychotherapeutic realm. What would Jeanette and the rest of the group expect from me based on my disciplinary background, despite the fact that I am not a clinically trained psychologist?12

Jeanette articulated what I will call ethical dilemmas of the ‘therapeutics’ and ‘matters of care’ in memory-work. As these dilemmas emerged in our practice, I suddenly experienced my position as ambiguous, and I considered how to solve them. During recruitment of participants and in our memory-work sessions, I framed memory-work as neither psychotherapy nor medical counselling. As previously discussed, memory-work distinguishes itself from psychotherapy. However, memory-work on gut and psyche will most likely touch upon fundamental issues about being human, living with disease and unpredictable digestive organs. Indeed, the memory-work might very well generate new self-realizations as we experiment with articulating what is usually unarticulated, thus teaching us to become affected by our own and each other’s experiences (Latour, 2004). However, it would be impossible to predict how such realizations would influence us emotionally here-and-now, and in the long term after ending our memory-work practice.

Rather than trying to predict and solve the ‘therapeutics’ and ‘matters of care’ of memory-work, we might understand Jeanette’s statement as a reminder of ethics in practice. Jeanette created and performed her self (Cornett, 2003) in new ways when she talked about something she did not usually put into words, and even invited a group of memory-workers to analyze her experiences of gut-psyche connections. This might

12 In Denmark, a university degree in Psychology can be followed by a two-year practical training, enabling the psychologist to become authorized by the Ministry of Social Affairs. The practical training is not a prerequisite for practicing as a psychologist, but is seen as an assurance of competence (Danish Psychological Association, 2015).
change the way she perceived herself, her digestive organs and disease, family, friends and newly acquainted memory-workers. Even though the focus in this article is on Jeanette’s self-realization, this practice of articulating gut-psyche connections might also change the way the rest of us—listening to and analyzing Jeanette’s text—see ourselves and our human and nonhuman others.

Instead of drawing attention to a procedural framework for dealing with ethical dilemmas, the ambiguity of my position highlights the emerging dilemmas and their potentials in our practice. In the analysis of this article, it therefore becomes a matter of understanding how the emerging ethical dilemmas were recast as a collective concern in our conversations and what difference this recasting made in our memory-work practice.

“I have never talked about this,” Jeanette stated. While writing memory-work texts, we each revisited a previous experience and described it in detail. Smells, sounds and emotions reappeared when we revisited and reinterpreted our experiences, reminding us of how it was to experience that specific moment. While constructing our experiencing self in text, we engaged in a narrative practice that reconstructed our past experience in light of the present moment. Remembering and writing down experiences thus became interpretative processes encouraging self-reflection (Haug et al., 1999; Middleton & Brown, 2005).

The words of Lisbeth, another memory-work participant, reminded our memory-work group that qualitative, participatory research fundamentally differs from research where one hands in blood samples that are subsequently analyzed by a stranger somewhere in a laboratory. Lisbeth continued: “Taking care of yourself [in qualitative research] is important because we have different boundaries. Actually, it’s entirely okay to take care of yourself because it awakens different things for us. We all have different backgrounds, and we’re here with different motivations.” Care became an individual issue of being careful of personal boundaries, rather than a matter of seeing a psychologist.

Lisbeth emphasized that we participated in memory-work as humans, relating to each other, but also to our own experiences; we were not just neutrally listing the content in our medical journals (where health professionals noted our medical conditions etc.). Because of this, we had to understand and engage in memory-work as a practice of care. “Otherwise, we can’t be here,” as Jeanette said. In Danish “Ellers kan vi ikke være her”, this sentence reflects a multiplicity of meanings. It might indicate there would be no room for us if we did not allow rest-breaks. With the word “here,” she might also be referring to Medical Museion where we met to share our experiences of gut-psyche connections. In this sense, giving oneself or another participant rest-breaks indicated the necessity of stopping our investigation before we became too overwhelmed in the analytical process. However, we can also understand the ‘being here’ that Jeanette mentioned as a reference to existentially ‘being here.’

Our written texts and spoken analyses might challenge our current conceptions of our bodies and selves. They might even surprise us and create ruptures in our pre-existing understandings, forcing us to ascribe new meanings to what we have experienced so far (Zittoun & Gillespie, 2016). In this sense, giving breaks is not simply necessary from a research ethics perspective; it is imperative so as to take care of each other and ourselves, because we were experimenting with articulating, in words, together, our human projects of becoming. This seems to challenge the previous notions of care as an obligation of a psychologist, and as an individual concern.

We might reformulate Jeanette’s statement thus: ‘Otherwise, we cannot exist as humans.’ According to Haug, we “attempt in our everyday life to give coherent meaning to ourselves” (Haug, 1999, p. 25). We do so in relation to other humans and nonhumans and to ourselves whether we engage as participants in qualitative research or, as in Jeanette’s comparison, share joys, vulnerabilities and frustrations in a mothers’ group. In memory-work, we paused in the midst of our experiences and conceptions of care, challenged the very coherency of them and found glimpses that did not really fit together. This implied that we broke with “any assumption of the identity between an utterance, its motive, and its [emotional] impact on the receiver” (Nissen & Friis, 2020, p. 110).
Moreover, it drew attention to our experiences as ‘dissensual’ (Nissen & Friis, 2020): staging conflicts between different perceptions and significations of our bodies and selves.

Matters of care became a central subject of investigation in light of writing a personal experience in the third person. Although the ‘third person’ enabled us to analyze our memories from a distanced position, the very distance between ourselves as ‘she’ in the text and ‘I’ doing memory-work afforded us to consider how to care for each other and our past and present selves, rather than whom to care for (Puig de la Bellacasa, 2017). The matters of care in our memory-work emphasized the positions from which we were enabled to care, and the forms of care they made possible. Jeanette mentioned how she initially wrote her text with a first-person narrator and then changed it into the third person, editing the text accordingly:

Jeanette: I wrote it as an ‘I’ narrator, and I thought maybe it’d be a bit more impersonal [in the third person]. Maybe you could distance yourself from it a bit. When I wrote it, I was affected by it, and I thought, I could rewrite it in the third person. Then it might distance itself a bit, but I actually think it had the opposite effect because my adult I started saying, “aww.”

Lisbeth: [laughs]

Jeanette: That was a way of saying: There is me and my rational ‘I’ and there is how I rationally make sense of this. After all, I don’t walk around the hallways and scream everyday just because I have an autoimmune disease [laughs]. You learn that’s how it is. That’s life. It’s not fair. But on the inside, you sometimes still need a hug. When it turned into the third person, it became more isolated. There was a sadness about it, and then there was how you live with it on a daily basis. It’s not very often the other shows up in our ordinary life. Even though you might need it.

Tine: Interesting. I haven’t previously heard about it having this effect.

Lisbeth: I really get you. I don’t know about you—

Beatrice: Yeah, I feel exactly the same way. I’m here as an ‘I’ and it’s easy to just say “I, I.” When I started writing Beatrice, a completely different feeling emerged like, “ugh, I am really sorry for her”. It was more personal compared to writing ‘I’.

Lisbeth: Oddly enough.

Jeanette: That’s interesting to hear because for me it was a strange feeling. I actually expected the opposite.

Tine: Yeah. Yeah.

Beatrice: To distance yourself, but you didn’t. You engaged yourself more because you saw your own name and looked at it from the outside.

Writing from a first-person perspective is a familiar way of revisiting experiences. It is “easy to just say ‘I,’” as Beatrice commented. Although the ‘I’-perspective seems personal, the personal relation to what was remembered changed unexpectedly when we wrote in the third person. Instead of extending the distance from the experience and reducing its affective intensity, the third person allowed us to engage with the experience through new relationalities. As an adult, Jeanette seemed to relate to her eight-year-old self through a caregiver relation. Beatrice articulated this as a form of externalization: The materiality of the written text objectivized the relation to herself, and she found herself looking at her name and actions from the outside, now able to respond to herself as a caregiver.

Maria Jansson and colleagues (2008) describe a similar externalization using the third person in memory-work. They emphasize how it “enables us to approach this ‘she’-person [sic] with greater empathy and understanding; it is a form of textual distance that makes it possible to stay near ‘her’ and take her experiences fully seriously, in a way...
that is more difficult when using ‘I’” (Jansson et al., 2008, p. 235). They describe this mode of writing as an important aspect of allowing memory-workers to theorize their affective and embodied experiences, thus minimizing the risk of normalizing analyses as these might restrain our sense-making processes.

The memory-work text, written in the third person, is an experimentation with articulation—and a particularly valuable one for topics that are hard to articulate. It was not only Jeanette and the rest of us who tried articulating our experiences of gut-psyche connections. Our memory-texts articulated them, too (cf. Latour, 2004). The texts coexisted (materially) with the bodies and selves we were remembering while writing down our memories, and were not simply an intermediary of our experiences. They articulated our experiences in the third person and thereby enabled us to learn to be affected by each other and ourselves anew with this strange, yet familiar, 'she'-person. To return to the idea of 'giving breaks' discussed above, writing down memories and reading them aloud perhaps created a space where the process of articulation could unfold more slowly than in ordinary conversation, thus allowing us to pause and tinker in our experimentation. This underscores the importance of 'the collective' (e.g. memory-work group or research community) in experimental practices, as also shown in several studies of experiments in STS (see for example Danziger, 1990; Shapin, 1984).

Jeanette mentioned that the third person seemed to depict a sadness related to her autoimmune disease, but also illuminated the modes in which she makes sense of her disease and lives with it in her daily life. The modes of making sense seem rarely to crystallize in her daily life, although "you might need it," as she said. Jeanette described a rational 'I' who makes sense of what she goes through. At first, this might seem to indicate a contrast to her sadness, denoting it as an irrational emotion in the raw reality of life, yet she used a humorous tone as she added, “I don’t walk around the hallways and scream everyday just because I have an autoimmune disease.” She seems to meta-comment on stereotypical and prejudiced images of how it is to suffer a chronic disease. By doing so, she distances herself from such images (Martin, 2009). Instead of its being an irrational emotion, we might thus understand her sadness as a form of despair, which almost paradoxically makes it possible for Jeanette to keep trying to make sense of living with disease: “You learn that's how it is. That's life. It's not fair. But on the inside, you sometimes still need a hug.”

With her memory-text, Jeanette invited us briefly to take part in her sense-making processes. It is a vulnerable process, and she was vulnerable in it, surprising herself and the rest of us with the affective intensity of doing memory-work. Jeanette’s statement “I think I need some help from a psychologist” became an ethically important moment in our practice as it allowed us to address how this intensity emerged, and how we could take care of each other and ourselves while doing memory-work—although we were strangers, our only shared interest being our gut and psyche. Discussing the ‘therapeutics’ of Jeanette’s self-realization and the ‘matters of care’ that emerged in our analysis and in the use of the third person, we as a memory-work collective were able to find ways of responding to our own and each other’s reactions to the ethical dilemmas we faced together. Responding appropriately to Jeanette’s statement was not only for me as the researcher to decide, although this was my initial impulse. As we were all both subjects and objects of our memory-work, her statement resonated with all of us, calling for a collective response. In my analysis, our response became the discussion of the ethical dilemmas of the ‘therapeutics’ and ‘matters of care’ which supported us in developing sensibilities towards each other. The discussion of our ethical dilemmas became a collective research tool for cultivating response-ability as we experimented with articulating gut-psyche connections.

Concluding reflections

The opening quotation from my participant Jeanette displays empirically an ethically important moment in our collective memory-work practice.
By stating she might have to see a psychologist, she drew our attention to ethical dilemmas about the ‘therapeutics’ and ‘matters of care’ in participatory research practices. Our participation as both subjects and objects of memory-work let us investigate our written memories of gut-psyche connections from unfamiliar care positions, creating new realizations and experimenting with articulations of the ways in which we more broadly make sense of each other and our bodies, selves and health. By applying Haraway’s concept of response-ability and Brinkmann and Kvale’s (2008) notion of ethical fields of uncertainty, I have traced how the conversation that followed Jeanette’s statement recast the emergent ethical dilemmas as uncertain matters for my participants, and also for myself as the researcher, to navigate. We discussed the dilemmas as an obligation of the invoked figure of a psychologist, as an individual issue of being careful of personal boundaries, and as an effect of writing our memories from the third person perspective. As we did so, we articulated how to care for each other and our past and present selves, thus developing our sensibilities, and capabilities for attuning to the situation, and responding accordingly.

Jeanette’s statement called for action while refusing the need for a response when she stated, “I’m not sure everyone would react like me.” My immediate response was to draw on a classical way of conceptualizing ethical dilemmas as a responsibility that rested upon my role as a researcher. If I had acted on my impulse and somehow tried to solve the dilemma, I might have cancelled out the potential of our ethically important moment, and perhaps even failed to recognize the affective intensity and ambivalence of Jeanette’s self-realization. Instead of trying to solve the ethical dilemmas, our conversation recast the dilemmas as a collective concern. Recasting the dilemmas in this way became our response to Jeanette’s statement and an important research tool for cultivating response-ability together.

The dialogue between the fields of STS and Psychology in this article enables me to problematize researcher and participant positions, our responses and capabilities to respond in ethically uncertain moments, and to question what place ethics and response-ability have in participatory research practices. In doing so, I seek to contribute to conceptualizing ethical dilemmas within STS and Psychology, hoping to inspire future research to engage with ethical dilemmas as useful research tools for cultivating response-able practices. The concept of ‘response-ability’ provides the analysis with a vocabulary for understanding why it matters that we (memory-workers) discuss the emerging ethical dilemmas together. Our conversation reconfigured not only the ethical dilemma, but also our relations as a collective of women and the specific ways in which our memory-work practice made our memories available for interpretation through unfamiliar care positions. While the notion of ‘ethical fields of uncertainty’ emphasizes my responsibility to facilitate an ethical research practice, it also underscores the ambiguity and insolvability of ethical dilemmas and turns my discomfort and experience of the dilemma into an object of the analysis. If I had left out my presence in the analysis, I would have neglected the collaborative methodological underpinnings of memory-work. I might also have failed to understand the ethical dilemma as a collective concern, and as reflecting the emancipatory aim of the memory-work practice, thus dismissing Jeanette’s affective self-realization, so central in our experimentation with articulating in words our gut-psyche connections (cf. Latour, 2004).

The textual objectification of our experiences, written in the third person, externalized the relation to ourselves and to each other; and turned out to be the point at which the ethical dilemmas of ‘therapeutics’ and ‘matters of care’ crystallized. The textuality of memory-work raises new questions worth pursuing in future forms of participatory research. Today, the common and familiar use of first-person perspectives in for example SoMe or genres like autofiction and New Journalism might be reinforcing the first-person perspective as being, or appearing to be, more distanced than that of the third person. This seemed to be the case in our memory-work collective. We were surprised by the intimacy of the third person perspective. This indicates a historical difference in the meanings ascribed to narrator perspectives since Haug and colleagues developed the method in the 1980s, when they...
highlighted the third person as more emotionally distanced than that of the first person, and as enabling unfamiliar positions from which to investigate personal memories.

In some ways, the first-person perspective already seems to externalize ourselves in the text, just like my chosen memory-work topic *One time my gut and psyche talked to each other...* externalizes gut and psyche from 'my' body and from 'my' self, almost ironically invoking a first-person perspective despite my suggestions of writing in the third person. I wonder what would happen if we wrote memory-work texts from a second-person perspective, addressing a 'you' difficult to identify, or if we write from a plural perspective, articulating an 'us' and 'we' experiencing, remembering, reinterpreting. *One time my gut and psyche talked to each other* surely evokes different experimental potentials, ethical dilemmas and response-abilities than *One time your gut and psyche talked to each other* or *One time our gut and psyche talked to each other*...

**Acknowledgements**

I would like to express my very great appreciation to my fellow memory-workers for joining me to delve courageously and empathically into their own and each other's personal experiences of gut-psyche connections. I would also like to express my deep gratitude to Associate Professor Louise Whiteley and Associate Professor Adam Bencard, my supervisors at Medical Museion, and my supervisor Professor Morten Nissen, Aarhus University, for enthusiastic encouragement, critical comments and thought-provoking discussions. Furthermore, the Microbes on the Mind research group has supported developing the method, and the internal seminar group at Medical Museion has provided comments on initial drafts; both groups allowing new perspectives to emerge. My thanks are also extended to the wider staff at Medical Museion for facilitating the project.
References


Experimenting on the Enactment of Predictive AI: The Quest for a Future Proactive Healthcare Sector

Anne Henriksen
PhD fellow, Science-Technology-Society Centre, Aarhus University

Finn Olesen
Associate professor, Department of Digital Design & Information Studies, Aarhus University

DASTS is the primary academic association for STS in Denmark. Its purpose is to develop the quality and breadth of STS research within Denmark, while generating and developing national and international collaboration.
Abstract

Currently, a large number of AI projects are experimenting with the use of AI and big data for various purposes, especially in the public sector. In this article, we explore one such AI project. Specifically, we study a group of developers in Scandinavia and their efforts to enact predictive AI through the development of a clinical decision support system (CDSS) in pursuit of a future proactive healthcare sector. This yet-to-be system was envisioned to prevent unplanned hospitalizations by ‘turning’ what we term ‘potential patients’, i.e. the effective management of patient trajectories, in pursuit of a proactive healthcare sector. In the article, we investigate this particular project as an ‘experiment’ and conceptualize the developing CDSS as a ‘partially existing object’ with an uncertain ontological status. By studying the gradual enactment and emergence of the CDSS, we illuminate how this fuzzy data-driven object is performed and gradually attributed with solid reality: during its creation process, it advances from being a proactive device imagined to be used in primary healthcare to becoming a triage tool embedded in the prehospital emergency department. Along the way, the project developers are also transformed, learning what ‘moves’ and ‘actions’ to make, and, thereby, becoming skillful CDSS-operators. By using ‘experiment’ as our analytical lens, the article renders visible how persons, locations, and procedures have to be changed, revoked, and suspended in order for the AI project to succeed. Thus, the article contributes to showing how ‘social mangling’ is an essential precondition for predictive AI to succeed as a prolific solution to specific healthcare challenges, along with developers’ learning and transformation.

Introduction

With several recent technological advances and the explosion in digital data (Babak, 2015), artificial intelligence (AI) seems to pose a new “promissory technology” (cf. Hoeyer, 2019; Tupasela & Di Nucci, 2020) imagined to solve all sorts of challenges outside confined laboratory spaces. In Amsterdam and Helsinki alone, more than thirty AI projects are currently running (Olsen, 2021). Similarly, a recent report on automated decision-systems finds that a large number of countries are “experimenting” with the use of AI for various purposes, especially in the public sector (Chiusi et al., 2020: 6ff.). Precisely the word ‘experimenting’ seems apt to describe the situation, even though modern AI has been with us for some time through various applications (Bryson, 2019). As Stilgoe (2018: 26) suggests, modern AI is still very much “a work-in-progress laden with promises for what it might become”. Self-driving cars relying on big data and AI sometimes crash (Stilgoe, 2018), and development projects, which have otherwise been championed, are occasionally discontinued because the use of algorithms turns out to nullify the expectations (see e.g. Hao, 2020; Heaven, 2020).

It is not new that particular capacities and reality are attributed to technological objects even before they have a stable existence. By virtue of expectations, aspirations, and imaginaries (Jasanoff & Kim, 2015), they are enacted as partially existing objects (Latour, 1999; Jensen 2010). Following Bruno Latour, we explore AI algorithms and the computer systems in which they are embedded as partially existing objects; objects which have an uncertain ontological status, and which exist and are defined only relatively to their networks of construction (Latour 1999; Latour & Weibel 2005; Jensen 2010).

How do partially existing objects as fuzzy as developing AI algorithms, with only limited materiality, garner existence? How do they and the computer systems they are built into go from being ‘weak’ objects to gradually becoming more and more ‘real’ material devices that have particular uses and are woven into well-established practices? How are they enacted as specific versions and realities? By means of
what politics? And as part of which imaginative spaces of opportunity enabled by particular actors?

These questions motivate our study and are explored on an empirical level based on one particular Scandinavian AI project, studied ethnographically by the first author: This project strived to design and develop a clinical decision support system (CDSS), based on modern AI techniques and big data. In the article, we investigate this particular project as an experiment. We do so to better grasp the emergence and enactment of the CDSS, and the dynamics played out along the way, i.e. how and why the AI project developed in a certain manner: This strategy reveals the preliminary and predetermined existence of CDSS before its actual development, and yet it suggests that its ontological status is uncertain due to a lack of tangible qualities, materiality, and embeddedness into specific practices. By conceptualizing the AI project as an experiment with a partially existing object – the CDSS – we accentuate the question of how such objects are performed and attributed with existence in highly local design spaces, and how they attain stability.

Our article does not provide a detailed analysis of how the AI project progressed during the period of ethnographic inquiry. Instead, it seeks to understand how the project and the developing CDSS ties to larger societal transitions and the social and political shaping of society in virtue of their immersion into particular socio-technical settings. For the sake of the anonymity of the informants in the AI project studied, we do not refer to documents etc. that may disclose information about the particular project.

While much STS research has studied developing technologies and technological futures not yet ‘boxed in’ (cf. Latour, 1987), especially in the healthcare domain (see e.g. Jensen, 2010), only a limited number of STS studies have focused on modern AI systems and furthermore studied them empirically. This article contributes to the literature by filling this gap. In particular, it aims to show how an experiment-based analysis can contribute to illuminating the ongoing construction process in AI development projects, or the social mangling (cf. Pickering, 1995), through which AI algorithms and computer systems are enacted.

Experiments, Experimentation & Performativity

What is an experiment? Historically, a scientific experiment is a particular step in epistemological inquiries to create knowledge about a delimited phenomenon in nature. Reality must be manipulated so we can learn about it; we must “twist the lion's tail”, as Francis Bacon taught us (Hacking, 1983: 149). The classical experiment is thus modelled after natural science ideals as a method to test hypotheses about a delineated natural phenomenon in a controlled manner within well-defined laboratory spaces. This classical model of experiments and experimentation, however, does not help to explain the big data and AI-based experimental practices in our study.

The philosopher Ian Hacking argues that experimentation needs to be investigated in its own right as a practice. It is not just a “step on the royal road” to knowledge. Experimentation is doing rather than thinking, and the experimental method is not just another name for the scientific method” (Hacking, 1983: 14f). Hacking’s arguments offer a fruitful entry to conceptualize experimentation with AI and big data.

Experimentation regards “the creation of phenomena”, not their discovery, Hacking argues (Hacking, 1983: 220). It is an extremely complicated task to refine and stabilize phenomena as sources of relevant data. Not least, it is difficult to refine what should count as ‘data’ in an experiment. This difficulty traverses a long road from talking about data in a specific context to presenting universal phenomenal statements about the world. The task involves a significant learning process, requiring practical rather than theoretical abilities, where the experimentalists must patiently train a range of skills before they are able to make reliable observations. These are, for instance, the skills of turning, cutting, extracting, preserving, pressing, and repeating. In the process, they must learn when the experiment has succeeded, i.e. when ‘nature’ has spoken. Only when the whole setting and the apparatus work in the ‘right’ way is it possible to observe specific phenomena (Hacking, 1983: 230). It follows that observation plays a relatively modest role in experimental science compared to other
tasks. Endurance and practice create the experimentalists’ ability to distinguish artefacts produced by the instrument from the effects produced by the observed entity.

A widespread paradox in the sciences is that “[…] most scientific experiments don’t work most of the time […]” (Hacking, 1983: 230). Because of the complexities and tough learning process involved in creating phenomena in an experiment, the risk of failure is high. The sociologist of science Andy Pickering agrees and suggests that we consider experiments as complex events in which ‘dances of agency’, or dialectics of resistances and ‘accommodation’, happen:

My basic image of science is a performative one, in which the performances, the doings of human and material agency come to the fore. […] The dance of agency, seen asymmetrically from the human end, thus takes the form of a dialectic of resistance and accommodations, where resistance denotes the failure to achieve an intended capture of agency in practice, and accommodation an active human strategy of response to resistance, which can include revisions to goals and intentions as well as to the material form of the machine in question and to the human frame of gestures and social relations that surround it. (Pickering, 1995: 21f.)

We will return to the image of the dances of agency later. It appears that both Hacking and Pickering exclusively write about experiments situated in concrete settings. Bruno Latour has pulled the experiment out of designated spaces and buildings, and argued that experiments and laboratories are movable devices. This is a significant analytic suggestion because it allows us to follow the ‘CDSS-experimentalists’ as they move around with their partially existing object: the developing CDSS.

Nothing extraordinary or distinctly ‘scientific’ happens inside the walls of the laboratory, Latour (1983: 141) claims. So, why are laboratories considered to be extraordinary places? The explanation is simple. Any notable laboratory has run through a series of displacements in order to achieve its current status. A ‘displacement’ is here understood as a semiotic movement from one position to another on a flat surface. It is misleading to ask, where is the laboratory and where is society? The lab and the society are mixed up from the beginning.

In a dynamic process of displacements, things and humans are transformed. Latour illustrates this through his material-semiotic interpretation of a famous historical event (Latour, 1988). The agricultural system in France was transformed when the microbiologist Louis Pasteur in May 1881 displaced his lab, moving it from École Normale Supérieure in Paris to the village Pouilly le Fort, thereby moving it into the center of French farmers’ interests. When Pasteur returned to Paris, he brought with him two things of utmost importance to the farmers: a cultivated specimen of the anthrax bacterium and the interest of the farmers, who wanted a cure to save their cattle. Thereby, Pasteur’s laboratory was transformed from being a rather secluded setting in Paris to becoming a nationally significant experiment to save farmers’ livestock. Latour’s reading of Pasteur’s achievements is a semiotic overruling of the contrast between text and biological material. Latour asks how the laboratory was made relevant. The answer is that the village was turned into a laboratory: “The only terrain in which a laboratory scientist is a master is that of experiments, of laboratory logbooks, test tubes and dogs” (Latour, 1988: 61). Hence, society must be transformed into such terrain if the scientists are to have relevance in society. We bring Latour’s analysis of the laboratory-society transformations into our investigation as a remedy to learn how the CDSS-experiment was displaced to become a meaningful device in healthcare contexts.

We use Hacking’s detailed exposition of experimenters’ learning approach, combined with Pickering’s evocative notions of the mangle, and Latour’s semiotic analysis of transformative displacements in our own study of the AI-based CDSS. We additionally apply further theories foregrounding performativity in technology development...
and experimental practice.

**Decision Support Systems & AI in Healthcare**

What is a clinical decision support system, and what is experimental about the development of such a system and the use of AI for this purpose? The history of clinical decision support systems tells us that they are “a class of computer-based systems that aids the process of decision making” (Ozaydin et al., 2016: 46) and includes some kind of “decision support capabilities” (Berner & La Lande, 2016: 2). Reading this, one quickly realizes that such systems are not new. In fact, they have been used for more than 50 years as parts of healthcare information systems with a view to “change the way medicine has been taught and practiced” and, in particular, prevent medical errors and improve diagnoses (Berner & La Lande, 2016: 2). Not least, clinical decision support systems have played a crucial role in making electronic healthcare records (EHRs) useable in practice (ibid.). Probabilities and probabilistic knowledge are also not new to clinical practice (Spooner, 2016). On the contrary, they have been instrumental in shaping medical science, especially epidemiology (Tversky & Kahneman, 1974). Finally, it is not novel to use AI approaches to develop computer systems for clinical use, including decision support systems. Several earlier AI approaches, e.g. Bayesian networks forming a part of the artificial neural networks characteristic of AI (see e.g. Press, 2016), were developed in relation to the medical domain through work on knowledge-based systems (Spooner, 2016; Liu et al., 2020).

What is new, however, is the use of automation and big data as means for building probabilistic knowledge and, more specifically, predictions used for making decisions and actions (Mackenzie, 2015; Mackenzie, 2017). In the current new ‘era of AI’, the “inference engine”, as it is called in knowledge-based systems (Spooner, 2016: 31), is learned through data rather than programmed by humans. The result is a machine learning model, or ‘algorithm’, that can automatically process and interpret huge volumes of data, by recognizing patterns for the purpose of predicting future behavior of “entities” (Ozaydin et al., 2016: 46, 50). In this way, data items representing humans (Bechmann, 2019) are automatically classified and mapped into predesigned categories (Bowker & Star, 1999), e.g. sick/not sick, which become pivotal in predictions of, for instance, which people need treatment. This new wave of AI techniques draws on the branch of machine learning methods called *deep learning* (Liu et al., 2020; see also Alpaydin, 2016), and is also referred to as *data mining* or *predictive modeling*. Here, the key objective is to “infer from a collection of data/measurements mechanisms to facilitate decision-making processes” (Ozaydin et al., 2016: 48). Hence, as this quote suggests, data are used as proxies for certain behaviors put under scrutiny.

Today, clinical decision support systems, fueled with big data and algorithms, are often envisioned to improve the management of treatment, medication, and screening of patients (see e.g. Galetsis & Katsaliaki, 2020; Raghupathi & Raghupathi, 2014), for instance by providing and supporting preventive care to individuals through predictions, patient profiling, and segmentation (ibid.; see also Mønsted, 2019). In this sense, modern clinical decision support systems come with a particular *ontology*. This ontology may be characterized by a vision of the world as utterly stable, determinate, and knowable (cf. Pickering, 2016; Law, 2004), composed of data ‘out there’ that merely have to be mined and processed in order for behaviors to be predicted and subsequently tamed and controlled (cf. Berg, 1997). Yet, research shows that extensive work is necessary in order to make data ‘ready’ for uses other than those they were originally produced for and as parts of (Bonde et al., 2019; Møller et al., 2020; see also Loukissas, 2019). Such work and the demand for high-quality ‘reusable’ data have to be viewed in the context of the increasing need for proper data infrastructures, thanks to the gradual turn in the healthcare sector and society to data and data-driven computer systems (Kaun & Dencik, 2020; Bossen & Piras, 2020). In the case of data produced in

---

1 This is referred to as *unsupervised learning* which draws on target variables in the training of learning algorithms (Russell & Norvig, 2016).
In medical contexts, however, it may be that, no matter how much such data are ‘recooked’ (cf. Gitelman, 2013), they will still be ingrained by “uncertainty” due to the approximation, inadequacy, complexity, and ambiguity that generally permeate clinical practice, especially in primary healthcare (Cabitza et al., 2018). This intrinsic uncertainty can create problems for AI developers to achieve veracity of, and ‘truth’ in, algorithmic outputs (Mønsted, 2019; Henriksen & Bechmann, 2020).

In the following, we focus more closely on the empirical basis of our study – the emerging CDSS – by accounting for how it was studied ethnographically.

**The Empirical Basis: Studying the AI Project & the Developing CDSS**

The AI project we study took place in Scandinavia over a three-year period in the form of a research project, initiated and managed by a publicly funded regional innovation incubator; this is described on the company website as an institution with an aim to bridge public and private organizations in order to “develop, test, and implement welfare and healthcare technology, solutions, and services in the national healthcare system”. Furthermore, the project involved a private AI company, a municipality, a state university, and a regional hospital that had already been engaged in a number of technology development projects and was generally noted for its ‘innovation-readiness’. In this way, the project formed a public-private partnership which is characteristic of how digitalization of public services is achieved today in Scandinavia (see e.g. Hockenhull & Cohn, 2021). By asking the steering committee of the project for permission to follow the AI developers’ work on the design and development of the CDSS, the first author was allowed to study this work close-up. She primarily followed the activities inside the AI company. Here, a small team of data engineers, designers, and a physician met on a daily basis, often with the involvement of the project manager from the innovation incubator and the managing director of the AI company.

All in all, the first author followed (Latour, 1987) the developers’ work and thereby the developing CDSS from late 2018 until early 2020, when the computer system was at its infant stage. She investigated the developers’ work by drawing on an ethnographic approach as in earlier ethnographic studies within STS that focus on AI development (Agre, 2016; Forsythe & Hess, 2001; Suchman, 1987). Various methods were used to generate “thick descriptions” of the work (Geertz, 1973). Participant observation was conducted at numerous meetings and workshops held in relation to the AI project during the entire period. Furthermore, the first author stayed with the company on an everyday basis from March to August September 2019, where she performed several spontaneous on-the-spot interviews with developers and conducted day-to-day observations of their everyday work. Additionally, she conducted more than 20 semi-structured interviews (Kvale, 2008), for instance with managers, business developers, data scientists, data modelers and so forth. These interviews were primarily conducted during two periods: August-September 2019 and January-February 2020. Two interviews were co-conducted with the second author of this article. Almost all developers participating in the development of the CDSS were interviewed. As the first author had agreed to assist the CDSS development team in exchange for access, she made the observations and spontaneous interviews in the role as a participant observer (Hammersley & Atkinson, 2007). For instance, she helped to investigate work procedures among general practitioners (GPs) in the very beginning of the project, and to take notes at meetings and participate in discussions. However, she never played any crucial role in the project or in the AI company in general. In fact, she found it difficult sometimes to be truly involved in the daily work on the design and development of the CDSS. Such difficulties may be seen as a fundamental premise of ethnographic research (Hammersley & Atkinson, 2007). All the data used in this article have been fully transcribed and subjected to an iterative analysis process, open to new themes emerging from the data and yet informed by our research interest. More specifically,
the data have been analyzed by means of initial categorization based on “the participants’ voice” resulting in preliminary themes and topics (Malterud 2012, p. 796), repeated readings to generate more condensed meaning units (Davies, 2008), and simultaneous writing and thinking to generate more well-found interpretations of the data (Denzin, 2013; St. Pierre, 2011). Furthermore, the first author analyzed documents collected in the field with the aim to gain a better understanding of the AI project and the developing CDSS. These documents included different descriptions, e.g. the project description in the application for The National Innovation Fund, which granted the AI project its initial funding. Latour (1986) argues that such material objects actively take part in the construction of new ‘things’, both in terms of knowledge and material. Using technical drawings as an example, he contends that material objects serve as visualizations of ‘the future’ because they are used by scientists and innovators in their attempts to convince audiences how their proposed ‘thing’ functions like a roadmap to the future, i.e. the one and only way. From that perspective, the above-mentioned project description would be considered a crucial vehicle in the attempt to mobilize funding for the AI project.

In the following section, we explore which future this project description suggested was engendered by the CDSS and how. In doing so, we make references to this description.

The CDSS Envisioned

According to the project description in the application to The National Innovation Fund, the stated goal of the AI project was to “predict unplanned admissions including readmissions” and thereby “identify” individuals in the risk of such admissions “before they require acute treatment”. In that sense, unplanned admissions were viewed as avoidable and, consequently, manageable admissions. The project was to achieve this goal by means of the proposed CDSS. This yet-to-be system was defined as “a machine learning-based clinical decision support system for proactive healthcare”. It was furthermore denoted a “predictive system” – a system building on machine learning-processed “predictions” generated by an AI algorithm. This was “an algorithm for early identification of unplanned admissions” developed with the use of deep learning methods, namely “deep neural networks (DNN)”. It was the development of this algorithm and an appertaining “explanation engine” and “simulator” which made up the “primary research goal” of the proposed project. In other words, the developers proposed an applied research project highly focused on technology development. Furthermore, the final CDSS was aimed to be marketed as a product; thus, “commercialization” was also to form a significant aspect of the project. In consideration of these various different parameters, ‘experimentation’ seems a fitting headline for the complex line of work in this AI project.

The predictions generated with the CDSS should additionally be coupled with “clinical aspects”, e.g. “early screening, preventive care, and ultimately diagnosis”, which could be put in place by health professionals in order to prevent detected citizens from potentially being (re)admitted to the hospital as emergencies. Hence, with the CDSS, it would be possible to “screen each individual citizen at very high intervals, determining which people require care and help” and “predict and change patient trajectories”. The CDSS would thus not only help bring down a big bulk of admissions, expected to increase even further, and reduce the great costs of such admissions; it would also, and perhaps more importantly, “position the healthcare sector in a proactive role instead of a reactive role”, thereby supporting new national strategies for the healthcare sector. In this way, the CDSS was envisioned as a key ingredient and future-generating device (cf. Jensen, 2010) in the creation of a specific healthcare future which several actors imagined to become real: the proactive healthcare sector. Researchers in the sociology of expectations have argued that it is by articulating such futures through visions and expectations as forms of “wishful enactments of a desired future” (Borup, 2006: 286) that entrepreneurs contribute to the materialization and performance of such futures.

We suggest that a particular configuration, or version, of the patient
was woven into the CDSS in this envisioned future, namely the potential patient, i.e. a patient who closely resembles a rational citizen with a moral standing who had not (yet) been hospitalized and was willing to do as prescribed in order not to be so, for instance attending smoking cessation courses. Thereby, potential patients were assigned an important role which they – in the envisioned future – had to fulfil in order for the CDSS-experiment to succeed, and for the imagined new automatized data-driven procedures to become real.

Data also had a significant role to play – a rather fundamental requirement in order for AI algorithms to function: "Algorithms are inert, meaningless machines until paired with databases upon which to function" (Gillespie, 2014: 169). Different approaches and methods to preventing (re)admissions and improving the cross-sectoral collaboration for this purpose have been tested over the years (see e.g. Wadman et al., 2009), also including the use of statistical methods (see e.g. Data Study Group team, 2019; National Services Scotland, 2008). Seemingly, it was the commitment and endeavor to take big data and the data-intensive method of deep learning as the starting point for knowing and intervening which constituted the novelty of the AI project and its ‘innovation power’. By doing so, the project provides yet another example of how public services are increasingly organized around data and data-driven computer systems (Kaun & Dencik, 2020), not to mention automated data analysis-methods. It is argued that this data-centered ordering causes a huge transformation of public services, the relation(ship) between citizens and state institutions, and the welfare state in general (ibid.: 2; see also Ruppert, 2019).

The data made available to the AI project in order to construct the ground truth (Jaton, 2017) of the algorithm came from a research database. This database was managed by the regional hospital participating in the AI project and was developed within a framework of another publicly funded research project. The broad aim of this project was shared by the AI project in our study: to use the vast quantities of data collected by national and local authorities – a hallmark of the Nordic welfare state (Tupasela et al., 2020) – to build knowledge on the causes for unplanned (re)admissions. This was in order to focus and strengthen the cross-sectoral collaboration and identify the ‘ideal’ coherent continuity of care across primary and secondary healthcare. According to a description of the database, it consisted of four sources of data: (1) clinical and administrative data from the regional hospital; (2) prescription, telephone/mail, and attendance services-data from GPs; (3) data from municipalities containing information about health and social services provided for citizens as well as address register-data; and (4) data from national records such as CPR-data, electronic patient record-data, and socioeconomic data. Field observations show that the database was continually supplied with more data over the years.

The Enactment & Emergence of the CDSS

We can conceive the CDSS as a “bounded something” (Jensen, 2010: 24) in virtue of the opportunities, aims, and methods stated in the application for the National Innovation Fund, the money granted by this fund, and the data made available to the developers. It was a ‘thing’ that partially existed by virtue of the conceptions that had been made of how it should be performed and thereby yield desirable outcomes and changes. Yet, it was also still a ‘thing’ that had to be further constituted through situated contexts of creation (Hacking, 1983), as it only made up a somewhat diffuse part of the healthcare sector. Its ontology had fairly blurred contours, so to speak.

When the first author began her study of the developers’ work in the AI project, the CDSS had just recently been featured in the national news with the message that engineers at the AI company had made a discovery. In their search for signals in the data, they had run some machine learning experiments and discovered that it was possible to predict, with great accuracy, the likelihood of citizens being hospitalized as emergencies within the next 100 days. Hence, the formation of the CDSS had started in a laboratory manner by examining data for patterns and, consequently, statistical correlations. This was in line with the work plan described in the project description, proclaiming
that the first step in the AI project would be to “develop algorithms to get early results”.

While the first author began to participate in meetings and workshops, it became clear to her that it still remained an open question to the developers regarding how to use predictions to stave off unplanned admissions. Arguably, predictions alone would not prevent potential patients from being hospitalized; they needed to be enacted and performed as a part of well-established practices. Hence, the discovery that it would be possible to predict potential patients within the next 100 days only posed a success on paper. It was still quite unclear how to apply the predictions in practice. In this way, the situation was different from Pickering’s (2018, p. 7) Glen Canyon Dam example, where the engineers’ calculations of the future imply that “there is nothing left to find out in dam design”. In this case, the developers had learned that there was yet much to find out. What they particularly needed to explore was how to “turn what by engineers is regarded as a great algorithm into an algorithm that is applicable in a clinical setting”, as the project manager contended at a meeting. A little later, when the first author started to spend time with the AI company on an everyday basis, it turned out that the developers had begun to consider their data-driven approach somewhat mistaken or at least insufficient. The reconsideration of their standard modus operandi was later expressed by an engineer in an interview:

Informant: You see a lot of stuff being published on what one can do with AI… and 99 percent of these articles come from engineers who start the problems, or, I mean, the projects themselves, without considering: “Is this an actual problem, and is there a sensible intervention or action?” This is why we see that PubMed and ArXiv.org are crammed with trivial AI studies.
Interviewer: Which spring from data because these data exist?
Informant: Yes, technical fascination with problems that can be solved only because data exist. [...] Engineers and data scientists have a tendency to [...] just look at data first because it is data-driven [...] And then we don’t really consider, for instance, what sort of outcome do we actually want to look at. [...] Statisticians cannot just run multiple tests of various kinds because then they can detect correlations in all data. But this is the approach you use as an engineer, in principle, because you just look at data first. And that’s why you find: “Oh, there is a problem here that I can solve! Let’s do an article on that”. But starting with the data is just the wrong place to start out.
(Chief engineer, interview, Feb. 2020)

In order to understand how to prevent unplanned admissions predictively in practice and make the experiment succeed (Hacking, 1983), the project manager and the managing director of the AI company made a guess that there was a need for means and methods centered more directly around users and working practices. “Where in the healthcare sector are the good AI use cases?”, the project manager asked at a meeting, as if such cases merely had to be discovered. Based on her quest for such cases, she arranged a workshop to identify potential users of the CDSS. The first author attended the workshop and became aware that the majority of the participants2 had begun to hypothesize that unplanned admissions would be avoidable if the CDSS was used by GPs. This was based on the discovery that potential patients could be predicted 100 days prior to a hospitalization. Consequently, it was surmised that potential patients assumedly needed to be “turned” before being admitted to the hospital as emergencies. Hence, the CDSS should be integrated in primary healthcare, and most likely general practice, since GPs would typically be the first to see potential patients. Furthermore, GPs had the authority to refer patients to hospitals and

2 Participants came from both the AI company and the regional hospital, and, furthermore, included the project manager from the innovation incubator.
municipality services, and to prescribe medicine, unlike other clinicians in primary healthcare. Again, this idea of ‘turning’ potential patients on the basis of predictions and actions builds on the presupposition that potential patients would act as rational citizens and participants in the proactive healthcare sector.

Soon after the workshop, while the first author began to conduct day-to-day observations at the AI company, a number of new employees suddenly appeared: a design director, a UX designer trained as an anthropologist, and a physician. This marked the beginning of a new phase in the AI project, characterized by another more explorative “mode” of experimentation (Pickering, 2016). This mode was informed by design thinking – a methodology known as a user-centric and iterative process for development and innovation (see Brown, 2008), which has also been conceptualized as a practice of “opportunity creation” (Nielsen et al., 2017). The design director explained at initial meetings that the aim with design thinking was to be open to what the CDSS might become and avoid going into “solution mode”. “It could even be a mobile app”, he suggested. Hence, the idea was, in other words, to assume an “unpredictable becoming” of the emerging CDSS (Pickering, 2016: 4), as it was too early for it to be ‘black boxed’ (cf. Latour, 1987). We suggest that the CDSS now was to be regarded as a multiple object (Mol, 1999), in the sense that it could be many different ‘things’. As long as it would contribute to detecting and, in effect, reducing (un)planned hospital admissions as a marketable AI system, it was less important how – through which practices – it was performed and enacted. It is perhaps telling that the managing director of the AI company at one point noted the difficulty in adopting this almost naïve view on the CDSS: “data really biases one’s ideas”, he claimed.

The design thinking process was introduced to the development team as a non-linear process involving five phases: emphasize, define, ideate, prototype, and test (see e.g. Garcia & Lähdesmäki, 2019). The idea was to repeatedly make observations of clinicians’ work, use such observations for identifying the clinicians’ presumed “pains” and “gains”3, and then brainstorm on possible solutions, i.e. different enactments of the CDSS, followed by prototyping and testing (ibid.: 74). Hence in design thinking, it was ethnographic observations which posed the driving force for understanding the world and the emerging CDSS – not digital data and statistics. The integration of design thinking into the CDSS-experiment thus had two important implications. Firstly, it meant that the AI-based ontology for constituting and developing the CDSS was supplemented by other ontological ideas as a means for making the experiment succeed. Secondly, it meant that the project was now to be performed at other locations outside the AI company, thereby distributing the experiment (Latour, 1983) and expanding the innovation space, i.e. the space of ‘opportunity’ (cf. Nielsen et al., 2017; see also Kjærdsgaard et al., 2016).

The first place where observations were made by the anthropologist, and partially also by the physician and first author, was among GPs with the aim to test the initial hypothesis: potential patients predicted within the next 100 days could be prevented from being hospitalized if GPs used predictions. Hence, the mode of experimentation was not entirely explorative but perhaps more a ‘search for answers’ and, consequently, opportunity. Subsequent sessions to identify ‘pains’ and ‘gains’, brainstorm on possible solutions, and create initial prototypes followed on the basis of the observations brought back to the AI company, i.e. notes and statements from the doctor’s practice. Especially the brainstorming on possible solutions and prototyping played out in a manner we choose to describe as virtual dances of agency (Pickering, 1995). In its original sense, dances of agency are characterized as time-extended struggles between the scientist and obstinate machinery, where the resistance of the machinery is viewed as a kind of “material agency” that struggles with the human agency of the scientist in a dynamic process. But as the lack of so-called “live data” made it highly difficult for the developers to explore and test how the CDSS would perform – what the apparatus would do – they

3 “Pains” are explained as “the negative outcomes from current situations” and “gains” as “positive outcomes that users are trying to achieve” (Garcia & Lähdesmäki, 2019: 74).
A.Henriksen and F. Olesen: Experimenting on the Enactment of Predictive AI

had to imagine it. In this sense, the struggles rather took place in the form of thought experiments, where the developers tried to make the CDSS accommodable, or at least figure out the opportunity for making it accommodable, while constraints would turn up as resistances. For instance, during these experiments, it turned out to pose a significant constraint, which was difficult to get round, that GPs had no reasons not to refer potential patients to the hospital – not even profit.

The idea of the CDSS as a ‘thing’ used by the GP gradually vanished through these virtual struggles, while new hypotheses emerged. This fueled the continued exploration among, first, on-call GPs, and then prehospital emergency nurses and home care workers. Further constraints emerged as resistances while the processes of pains and gains-identification, solution brainstorming, and prototyping were repeated based on observations. For one thing, it was doubted whether it would be possible to ‘capture’, i.e. detect, potential patients on the basis of the data generated in primary healthcare which had turned out to be of poor quality. This was a significant hurdle and prompted the project manager to contend at a meeting that municipalities should require home care workers to work in a more “standardized” manner. That was to ensure they would use digital systems more identically and thereby produce more complete and consistent data, thus providing a proper data infrastructure for the CDSS (Kaun & Dencik, 2020; Bossen & Piras, 2020). Also, it was difficult to imagine how the CDSS could be used to detect potential patients and propose preventive care initiatives when citizens were generally evaluated against individual baselines:

It became evident that habitual aspects really count in primary healthcare; everything is measured against what is habitual. You may have some high, crazy values, but, if they are habitual for that patient, there is no need to react. So, it’s always, you know, measured against what? [...] And that’s what’s making it extra difficult because which actions should we then propose are put in place? (Anthropologist, interview, Feb. 2020)

Consequently, it became difficult to change patient trajectories and enable the ‘turning’ pursued, as the quote suggests. Hence, the idea to predict potential patients within 100 days and prevent them from being hospitalized by enacting the CDSS as a part of primary healthcare gradually vanished during the virtual dances of agency that the developers engaged in. The project manager, meanwhile, learned that a physician executive consultant from the regional hospital, who participated in the AI project, was working on problems that might pose an opportunity. The development team was motivated by this news and initiated observations at the prehospital emergency department of the hospital. The experiment was now further distributed and took a new turn (Latour, 1983). Additionally, the team invited the physician executive consultant to a meeting, where he showed great interest in the emerging CDSS and shared his visions with the developers. He envisioned that patients hospitalized as emergencies could have their condition predicted during preadmission evaluation. Statistics showed that the majority of such patients were hospitalized as patients with unstable conditions, although most of them were discharged as patients with stable conditions. Hence, he claimed that these patients would have had too many examinations carried out compared to their ‘actual’ condition. In this way, it was not only observations that were imported into the experiment (cf. Galison, 1987) but also the interests of other actors (Callon & Latour, 1981). Motivated by a declaration of interests from at least one external actor, the physician executive consultant, further meetings were conducted to retain and translate his interests in the project. Not least, further meetings were conducted to retain and nurture the opportunity to enact the CDSS as a part of the prehospital emergency department (cf. Nielsen et al., 2017). Thereby, the content and context of the experiment were stabilized for the time being.

At the time when the first author left the experiment in February 2020, the CDSS performed more and more as a triage tool with an ontology based on binary logics (stable/unstable), leaving the developers with the feeling that their experiment would soon succeed. In
this regard, one particularly factor to which the developers assigned much significance was their own learning. This was expressed by several developers, including the director of the AI company, but especially also by an engineer involved in the project on a regular basis:

I believe many AI projects start out at the wrong place; they begin by providing data to some engineers who then learn that they can do something with these data: make predictions or classifications. And then they measure how well they do it. If they do it really well, they'll think, "This is awesome; maybe we're just as great at this as clinicians are, or perhaps even better – we've solved a big issue!" But that process should start by considering where there are problems to be solved. What is it that doesn't work? [...] You know, begin with the clinical professionals or such an anthropological approach in order to observe problems [...] Only if you've found things that don't work well [among clinicians], can you start to consider if you can do something about this at all. If yes, then you can start to look in the toolbox, and one of the tools you'll have in that box; that's AI. And only then can you start to consider if AI could help to solve the problem rather than starting with AI. [...] Then the next step will be action… because if the prediction or classification wouldn't lead to an action which will change the workflow or care trajectory, it makes no difference.

(Chief engineer, interview, Feb. 2020)

As the quote suggests, the developers had learned that they might be more successful in enacting predictive AI in the healthcare sector by drawing on designers, user-centric approaches, and ethnographic observations in addition to data, statistics, and engineering, and by exploring the working practices first rather than the digital data produced through such practices. As previously noted, Hacking (1983) contends that experiments are about learning how to use an apparatus or instrument in the right way, and knowing when the experiment succeeds. In our case, we might say that it was about learning how to perform an experiment in order for predictive AI, and thus automatized data-driven procedures, to succeed as a solution to specific healthcare challenges.

Conclusion

In this article, we have studied the experiment carried out in one particular AI project in Scandinavia. This particular project strived to enact a big data and AI-based clinical decision support system (CDSS) to prevent unplanned hospitalizations in pursuit of a specific healthcare future: the proactive healthcare sector. We have argued that the enactment of this CDSS relies on the effective ‘turning’ of what we have called ‘potential patients’, and the presumption that such patients will act as rational citizens. We chose to regard the developing CDSS as a ‘partially existing object’ (Latour, 1999; Latour & Weibel, 2005) with an uncertain ontological status. It was a ‘thing’ in virtue of its stated aims, data, methods, and funding, and the conceptions made of it beforehand, and yet a ‘thing’ that had to be further constituted through situated contexts of creation (Hacking, 1983). That creation was undertaken by a small-scale development team at an AI company followed by the first author in an ethnographic case study over the course of a year.

By studying the gradual enactment and emergence of the CDSS through its construction process, we have shown how it went from being a future proactive device to becoming a triage tool to be integrated into medical triage at the prehospital emergency department. Paradoxically, this was the department that it had previously been envisioned to keep potential patients out of by its imagined enactment in primary healthcare and especially general practice. This imagined enactment was based on a discovery made in data that it would be possible to predict, with great accuracy, the likelihood of citizens being
hospitalized within the next 100 days. It is important to note that the CDSS did not simply 'move' to the prehospital emergency department and thus the hospital by itself. The developers learned en route how to perform the CDSS-experiment so that it could succeed (Hacking, 1983), and, as they learned this, they and the AI company were transformed.

First, the developers learned how to import (cf. Galison, 1987) methods and ideas into the AI project that would change the horizon and modus operandi by which they had attempted to conceive and develop the CDSS initially. This became evident in that they began to draw on a more explorative and user-centric mode of experimentation (Pickering, 2016) informed by design thinking which incorporated ethnographic observations (Brown, 2006). Through so-called thought experiments, the developers engaged repeatedly in ‘virtual’ dances of agency (Pickering, 1995) with a focus on primary healthcare. The choreography of these dances, implying multiple resistances, made the developers see new opportunities, and, perhaps more importantly, it made the developers revise the apparently revisable aims and scope of the AI project in their attempt to accommodate the CDSS to healthcare contexts. Furthermore, the developers learned how to draw the experiment into a wider, situated context, and how to displace the CDSS-experiment (Latour, 1983) so that it could recruit new actors and interests and allow for new opportunities for ‘enactment’ of the CDSS to emerge (cf. Nielsen et al., 2017). Arguably, it was this transformation and learning that enabled the developers in the project to constitute the CDSS as a concrete device in medical triage. Through the transformation and learning processes, the developers changed from being beginners to becoming competent CDSS-operators, knowing what ‘moves’ and ‘actions’ to make in order to enact predictive AI in healthcare.

In the end, we can understand the CDSS-experiment in our study as a dynamic process through which actors strive to re-enact the healthcare sector by means of particular roles assigned to patients, new automated data-driven procedures based on modern AI techniques, and the involvement of private AI companies in decision-making processes, i.e. triage. By studying the AI project and emerging CDSS through an analytical lens focused on ‘experiment’, the article has rendered visible how persons, locations, and procedures had to be changed, revoked, and suspended in order for the AI project to succeed. Thus, the article contributes to showing how ‘social mangling’ (cf. Pickering, 1995) is an essential precondition for succeeding in experiments on the enactment of predictive AI, along with developers’ learning and transformation. Despite their presumed powers, it is not data, methods, and technology per se that make such experiments succeed as an instrumental stance would suggest. Rather, it is what is done with and around such machinery that matters, and through which fuzzy AI algorithms and data-driven computer systems become ‘real’ material devices with concrete uses in particular healthcare contexts.
References


Loukissas, Y. A. (2019). All Data are Local. Thinking Critically in a Data-Driven Society. The MIT Press.


“We Can Be Pioneers”
– Exploring experimental knowledge sharing in an online peer-support forum for non-offending pedophiles

Mikkel Rask Pedersen
PhD fellow, Department of Digital Design & Information Studies,
Aarhus University
Abstract

Online peer-support forums can enable those alienated from society to engage in experimental knowledge sharing practices in the absence of professional and public interaction. This, in turn, helps them manage problems in new ways. However, for pedophiles, it is unsure when peer-support help forum users cope in healthy ways, and when they justify harmful behaviors. The article argues, that exploring the potentials of peer-support forums for pedophiles to help combat child sexual abuse, this produces an analytical challenge to represent the users' experimental knowledge sharing practices without tying them to the very categories they try to escape. To accommodate this challenge, the article proposes to view these practices as politically contingent experiments rather than fixed results, and uses Donna Haraway's concept of situated knowledge to do so. Based on a grounded theory analysis of an observational and ethnographic fieldwork on the Virtuous Pedophiles online peer-support forum for non-offending pedophiles, the article proceeds to analyzes relations and networks between the users' personal experiences and the political situation that surrounds. The results are discussed in relation to classical frameworks of experiments in STS. Finally, experimenting as a mode of practice and analytical tool is argued to contribute to child sexual abuse prevention research, due to its attention to performances, resistances, and constraints, important to understand the challenges of being a non-offending pedophile.

Introduction

We can be pioneers in daring to confront our struggles. We can refuse to accept their labels. We can support each other in keeping our boundaries and our dignity, knowing that is what is going to protect children and make the future better.

Anonymous User, Virtuous Pedophiles forum

Online peer-support forums have been praised for being powerful platforms for self-disclosure and social support seeking (De Choudhury and De 2014, Manikonda and De Choudhury 2017), as well as being valued for their ability to help people experiment with new solutions to their problems in the absence of professional interaction (Mead & MacNeil 2004). Scholars researching child sexual abuse prevention have started to explore how online, anonymous peer-support forums for pedophiles can be utilized based on their ease of access and anonymity to cope with the detrimental psychological effects of living with sexual attraction to minors (Nielsen, Aaskov & Larsen 2020). This is thought to have important implications for child sexual abuse prevention, as mental ill-health could be a key risk-factor in committing abuse (Jahnke & Hoyer 2013, Lasher & Stinson 2017, Cantor & McPhail 2016).

The idea that sexual thoughts about minors inevitably lead to abuse is at best a misunderstanding, but at worst it becomes a self-fulfilling prophecy. When pedophiles first experience sexual thoughts about children in early-to-mid adolescence (Tozdan & Briken 2018), many pedophiles report that this experience is often accompanied by severe psychological distress resulting in loneliness, depression and suicidal thoughts, as they fear their lives could have no positive outcome (Rask Pedersen 2017, Houtepen, Sijtsema & Bogaerts 2016). Owing to a lack of support in society, many pedophiles thus turn to online communities to manage the mental ill-health problems of living with pedophilic attractions (Stevens and Woods 2019). However, turning to online forums may not always have the desired effect in reducing child sexual
M. Pedersen: We Can Be Pioneers

abuse.

Even though the chance to engage with and discuss serious mental health concerns online and anonymously among peers provides an opportunity to circumvent barriers to seeking professional support (Stevens and Wood 2019) and experiment with new solutions to manage the stigma (Holt, Liggett, Holt & Lee 2020), there is a risk that the absence of professional support in such forums might allow room for justifications and rationalizations of child sexual abuse (Halloran & Quayle 2010, Holt, Blevins & Burkert 2010, Malesky & Ennis 2004, D’Ovidio, Mitman, El-Burki & Shumar 2009). Holt et al. (2020) argue that “exposure to and internalization of subcultural norms within pedophile support forums may lead to users accepting pro-offending justifications, beliefs, and denials that may increase risk for contact offenses” (Holt et al., p. 302). Despite the fact that these forums may be powerful platforms to combat the risk of committing abuse, there is therefore also some concern about how to approach this double-edged sword of online peer support for pedophiles, as the forum user’s ‘experimental’ solutions prove no guarantee to be in the child’s favor.

The issue of online peer-support forums, however, does not only present itself in terms of evaluating pros or cons or quantifying the amount of risk of committing child sexual abuse, but also glares at the very instance we suggest researching peer-support forums for pedophiles in an attempt to engage this double-edged sword of online peer support for pedophiles, as the forum user’s experimental solutions prove no guarantee to be in the child’s favor.

The issue of online peer-support forums, however, does not only present itself in terms of evaluating pros or cons or quantifying the amount of risk of committing child sexual abuse, but also glares at the very instance we suggest researching peer-support forums for pedophiles in an attempt to engage this double-edged sword of online peer support for pedophiles, as the forum user’s ‘experimental’ solutions prove no guarantee to be in the child’s favor.

Based on observational and ethnographic fieldwork relating to a peer-support forum for pedophiles called Virtuous Pedophiles, I will introduce, engage, and discuss a theoretical and analytical framework to explore the peer-support interactions of pedophiles as situated and experimental, and argue for the value of doing this. Firstly, I introduce peer support in more detail as a general practice of experimental knowledge sharing defined by the absence of professional knowledge and the use of real-life experience. I argue that a classic critique of experimentation in laboratories from an STS perspective may be adapted to explore how users of such forums experiment with producing and sharing knowledge. I propose to use Donna Haraway’s ‘Situated Knowledges’ (1991) as a theoretical and analytical framework to do this. Specifically, I argue that peer support should be explored as a practice situated in both real-life experience and the political categories in which these experiences occur. I introduce the results from a grounded theory-based content analysis on the Virtuous Pedophiles forum, and build Haraway’s theoretical perspective into analysis, using Marilyn Strathern’s work on relations (2020), hybrids and networks (1996) to visualize the practices of forum users. I then proceed to discuss the
results in two ways: Firstly, I discuss how and why we can regard these practices as experimental. Secondly, drawing on Pickering’s concept of ‘experimenting in the wild’ (2018), I discuss why an emphasis on performance and experimentation is valuable for child sexual abuse prevention.

**Experimental Knowledge Sharing**

The goal of the Virtuous Pedophiles forum is to help provide pedophiles with peer support to help them lead happy, productive lives. Although this forum believes that this can be achieved by living without any physical form of sexual relations to children, other forums for pedophiles aim to achieve the same goals by presenting a wide range of well-rehearsed arguments to justify child sexual abuse (Holt et al. 2010, D’Ovidio et al. 2009). The troublesome or challenging aspect of relying on online forums for people who are sexually attracted to minors to prevent child sexual abuse is that the ability to be anonymous and engage with peers on online forums is effective in supporting and changing attitudes regardless of moral standpoints.

The ambivalence of the practices of online peer-support forums for pedophiles is rooted in the fact that peer support is a resource without any formalized treatment goals, and, importantly, that it is defined by its ability to provide knowledge and solutions outside of medicalized and professional frameworks. Mead and MacNeil describe peer support as a general practice and resource to be used in any treatment program which:

> [...] doesn’t assume a medical definition of the problem and opens us to exploring other ways of thinking about the experience rather than dealing with it (Mead and MacNeil 2006, p. 10).

In Natasja Kingod et al.’s (2017) qualitative but systematic literature review of peer support in online communities and its influence on the daily lives of people living with chronic illnesses, they identify four themes: 1) illness-associated identity work; 2) social support and connectivity; 3) experimental knowledge sharing; and 4) collective voice and mobilization (Kingod, Cleal, Wahlberg & Husted 2017). The particular ability to open up new ways of thinking about experiences is what they identify as experimental knowledge sharing, which they define as:

> Peers [who] exchanged knowledge that emerged from their own experiences of living with illness. This knowledge was not something that could be generated by health care professionals because it arose from real-life experiences and situations (Kingod et al. 2017, p. 18).

As in Mead and MacNeil’s description of peer support, experimental knowledge sharing is defined as a practice somewhat opposed to or in resistance to the (medical) expert systems which typically enroll peers. Notably, Kingod et al. do not describe experimental knowledge sharing as a therapeutic tool in offline settings, as Mead and MacNeil do, but identify this practice as a prominent feature on online peer-support forums. Additionally, Kingod et al. claim that their experimental knowledge sharing is unique because it is generated from real-life experiences and situations. At this point, I should like to return to what I have described as the tensions in online peer-support forums from a child sexual abuse prevention perspective. From a political or societal perspective, there is a motivation to understand how online peer-support practices for pedophiles help them to live without committing abuse – or, alternatively, escalate the risk of committing abuse. As introduced briefly above, this motivation holds its own tension to be able to sort between what helps pedophiles cope and what helps them to justify deviant actions. However, if the value of peer support is to experiment with matters outside of medical definitions and generated

---

1For more detailed discussions of this particular issue, see both Stevens and Wood (2019), Holt et al. (2020), and Nielsen & Aaslov et al. (2020).
from personal situations and real-life experiences, how do we understand these practices as part of a medical framework regarding child sexual abuse and risk, while still taking the experiments seriously? Or, said otherwise, how do we take pedophile’s negotiations of real-life experiences and situations as credible knowledge in itself, while still being taken seriously as producing knowledge relevant for child sexual abuse prevention?

Despite taking a major leap in terms of themes, this question is not unlike discussions of power, representation and experimentation within the field of Science and Technology Studies (STS), and I propose that these discussions can be extended to lend support here. For the purpose of binding STS perspectives to the tension of exploring and representing online peer-support interactions for pedophiles, we might recognize that this tension is a question of how to place agency on who has the ability to define what peer-support is and what it does. Is it the users sharing knowledge in new ways, which does not assume a medical definition of problem who hold this agency, or is it the researchers or scientists with extensive knowledge on child sexual abuse and risk-factors? As I see it, one of the purposes of most STS studies, ranging from the widely used Actor-Network Theory (Callon & Law 1989, Latour 1987), to studies of multiple ontologies in healthcare practice (Mol 2002), and the myth of the cyborg (Haraway 2006), has been to escape an essentialism in which someone or something holds power over others a priori. Instead, a common acceptance in many STS studies is that the knowing world never exists from an objective standard and perspective, but is something that is explored and/or experienced in relation to the wealth of both human and non-human, physical and abstract elements which surround us. Historically, this has been especially prominent in showing that science within laboratories, or experiments, were not exempt from a social/political world, but produced knowledge and facts which were neither neutral nor objective, but bore a relation to both material and political forces (Pickering 1992, Pickering 1995, Latour 1990). One of the most influential works produced by Bruno Latour to date is entitled “Give Me A Laboratory And I Will Raise The World” (1983). This was a fresh critique of the way in which laboratories, scientists and experiments did not in fact show us the facts of the world, but produced facts in our so-called modern world (Latour 1993), and it is the same type of critique I now propose to explore the worlds raised by the users on peer-support forums. Specifically, the performative notion of how knowledge is produced, and that we might learn more about the world from placing more emphasis on how experiments produce facts, than on the facts themselves.

Building on the proposed notion that peer support includes the ability to engage in experimental knowledge sharing sourced from real-life experience and situations, I argue that we might try to dispel the tension of explorations and representation of facts by exploring the forums users’ experimental knowledge sharing analogous to how scientists perform experiments. In other words, if we view experimental knowledge sharing as the performance of producing knowledge, we can learn a great deal about these practices if we do not immediately try to sort the results of these, but look at the performances of the practice itself. Like so, it is my argument that we as researchers can also adapt to an explorative process which does not assume an immediate medical definition of a problem, but remains open to the perspectives of the forum users. Crucially, this perspective does not seek to romanticize the real-life experiences and situation of the forum users as if they are the only experts on peer-support, but accept to explore these experiences seriously within their political/material relations.

I propose that a starting point to adapt this perspective of experimentation as practice between both individual and the political world around them, is well founded in Donna Haraway’s (1991) introduction of the concept of ‘situated knowledges’.

**Situated Knowledges**

Situated knowledges is Donna Haraway’s (1991) attempt to discredit what she calls the ‘god trick’ where we are compelled to believe that there exists an objective position to describe anything. The critique
of the god trick is not unlike the tension introduced in this paper. In peer-support practice and representation, the concept of situated knowledge reminds us that neither real-life experience and situations nor scientific expertise has the power to describe the value of peer support from a neutral or objective position. As Haraway describes it, both relativism and totalization (real-life experiences or medical frameworks) are god tricks, which promise vision from everywhere and nowhere, but neither is less biased that the other. Showing that knowledge is never truly objective and detached from the social and political world, but is always situated, particular and condensed in an antagonistic field of power, her particular political focus on situated knowledges from the peripheries of societies, I believe, generates valuable insights here. I argued in a previous section that viewing users as scientists experimenting with new knowledge is a valuable perspective with a view to easing the tension of exploration and representation. Even though Haraway does not focus directly on experimentation, situated knowledges still hold the same principles of critique, but place emphasis on the value of pursuing situated knowledge from subjugated groups. In the context of this paper, I take pedophiles to be a subjugated group.

Haraway argues that we acquire extraordinary and potent knowledge to describe the social weaves of the world in a more adequate, sustained, and objective way (Haraway 1991, p. 584) when researching subjugated groups. The potency of this knowledge is, that accounts from the peripheries are most likely to deny an all-interpretive account of the world – i.e. a god trick or a reductionist framework. She reminds us, however, that despite of the adequacy, sustainability, and objectivity in the subjugated’s accounts, they are not innocent or produce new god-tricks. Instead they are preferred precisely because of their guiltiness in making visible axes of domination and power in the world around them when fighting against it. Haraway also discusses the danger of romanticizing statements when we research subjugated positions. When ‘seeing from below’, as she describes the way of seeing the world from subjugated positions, we must therefore always be careful to remember that there are no short-cuts to truths here, either. Their knowledge is situated in equally complex ways as any other person (or scientist in a lab), and thus requires the same amount of analytical attention to understand and visualize. When I propose that we visualize the practices of how users of the Virtuous Pedophiles forum see their situation and challenges, it is therefore important to understand that when these users exchange knowledge of their own real-life experiences, these experiences do not appear in a social vacuum and do not pave a short-cut to the truth. In other words, discussing experiences and knowledge outside of expert networks does not make such discussions exempt from the power of the social and political world that surrounds you.

Taking seriously the (subjugated) accounts of non-offending pedophiles on Virtuous Pedophiles, this would then provide adequate and sustained accounts of peer-support practices, but only if we explore this from their point of view. That is, explore their situated knowledges.

To explore situated knowledges more specifically, this is for Haraway to look for “semiotic-material technology linking meanings and bodies” (Haraway 1991, p. 585). Revealing semiotic-material links between real-life knowledge and the political situation that surrounds the self-knowledge of pedophiles therefore involves visualizing the meanings and bodies, which are connected.

Approaching the users’ experimental knowledge sharing with an analytical focus on the semiotic-material relations produced through their self-knowledge, the very first analytical step now is to point out that accessing an online forum through a digital and pseudo-/anonymous avatar is a material relation underlying every single semiotic relation drawn in the users’ knowledge sharing practices. Indeed, the purpose of this article is based on the material-technological construction of a forum in which people who are sexually attracted to minors find both the courage and the resources to engage in these issues. Consequently, the attitudes and semiotic relations in their situated knowledge might seem very different if they were related to the materiality of a microphone on a podium in a physical group.
setting, thus being situated differently. However, although I do not wish to detract from or ignore the existence and semiotic properties of material relations in shaping the users’ situated knowledge, materiality rarely features prominently in the discussions on the forum, as I will show. I will therefore not force an analytical look to material elements and relations in the users’ accounts, other than what I have done by acknowledging the very material semiotic relation of the forum behind the pedophile’s knowledge sharing. For the purpose of this article, a focus on the way in which semiotic relations link meanings and bodies may come to ease the tension of exploration and representation sufficiently.

Following a short introduction to the methodology supporting the observational data, I will take these theoretical perspectives into consideration when analyzing and exploring how the users of the Virtuous Pedophiles forum experiment with new knowledge in order to live non-offending lives.

**Method**

The data for this paper has been collected through ethnographic observation and sampling on the online peer-support forum Virtuous Pedophiles. This data consists of text-based quotes from various discussions and topics on the forum, and was collected primarily in the period August-December 2019. I did not participate or interact in any of the debates on the forum included in the data, but I did – by arrangement with the owners of the forum – make my presence known when I described my project and my observational activities.

The forum has not been observed and analyzed in its entirety because it contains so many categories, topics, and posts that they cannot all be explored using qualitative methods. Even so, the observational activities were directed towards the most popular and relevant topics. The primary topics on which the focus was placed were “Research and Ideas”, “Member Introductions”, “Reaction to the News”, “Ordinary Discussions”, “Life Experiences”, “Request for Support”, and “Keeping Kids and Ourselves Safe”. Approximately 75 topics – containing as few as five posts and as many as 500+ posts – were included in the analysis. Despite a directed observational gaze towards some topics, there were no pre-defined inclusion criteria for analysis other than the criterion of being related to the experience of being a pedophile. Instead, the method for observing the forum was to build a grounded theory by including elements and criteria for analysis on an ongoing basis. If a discussion was included for analysis, it was bookmarked to give prompts between log-ins if new posts were added.

To build a grounded theory with no pre-defined inclusion criteria, I relied on Adele Clarke’s work to sort the data by creating situational maps of prominent human, non-human, and discursive elements to continuously evaluate and provoke analysis of these elements (Clarke 2003). In this process, what started out as somewhat confusing maps of a multitude of actors and elements jotted down in observation with no apparent relations, were categorized slowly into more finished constructs until it became possible to conduct relational analysis.
The first categorized and ordered situational map that supported the analysis of this article appears in figure 1, with the most prominent actors included in this article highlighted for convenience.

![Structured situational map of observations of the Virtuous Pedophiles forum](image)

Collecting data in this way through situational maps is a focused method which should be taken as a first analytical step to navigate the difficulties of exploring peer support for pedophiles in a non-reductionist way by not using pre-defined inclusion criteria. The analysis in this article is thus a next step towards provoking analysis of the relations between all the different actors. However, rather than claiming to build a grounded theory even further, this second-step analysis of relations is done with an attention to situated knowledge to provoke specific attention the subjugated and political position of being a pedophile. I do not mean to derail into new methodological discussions here, but I will add, that I do not think an analytical adaptation to situated knowledges from grounded theory is incompatible, but rather I take it as a comprehensive inclusions of Clarke’s (2003) concept of ‘sites of silence’ to work constructively with all the matters around the pedophiles’ articulations as well.

Continuing, I will present the way in which the data from the forum in the situational map has undergone further analysis to provoke analysis between relations and visualize the users’ situated knowledge.

Even though the users of the Virtuous Pedophiles forum are encouraged to be anonymous, all quotes in the following sections are still presented without user names, and are paraphrased to avoid direct text searches which might reveal the pseudonyms of the users concerned.

**Analysis**

To look for and visualize the semiotic relations that link meanings and bodies with a view to understanding the users’ discussions as experimental and situated practices, I propose to draw on the concepts of ‘relation’, ‘hybrid’, and ‘network’ in Marilyn Strathern’s work. A very first step here is to explain the difference between connections and relations. She describes relations as being the fabric that has an “effect on – or pose problem for – actors far beyond the scope of their connections” (Strathern 2020, p. 8). Entities can thus be related without being connected, and a relation should be understood as existing when the “conception of an entity’s self-referential ‘identity’ becomes modified when that entity is thought of ‘in respect to’ another.” (Strathern 2020, p. 7). However, in my analysis I will not visualize relations between single entities only, but will show how the users’ experimental knowledge sharing feature in multiple relations simultaneously. I propose to call these multiple relational constructs hybrids in political networks. I do
this for two reasons: 1) The logic emerging from the users’ statements is not to be found in the logic of the parts, but in the relations that exist between them, thus making their ‘total’ logic of being a non-offending pedophile a whole or a hybrid consisting of relational activity between several parts (Strathern 1996). As I continue my analysis, I will therefore distinguish between when a relation is a part in itself and when it features as a part in a larger whole. 2) The larger wholes of their logic should, as per Haraway too, be seen in analysis with political categories. Strathern argues:

It is as though the politics that lies within the image of hybridity does not do sufficient analytical work - politics is re-created as though it were also ‘outside’ the analysis of representations (Strathern 1996, p. 520).

Encompassing this, rather than being constructs inside political categories, the hybrid (or the cyborg (Haraway 2006)) of the non-offending pedophile should also be seen in relation to political categories, thus doing sufficient analytical work to show how they change into new constructs together. We might call this seeing the hybrid in a political network, or in a ‘socially extended state’ (Strathern 1996, p. 521). For this reason, the relational work analyzed in the data is visualized in two-dimensional figures, relating parts to wholes on an x-axis, and the wholes in a socially extended state or network on the y-axis. I will present three different analytical figures in this way, one in each of the following themes, which I call: Monsters, Humans, and Peers.

**Monsters**

Believe me, you are not the first, nor the last person to read the Monster Monologue here at Virtuous Pedophiles.

One of the topics discussed most frequently on the Virtuous Pedophiles forum was the idea that pedophiles were monsters or abusers. The initial quote here is a response given to a user who described a lot of self-loathing and felt like a monster, and the respondent even described this sort of narrative as the ‘Monster Monologue’ owing to its frequency. In many instances on the forum, one of the most important ways of managing the stigma attached to pedophilia seemed to involve realizing or helping people to realize that they were not monsters or abusers. Here are some of the quotes reflecting this tendency:

- It really made things easier when I understood the difference between an abuser and a pedophile. Once I could separate the two, I began not hating myself.
- I had my attractions define who and what I was, instead defining them myself and choosing how it should affect my life.
- She [mental health professional] said I wasn’t a pedophile because I never abused anyone. She said pedophiles are monsters who only mean to cause children harm.
- When I became a user here I also let my pedophilia define me. I thought only a monster would be sexual aroused by children. But through the help I’ve gotten here I’m starting to realize that I am much more than a pedophile.

All four users in these quotes use real-life experiences to reflect on the relation between being attracted to minors and being a monster. I believe it may be useful to present the semiotic relations, the wholes, and the network found in these quotes as follows:

<table>
<thead>
<tr>
<th>Monster/Pedophile/Abuser</th>
<th>Being defined</th>
<th>Out of control</th>
<th>Offending</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Me”</td>
<td>Defining</td>
<td>“Much more” / “Not hating”</td>
<td>Non-offending</td>
</tr>
</tbody>
</table>

Figure 2 – Author model #1
When Haraway asks “with whose blood were my eyes crafted”, it would make some sense to say that to see as a non-offending pedophile is to slay and spill the blood of the monster (fied stigma). In this figure, I try to visualize how the political network of the monster feature important relations to the relations between parts of the non-offending pedophile. In the quotes we see how disconnecting yourself from the monster typology involves creating relations that situate yourself as someone who is in control of defining yourself and thus becoming more than the monster. Their "Me" (as depicted in the table), thus become modified when related to or thought of in relation to a monster and an abuser. The testimonies of their self-knowledge as not being a monster, as something more, as a person capable of defining themselves, is thus shown not in a logic in itself, but as a situated practice in a political network, exemplified particularly well in the following quote:

It really made things easier when I understood the difference between an abuser and a pedophile.

The network of relations shown in the figure is an example of an experimental and situated practice of being a non-offending pedophile in the Virtuous Pedophiles forum. I will continue to consider more quotes and a new network, to add to the complexity of this larger vision.

Humans

I’m a good person. I try every day to do right, but I’m still totally alone and I don’t have anything to show for any of it.

As this quote suggests, being a non-offending pedophile is a lonely and unrewarding existence, as you have nothing to show for doing nothing. I will argue here that experimental and situated knowledge sharing on the Virtuous Pedophiles forum may, in fact, be a way of visualizing the performance of doing nothing. Because you cannot find a connected performance to your missing actions, the performance of a non-offending pedophile may still be related to the action of a choice. Here are some quotes to illustrate this point:

This is the difference between a monster and a human being. All of humanity are tempted, but we can evaluate whether that temptation is safe or not. We can choose whether or not we want to abuse another person. A monster would not experience this difference. They would always act in self-interest.

Your morality isn’t defined by what you want, but what you do. To think you control your feelings is, well, bluntly, ridiculous thinking created by ridiculous institutions that use guilt as ways to control you. Morality is a choice of action.

Homosexuality was seen the same way as pedophiles are today. They assumed it was a sickness that made men rape boys. But what did they do? They taught the public and changed the way people saw homosexuality. They made people see that homosexuals are not violent people, but just like with straight people, some are rapists. We need
to do the same. I am a pedophile and not a rapist.

I guarantee you there is an abundance of experience and knowledge here that you can dig into whenever you feel the need. I think what you can learn about your ‘challenges’ will surprise you. And please never feel ashamed, because what you are is not who you are. Always remember that.

As in the previous section, I will consider the hybrid of the non-offending pedophile bottom row, and its socially extended network top row. Here, the political whole of a rapist:

<table>
<thead>
<tr>
<th>Actions</th>
<th>Responsible</th>
<th>Temptation</th>
<th>Rapists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoughts</td>
<td>Not Responsible</td>
<td>Choice</td>
<td>Normal</td>
</tr>
</tbody>
</table>

Figure 3 – Author model #2

This time in the bottom row, we see how the users relate elements such as ‘thoughts’, ‘responsibility’ and ‘choice’ into what I analyze as a related whole of being ‘normal’. As before, the total logic of this practice of forming a larger vision of a non-offending pedophile is found in its situated practice when it is related, or extended, up against an opposite. Each part’s logic – that being ‘thought’, ‘non-responsibility’, and ‘choice’ - are all modified, mobilized and visualized as elements in a performance of seeing as- and being a non-offending pedophile when it is related to the ‘actions’, ‘responsibilities’ and ‘temptations’ of the abuser/rapists. In this way, seeing yourself as a non-offending pedophile means accepting that you are responsible for your choice of not acting on your thoughts by relating the choice not to act to a choice of morality, which is then extended as a part in a larger network of being a human rather than a rapist/monster. The users are thus regarded not as non-offending pedophiles, but as moral human beings – which (as indicated in the quote at the start of this section) is very useful because it gives them something to show for their struggle.

I will consider one last analytical network.

Peers

At the beginning of this article, I argued that many pedophiles report a sense of psychological distress when acknowledging that their sexual feelings probably categorize them as pedophiles. However, despite this distress, many pedophiles do not seek any help. Studies indicate that pedophiles do not refrain from seeking help because they do not want it, but because they do not trust the relevant professionals to have sufficient knowledge to help them, or because they fear that they might be wrongfully reported to the authorities (Kramer 2011).

Even though laws differ from country to country and Virtuous Pedophiles is an international forum, mandatory, national reporting laws are seen as a prominent risk for many users, as they fear having their confidentiality violated if they simply state their feelings (and not their actions):

No one will try to get help if just saying that you are a pedophile will have you treated like an abuser.

Therapy is not an option for me, even though I really need it. I don’t think there is help to find where I live, and even if I could find someone, mandatory reporting laws would make it hard to be honest.

The biggest problem of being attracted to minors is that mandatory reporting laws discourage us to get help. We will be reported and wing up in a register, simply for wanting to get professional help.

The fear of getting help is not knowing what is safe to
discuss and always being alert.

Seen from the perspective of the non-offending pedophile user on Virtuous Pedophiles, these quotes already seem to paint a picture of a somewhat antagonistic relationship to mental health professionals owing to their obligations (relations) to mandatory reporting laws or to their stigmatic preconceptions. Considering the different elements in the quotes above, a political network of mental health professionals for the users could look like this:

<table>
<thead>
<tr>
<th>Professionals</th>
<th>Pedophile = Abuser</th>
<th>Registers</th>
<th>Not safe/Being alone</th>
<th>Alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peers</td>
<td>Reassurance</td>
<td>No judgement</td>
<td>Happy/Gratitude</td>
<td>Community</td>
</tr>
</tbody>
</table>

Figure 5 – Author model #4

I introduce the top row initially here, as I want the reader to understand this frequent theme and situated challenge before I present the next and final quotes reflecting the way in which users discuss being on a forum with peers:

I’m happy this forum exists and that young people are finding it. I was completely alone with my thoughts, and too scared to reach out to anyone for over 30 years. That was the hardest part. Finally finding this, totally by accident, seeing others like myself actually saved me from doing things I would regret.

To see you are not the only one and to be able to voice your concerns without judgement. Knowing you’re supported by so many others who go through the same emotions, weaknesses, victories, challenges. I am very grateful for that reassurance.

This forum was and still is the only place I can define my sexuality. Honestly, I thought that liking early teen boys while not wanting to hurt them was a completely unique attitude, so I wasn’t able to understand or fit it anywhere until I came here.

Considering these quotes and the quotes included previously, the parts and wholes and their relations all together could look like this:

In these examples taken together, we see how users relate parts such as ‘peers’, ‘reassurance’, ‘non-judgmental attitudes’, ‘happiness’ and ‘gratefulness’ to belonging to a community. The two-axis relational work that I portrayed here is, however, different from the other examples in the way that the different parts from each row were not directly connected in the quotes, but imposed by analysis. I still relate the political network of barriers in help-seeking to the users’ ways of relating the digital avatars of their peers to ‘reassurance’, ‘non-judgementality’, and ‘happiness’ for its direct oppositions to ‘being misunderstood’, ‘registers’, and ‘fear’. Even though this relational construct is not derived directly from user discussions, I argue that the semiotic relations between ‘national borders’, ‘laws’, and ‘registers’, and ‘peers’ and ‘anonymity’ are still relations with powerful implications for the ways in which non-offending pedophiles (peers) are seen on the Virtuous Pedophiles forum.

Discussion

There is a motivation to try to understand how non-offending pedophiles understand their situation and the challenges they face to further ways of preventing child sexual abuse. However, to do so was argued as providing a tension in how to explore these perspectives on online peer-support forums without assuming medical definitions of child sexual abuse and risk which they try to escape. I have argued that exploring experimental knowledge sharing among the users on Virtuous Pedophiles to try to understand their own perspectives on their situation and challenges, we should not mistake the emergence
of real-life experience outside of professional network as emerging in a social vacuum. Instead, I argued to adapt an analogous perspective to the forum users as experimenting scientists to see their practices as situated knowledges to include the actors surrounding the pedophile’s experiences for its potency to provide more adequate, sustained and objective accounts from the users.

In the analysis of data collected through observational fieldwork on the Virtuous Pedophiles forum, I have covered three different themes to show how the users’ quotes revealed relations to a multitude of actors, including political categories, which together formed a larger vision of the situated practices of being a non-offending pedophile. Using Strathern’s work on relations, hybrids, and networks, I have shown how their situation and problems were given meaning and modified through their relation to – and not to – elements situating their knowledge. Thus, despite being emergent from their own experiences and experimental from challenging common understandings, their knowledge was not free and untied, but made meaningful through relations to the categories and stigma situating them – i.e., the common understandings. Visualizing how the users on Virtuous Pedophiles saw their situation and challenges of living as non-offending pedophiles through this analytical lens, the results of this endeavor can be summarized in three key points: 1) That seeing as a non-offending pedophile is to proliferate relations from yourself to attributes opposite of a monster and an abuser, thus defining yourself rather than being defined. 2) That you are not responsible for your thoughts, and the action of making a choice not to act on your thoughts proliferate relations to a moral human being, capable of being a control, opposite of monsters and rapists, who act on their temptations. 3) That peers on the online forum come to be important resources for them to achieve a positive outlook on life, as they symbolize the opposite of the fears and insecurities in seeking help from mental-health professionals.

These results lead to two conclusions: First and foremost, I believe they provide a methodological, theoretical, and analytical example of how to approach experimental knowledge sharing to manage stigma on online peer-support forums for pedophiles, navigating a potentially dangerous dualistic separation between professional and political categories and the real-life experiences of pedophiles. Furthermore, I believe the results show the value of regarding the practices of marginalized groups as experimental. I will try to develop this last point here.

Exploring and researching marginalized groups – whether this is done in a context of peer support and experimental knowledge sharing or not – the power and interplay between identity and categories is something to behold critically to understand their realities. This, I believe, is a point which has particular resonance with Ian Hacking’s work on the ‘looping effect’ and how researchers and engines of discovery ‘make up people’ (Hacking 2006). In Hacking’s terminology, the motivation to understand the potential of peer-support practices in child sexual abuse prevention could be described as an engine of discovery to medicalize people and – in turn – regarding a group of people as being defined by definite properties. Hacking means not to criticize the truth in these practices of discovery, but observes that we tend to think that the targets we aim to discover are fixed and standing still, when in reality they are moving. As part of a looping effect, we might therefore also pay attention to resistance as the process in which the people who are: “medicalised, normalised, administered, increasingly try to take back control from the experts and institutions, sometimes by creating new experts, new institutions.” (Hacking 2006, p. 8). In this way, the situated practices visualized in the results could also be framed in terms of resistance to experts and institutions, but importantly, as moving, creating, and relating knowledge in new ways. Experimentation as a concept of a mode of practice, I then believe to be a way to fix our mode of discovery to a moving target, accepting the emerging interplay between people and categories. Seeing experimentation as a practice of emerging processes between identity and expert categories or standards is supported additionally by considering Charis Cussins’ (1996) ethnographical exploration of interactions between female patients in infertility clinics and reproductive technologies. Cussins discusses how the women in infertility clinics engage in what
she describes as an “ontological choreography” between their own personhood and objectifying their infertility. In this choreography, the women forge a functional zone in which they objectify themselves to the standards and cultures of the infertility clinic to bring about desires change in their identities. The power of this analysis, is when Cussins show empirically how the women’s agency to get pregnant, i.e. real-life experiences and situation, is not reduced to, but inseparable from the objectifying standards of the clinic. Similar to situated knowledges, but with a sharper attention to agency, Cussins show how to regard practices from (marginalized) groups as motivated and strategic, but still choreographed in specific objectifying relations to the standards and cultures around them to bring about desired agendas.

Experimentation in this regard then come to not only mean arbitrary interplay between subject and power, but in fact as a (moving) process of gaining agency through subscribing to standards and norms. Whether in resistance to, or in choreography with, regarding and exploring the pedophile forum user’s practice as experimental I value precisely for the attention the inseparability to and interplay with the political world of child sexual abuse, which is shown in the users’ interaction. Following Hacking and Cussins, this perspective is then important if we are to understand both the production of identity and the realization of agenda. Another point to exemplify the analytical power of looking at experimental practices in political fields may be lend from Gomart and Hennion’s (1999) analysis of social attachment in music amateurs and drug users. Regarding practice concerning knowledge and identity, Gomart and Hennion states that when subjects enter a ‘dispositif’, i.e. a powerful and political environment meant to structure action and knowledge, we might hold analytical priority to the ‘constraints’ in this environment as the generous aspects which create existence and initiate transformation, rather than something which reduces it (Gomar and Hennion 1999, p. 221). By looking for how these constraints – i.e. the medicalizations (Hacking) and the objectifying practices (Cussins) - are managed and experimented with reflexively by the people we study, Gomart and Hennion argues, how they in a peculiar way come to write a sociology for us. When the users studied here then experiment with the constraints of being a non-offending pedophile, such as abusers and monsters, these reflexive practices (or events as Gomart and Hennion would call it), is an important analytical source to understand how to live as a non-offending pedophile, rather than only looking for action of acting on their attractions or not.

Experiments, I argue, is thus as important explorative and analytical contribution to study marginalized groups for its attention to performances in situated knowledge, resistance, constraints, and ontological choreographies to visualize how the vision, identity, and agenda of being a non-offending pedophile is brought into existence.

My final and concluding point in this paper is to explain why I also believe it is important to adapt a focus on experimentation to understand child sexual abuse prevention specifically. Even though I have used experimentation as a concept to describe the performative and dialectic nature of producing and exchanging knowledge, it also often means producing something new. Even though this paper has been critical of the initial sense of seeing experimental knowledge sharing as generated from real-life experience and situation, I still believe it is important to call the peer-support interactions of pedophiles experimental in order to underline that these practices do in fact have a unique value of creating new ways of seeing. As new, complex emergent systems of how to live in ways that resist the inevitability of becoming an abuser. Finding and experimenting with new solutions to your problems in the midst of strands of knowledge from your own experiences and political categories we might call what Pickering (2018) describes as ‘experimentation in the wild’.

In Pickering’s work “Poiesis in Action”, he discusses a framework to engage what sort of knowledge is the alternative to “Strong Knowledge”. His argument is that the alternative is in fact not knowledge at all, but ‘poiesis’:

a hybrid assemblage of mangle-ish performance and knowledge, but within a gestalt defined by the former (Pickering 2018, p. 3).
When you experiment in the wild, the sort of knowledge produced is not a detached and universal knowledge, but something ancillary to the very performances between the actors from which it emerges. Performances which he calls ‘dances of agency’ (Pickering 1995). Thus, when he describes poiesis as an alternative to “Strong Knowledge”, the center of gravity in this hybrid assemblage is the performances, and not the strands of knowledge identified in them. With a focus on performance, poiesis thus actively explores emergence, whereas “Strong Knowledge” conversely seeks to hide it (Pickering 2018). Pickering makes a compelling argument of poiesis in action, using examples of erosion control in Japan, dam-building on the Colorado River, and natural farming in Japan to show how they each make decisions based on performances in ‘dances of agency’, and not by doing with any particular knowledge per se. We might therefore with poiesis in action and ‘experimenting in the wild’ speak of a way of doing in the world without knowledge.

In this discussion I have proposed that the users’ interactions and ways of being a non-offending pedophiles came to show experimental and situated performances through resistance and constraints to powerful political categories. Performances of defining oneself, choosing morally, and achieving a positive outlook on life. These may each represent new ways of being a non-offending pedophile – or, poiesis in that they are hybrids assemblage of mangle-ish performance and knowledge. Though the users’ experimental knowledge sharing was related through strands of knowledge as showed in the analysis - which was what made them situated - it is the way in which these strands emerge and are mobilized into performances that I argue might both conceptualize it as experimental, and be a contribution to child sexual abuse prevention research. One user was quoted as saying:

I’m a good person. I try every day to do right, but I’m still totally alone and I don’t have anything to show for any of it.

Finding ways of being in the world and mobilizing strands of knowledge into performances is a performance of managing the stigma and having something to show for it. Therefore, it should be a valuable point of attention in their situated knowledges. That situated knowledge describes their positionality, but it is the performances emerging between the dances of agency of elements in their situated knowledges that show how they see as non-offending pedophiles. Situated knowledge is then a way of directing attention to strands of knowledge to navigate a way of representing the users’ discussions, but in accordance with Pickering’s discussion of poiesis, also a way of experimenting in the wild to make their real-life experiences count. I will consider the initial quote of this paper once more:

We can be pioneers in daring to confront our struggles.
We can refuse to accept their labels. We can support each other in keeping our boundaries and our dignity, knowing that is what is going to protect children and make the future better.
Anonymous User, Virtuous Pedophiles forum

It is indeed wild to experiment with knowledge sharing when you are a pedophile. However, it is when we understand that another result of their experimental knowledge sharing is to mobilize performances in this situated wilderness between political categories and real-life stigmatic experiences that we come to direct our attention to the real value. Not knowing, but doing. Not seeing as a non-offending pedophile, but acting as one. To be pioneers, and to experiment with new ways of resisting the stigma, while maintaining their boundaries and dignity to make a better future for children and themselves.
References


