



Volume 9, Issue 7, July 2019

Notional Machines and Programming Language Semantics in Education (Dagstuhl Seminar 19281) <i>Mark Guzdial, Shriram Krishnamurthi, Juha Sorva, and Jan Vahrenhold</i>	1
Data Series Management (Dagstuhl Seminar 19282) <i>Anthony Bagnall, Richard L. Cole, Themis Palpanas, and Konstantinos Zoumpatianos</i>	24
Values in Computing (Dagstuhl Seminar 19291) <i>Christoph Becker, Gregor Engels, Andrew Feenberg, Maria Angela Ferrario, and Geraldine Fitzpatrick</i>	40
Mobile Data Visualization (Dagstuhl Seminar 19292) <i>Eun Kyoungh Choe, Raimund Dachsel, Petra Isenberg, and Bongshin Lee</i>	78
Secure Composition for Hardware Systems (Dagstuhl Seminar 19301) <i>Divya Arora, Iliia Polia, Francesco Regazzoni, and Patrick Schaumont</i>	94
Cybersafety Threats - from Deception to Aggression (Dagstuhl Seminar 19302) <i>Zinaida Benenson, Marianne Junger, Daniela Oliveira, and Gianluca Stringhini</i>	117

ISSN 2192-5283

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at <http://www.dagstuhl.de/dagpub/2192-5283>

Publication date

February, 2020

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

License

This work is licensed under a Creative Commons Attribution 3.0 DE license (CC BY 3.0 DE).



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Aims and Scope

The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.

In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Editorial Board

- Gilles Barthe
- Bernd Becker
- Daniel Cremers
- Stephan Diehl
- Reiner Hähnle
- Lynda Hardman
- Oliver Kohlbacher
- Bernhard Mitschang
- Bernhard Nebel
- Albrecht Schmidt
- Wolfgang Schröder-Preikschat
- Raimund Seidel (*Editor-in-Chief*)
- Emanuel Thomé
- Heike Wehrheim
- Verena Wolf
- Martina Zitterbart

Editorial Office

Michael Wagner (*Managing Editor*)
Jutka Gasiorowski (*Editorial Assistance*)
Dagmar Glaser (*Editorial Assistance*)
Thomas Schillo (*Technical Assistance*)

Contact

Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
<http://www.dagstuhl.de/dagrep>

Digital Object Identifier: 10.4230/DagRep.9.7.i

Notional Machines and Programming Language Semantics in Education

Edited by

Mark Guzdial¹, Shriram Krishnamurthi², Juha Sorva³, and Jan Vahrenhold⁴

1 University of Michigan – Ann Arbor, US, mjguz@umich.edu

2 Brown University – Providence, US, sk@cs.brown.edu

3 Aalto University, FI, juha.sorva@iki.fi

4 Universität Münster, DE, jan.vahrenhold@uni-muenster.de

Abstract

A formal semantics of a language serves many purposes. It can help debug the language's design, be used to prove type soundness, and guide optimizers to confirm that their work is correctness-preserving. Formal semantics are evaluated by several criteria: full abstraction, adequacy, soundness and completeness, faithfulness to an underlying implementation, and so on.

Unfortunately, we know relatively little about how non-experts, such as students, actually employ a semantics. Which models are they able to grasp? How useful are these as they explain or debug programs? How does their use of models evolve with the kinds of programs they write? And does studying these kinds of questions yield any new insights into forms of semantics?

This Dagstuhl Seminar intended to bridge this gap. It brought together representatives of the two communities—who usually travel in non-intersecting circles—to enable mutual understanding and cross-pollination. The Programming Languages community uses mathematics and focuses on formal results; the Computing Education Research community uses social science methods and focuses on the impact on humans. Neither is superior: both are needed to arrive at a comprehensive solution to creating tools for learning.

Seminar July 7–12, 2019 – <http://www.dagstuhl.de/19281>

2012 ACM Subject Classification Social and professional topics → Computing education, Theory of computation → Program semantics

Keywords and phrases computing education research, formal semantics, misconceptions, notional machines

Digital Object Identifier 10.4230/DagRep.9.7.1

Edited in cooperation with Philipp Kather



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Notional Machines and Programming Language Semantics in Education, *Dagstuhl Reports*, Vol. 9, Issue 7, pp. 1–23

Editors: Mark Guzdial, Shriram Krishnamurthi, Juha Sorva, and Jan Vahrenhold



DAGSTUHL
REPORTS

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany


1 Summary

Mark Guzdial

Shriram Krishnamurthi

Juha Sorva

Jan Vahrenhold

License  Creative Commons BY 3.0 Unported license

© Mark Guzdial, Shriram Krishnamurthi, Juha Sorva, and Jan Vahrenhold

A formal semantics is often intended as a tool to comprehend the behavior of a language or other system. Semanticists assume, for instance, that programmers can use a semantics to understand how a particular program will behave without being forced to resort to deconstructing the output from a black-box evaluator. Indeed, different semantic models vary in what aspects of program behavior they highlight and suppress.

Every semantics has an intended audience. Formal semantics typically assume a readership with high computing or mathematical sophistication. These therefore make them inappropriate for students new to computing. What forms of description of behavior would be useful to them? In computing education, the term *notional machine* is often used to refer to a behavior description that is accessible to beginners.

Our meeting therefore focused on what we know, and what we need to learn, about notional machines. In particular, we studied and discussed:

- Different formulations of notional machines for a variety of languages.
- The distinction between a general description of behavior, independent of a specific program, and the explication of behavior of a specific program. We argued for the value of having both the general and the specific, since learners might need to shift between the two.
- The different forms that a notional machine can take, and their styles: [MARK fill in]
- The many analogies employed in notional machines, with their respective strengths and weaknesses.
- The different forms of theories that apply to generating and understanding notional machines, including cognitive and social.
- Analogies to notional machines in other domains, from models in physics to rulebooks in board games.

We accomplished most of our stated goals: to bring together the semantics and education communities (though with much greater representation from the latter than the former); to create tutorials to educate each on the knowledge and methods of the other; and to formulate interesting examples. While there did not appear to be many long-standing “open questions”, and there was not enough time to engage in editing Wikipedia, groups did organized community-wide activities (such as surveys to be conducted at upcoming conferences) and large banks of research questions (which are concrete and valuable outcomes that we had not anticipated). In sum, we believe the seminar successfully accomplished its overall stated goals.

2 Table of Contents

Summary

<i>Mark Guzdial, Shriram Krishnamurthi, Juha Sorva, and Jan Vahrenhold</i>	2
--	---

Overview of Talks

Drawings of Notional Machines from Secondary School Teachers <i>Brett A. Becker</i>	6
Sketching Notional Machines with Meaning <i>Kathryn Cunningham</i>	6
Using the Structure Behavior Function Framework to Understand Learning of Computer Programming <i>Kathryn Cunningham and Mark Guzdial</i>	7
Notional Machines and Research from the 1970s and 1980s <i>Benedict du Boulay</i>	8
Runestone Interactive Ebooks with Adaptive Parsons Problems <i>Barbara Ericson</i>	8
Presenting Name/Value Mappings in Notional Machines <i>Kathi Fisler</i>	8
Empirical Studies <i>Robert L. Goldstone</i>	9
Making Programming Languages to Meet a Greater Need <i>Mark Guzdial</i>	9
Conceptual Change in Learning to Program <i>Matthias Hauswirth</i>	10
What Do Students “See” in Computing Contexts? <i>Geoffrey L. Herman</i>	10
Reading Code Aloud <i>Felienne Hermans</i>	10
Sensing and First Data <i>Matthew C. Jadud</i>	11
Giving Feedback and Hints in (Haskell/Java/...) Programming Tutors Based on Comparing Model Solutions to Student Solutions <i>Johan Jeuring</i>	11
Philosophical Concept Analysis in PL or SE or CSE or ... <i>Antti-Juhani Kaijanaho</i>	11
Towards Algorithm Comprehension <i>Philipp Kather and Jan Vahrenhold</i>	12
Code and Cognition Lab <i>A. J. Ko</i>	12
Language Levels <i>Shriram Krishnamurthi</i>	12

Explicit Programming Strategies	
<i>Thomas D. LaToza</i>	13
Conceptual Change & Knowledge in Pieces (KiP)	
<i>Colleen Lewis and Matthias Hauswirth</i>	13
Concrete Notional Machines	
<i>Colleen Lewis</i>	13
Reference-point Errors: Slips? or Misconceptions of the Notional Machine?	
<i>Craig Müller</i>	14
Notional Machines for Everyday Life	
<i>Andreas Mühling</i>	14
Making a Causal Diagram for Learning Programming	
<i>Greg Nelson</i>	15
Pointer Concepts in C	
<i>Andrew Petersen</i>	15
Semantics Tutorial	
<i>Joseph Gibbs Politz</i>	15
Activity Theory	
<i>R. Benjamin Shapiro</i>	16
Stuff We Wish We Knew (About Notional Machines)	
<i>Juha Sorva and Otto Seppälä</i>	16
Revisiting Two Past Publications through the Lens of Notional Machines	
<i>J. Ángel Velázquez Iturbide</i>	17
Working groups	
Breakout Group 1	
<i>Geoffrey L. Herman and Philipp Kather</i>	17
Breakout Group 2	
<i>Craig Müller and Franziska Carstens</i>	18
Breakout Group 3	
<i>Brett A. Becker, Neil C. C. Brown, Paul Denny, Rodrigo Duran, Robert L. Goldstone, Antti-Juhani Kaijanaho, Greg Nelson, Carsten Schulte, Otto Seppälä, and Steven A. Wolfman</i>	18
Breakout Group 4	
<i>Kathi Fisler and Kathryn Cunningham</i>	19
Categorizing Notional Machines and their Representation or Visualization	
<i>Franziska Carstens</i>	19
Instructional Design for Notional Machines	
<i>Barbara Ericson, Robert L. Goldstone, Matthias Hauswirth, Antti-Juhani Kaijanaho, Greg Nelson, André L. Santos, and Anya Taftiovich</i>	20
Concept Analysis for Notional Machines	
<i>Antti-Juhani Kaijanaho, Thomas Ball, Markus Müller-Olm, and Juha Sorva</i>	20
Notional Machines for Everything	
<i>Joseph Gibbs Politz, Mark Guzdial, and Philipp Kather</i>	20

Notional Machines for Scratch and Python
Otto Seppälä, Thomas Ball, Titus Barik, Brett A. Becker, Paul Denny, Rodrigo Duran, Juha Sorva, and J. Ángel Velázquez Iturbide 21

Notional Machines and Simulation
Steven A. Wolfman, Kathryn Cunningham, Kathi Fisler, Greg Nelson, and Jan Vahrenhold 22

Participants 23

3 Overview of Talks

3.1 Drawings of Notional Machines from Secondary School Teachers

Brett A. Becker (University College Dublin, IE)

License  Creative Commons BY 3.0 Unported license
© Brett A. Becker

I presented a poster consisting of drawings of notional machines from 9 Irish Secondary School Teachers (senior cycle – teaching students 15-18 years of age). These teachers were enrolled in a 30 ECTS postgraduate diploma in Educational Studies in Computational Thinking. Specifically, these teacher-students were in my course “How Computers Work” which is a post-programming basic architecture course. At the point that the teacher-students drew their notional machines they had completed course material on: Number systems; Logic; Boolean Algebra; Von Neumann Architecture; CPU; Memory; and the Bus. They had completed tutorials on: binary / decimal conversion (by hand and in Python); creating simple logic gates in a simulator (e.g. half-adder); “anatomy of a computer” where we opened up a desktop computer; the “Little Man Computer” which is a simulator that “has many of the basic features of a modern computer that uses the Von Neumann architecture”; and benchmarking their own laptops with software. I then gave the teacher-students a 20 minute primer on notional machines. Then they drew their own depictions of what a notional machine looks like to them. There were many interesting observations made by those at Dagstuhl who viewed these depictions. It is likely that more can be learned by having students produce visualisations of their mental models and their concepts of notional machines.

3.2 Sketching Notional Machines with Meaning

Kathryn Cunningham (University of Michigan – Ann Arbor, US)

License  Creative Commons BY 3.0 Unported license
© Kathryn Cunningham

Prior research has shown that sketching out a code trace on paper is correlated with higher scores on code reading problems. Why do students sometimes choose not to draw out a code trace, or if they do, choose a different sketching technique than their instructor has demonstrated? In this study, we interviewed 13 CS1 students retrospectively about their decisions to sketch and draw on a recent programming exam. When students do sketch, we find that their sketching choices do not always align with the way that experts illustrate execution of the notional machine. Sketching choices are driven by a search for a program’s patterns, an attempt to create organizational structure among intermediate values, and the tracking of prior steps and results. When novices don’t sketch, they often report that they’ve identified the goal that the code achieves. In either case, novices are searching for the functionality of code, rather than merely tracing its behavior. Student sketches suggest new notional machine visualization approaches that integrate the meaning of code with code behavior. Title – Cognitive Complexity of Programs and Notional Machines- Rodrigo Silva Duran Instructional designers, examiners, and researchers frequently need to assess the complexity of computer programs in their work. However, there is a dearth of established methodologies for assessing the complexity of a program from a learning point of view. In

this poster, I present theories and methods for describing programs in terms of the demands they place on human cognition. More specifically, I draw on Cognitive Load Theory and the Model of Hierarchical Complexity to extend Soloway's plan-based analysis of programs and apply it at a fine level of granularity. The resulting framework of Cognitive Complexity of Computer Programs (CCCP) generates metrics for two aspects of a program: plan depth and maximal plan interactivity. Plan depth reflects the overall complexity of the cognitive schemas that are required for reasoning about the program, and maximal plan interactivity reflects the complexity of interactions between schemas that arise from program composition. To generate the aforementioned metrics, instructors need to first supply a concrete program written in a given programming language. A second input for the model is the expected prior knowledge of a given learner, measured in terms of what kinds of plans have been automated. Third, a notional machine will describe the plans and elements from the programming language required to comprehend the program, how the semantics describe the interaction among elements and to which level of detail and abstraction the plans are represented by the learner. In this poster we explore questions regarding the design of notional machines aimed to different audiences, how much detail and abstraction a notional machine should have when aimed to a particular audience and what formats could be used to communicate a notional machine more clearly.

3.3 Using the Structure Behavior Function Framework to Understand Learning of Computer Programming

Kathryn Cunningham (University of Michigan – Ann Arbor, US) and Mark Guzdial (University of Michigan – Ann Arbor, US)

License © Creative Commons BY 3.0 Unported license
© Kathryn Cunningham and Mark Guzdial

Over the past few decades, many researchers have proposed that designed devices can be understood in terms of their structure (what they are made of), their function (why they were designed), and their behavior (how they work). We unified past definitions of the Structure Behavior Function (SBF) framework, and then applied the framework to the understanding of computer programs. We defined the structure of a program to be the program's syntax and its programming plans; we defined the function of a program to be its purpose in natural language; and we defined the behavior of a program to be the way that it executes on a notional machine. In the SBF framework, the ability to transition between structure, behavior, and function is crucial to the design process. In programming education, we explicitly teach the transition from structure to behavior through tracing exercises, but the transition between behavior and function is not typically taught. We interpreted three theories of programming knowledge using the language of the SBF framework, showing that SBF can organize and relate several areas of computing education research.

3.4 Notional Machines and Research from the 1970s and 1980s

Benedict du Boulay (University of Sussex – Brighton, GB)

License  Creative Commons BY 3.0 Unported license
© Benedict du Boulay

This presentation covered four areas: six influential pieces of work of the period, difficulties of learning programming, notional machines and conclusions. Feurzeig and Papert’s work on Logo was an example of an influential piece of work. The difficulties of learning programming covered four areas: orientation, notional machines and notation, structures, and pragmatics. For each area a couple of research papers from the period were identified. Notional machines were divided into two kinds: stories and (machine generated) representations. The conclusions offered a link between notional machines and semantics via the route of formally specifying learner misconceptions.

3.5 Runestone Interactive Ebooks with Adaptive Parsons Problems

Barbara Ericson (University of Michigan – Ann Arbor, US)

License  Creative Commons BY 3.0 Unported license
© Barbara Ericson

I have been creating interactive ebooks for Advanced Placement Computer Science using principles from educational psychology: worked example plus practice, multiple modalities, and adaptive learning. Advanced Placement (AP) Computer Science courses are intended to be equivalent to college-level courses. I have been researching the effectiveness and efficiency of solving Parsons problems versus writing and fixing code. I also created two types of adaptation for Parsons problems: intra-problem and inter-problem adaptation. In intra-problem adaptation if the learner is struggling to solve the current problem it can be made easier dynamically by disabling distractors, providing indentation, or combining blocks. In inter-problem adaptation the difficulty of the next problem is adjusted based on the learner’s performance on the previous problem. The goal is to keep the learner challenged but not frustrated.

3.6 Presenting Name/Value Mappings in Notional Machines

Kathi Fisler (Brown University – Providence, US)

License  Creative Commons BY 3.0 Unported license
© Kathi Fisler

For programs without assignment statements, there are several ways to capture the mapping from names to values. The tradeoffs are particularly interesting for programs that involve data with components or objects. We illustrate the tradeoffs of two models for program evaluation for programs with such data: one substitutes the value associated with each parameter name, while the other substitutes the heap address associated with each name. A preliminary user study shows that each has advantages in some contexts, suggesting that a combination of the models be used in program tracing tools.

3.7 Empirical Studies

Robert L. Goldstone (Indiana University – Bloomington, US)

License © Creative Commons BY 3.0 Unported license
© Robert L. Goldstone

Computer programming is one of the most cognitively demanding and complex tasks in which humans engage. It places challenging demands on working memory, abstraction, mental modeling and simulation, planning, problem solving, memory retrieval, and the creation of novel, robust and flexible structures and processes. There has been an extended literature on the psychology of computer programming. Some of this has focused on the syntactic, semantic, and strategic misconceptions that students and even experts possess. Other research has described performance factors related to fragile knowledge, cognitive load, natural language intrusions, limited working memory, schema-based misconstruals, perception, transfer of knowledge, and individual differences. A theoretically and important research agenda concerns how best to enable humans to produce sophisticated computer programs that push and even transcend human physical and mental limits. Pedagogical recommendations include: integrating role-based conceptions of variables, read-trace-explain-sketch curricula, incorporating concept inventories, combining worked-out examples and test items, optimal scheduling of worked-out examples, labeling, idealization, aligning natural and formal language, explicitly training for transfer, peer instruction, and game-based components. Transcending human limitations in programming will often involve the creation of human-machine distributed cognitive systems, featuring technological innovations such as: color coding/highlighting, visual editors, notional machines, algorithm visualizations, simplified languages/environments, human-consumable error and status messages, embedded assessments, learning analytics, and creating new programming languages explicitly design to fit and shape human mental models.

3.8 Making Programming Languages to Meet a Greater Need


Mark Guzdial (University of Michigan – Ann Arbor, US)

License © Creative Commons BY 3.0 Unported license
© Mark Guzdial

There is so little computer science in high schools today, in part because it's so hard to program. I suggest that we need to figure out what makes programming more accessible. New tools like Vega-Lite and Sarah Chasin's Helena suggest a different strategy – developing task-based programming languages that serve a specific purpose and can be used successfully within 10 minutes.

3.9 Conceptual Change in Learning to Program

Matthias Hauswirth (University of Lugano, CH)

License  Creative Commons BY 3.0 Unported license
© Matthias Hauswirth

Learning to program is hard. In this poster we show two approaches we used to investigate the conceptual change students undergo as novice programmers. We use the Informa Clicker tool where students construct responses, similar to visual program simulation, and we use the Informa Mastery Learning platform to support the detailed analysis of the development of a fine-grained set of specific skills. Based on these approaches we have identified a collection of 165 misconceptions about programming in Java. At USI we are now embarking on a project to investigate trajectories through that space of conceptual understanding and to connect learning between different programming languages.

3.10 What Do Students “See” in Computing Contexts?


Geoffrey L. Herman (University of Illinois – Urbana Champaign, US)

License  Creative Commons BY 3.0 Unported license
© Geoffrey L. Herman

Cognitive research has shown that gaining expertise in a subject area both changes what an individual sees when shown a visual representation and it also changes how that individual searches for information in that visual representation. The goal of my research is to explore the connection between perception and students’ knowledge of computational notional machines. Using a mixture of eye-tracking methods and qualitative interviews, we are seeking to describe how students learn to read and trace code.

3.11 Reading Code Aloud

Felienne Hermans (Leiden University, NL)

License  Creative Commons BY 3.0 Unported license
© Felienne Hermans

When children learn to read, they almost invariably start with oral reading: reading the words and sentences out loud, not just to demonstrate their newly acquired skill, but also because they simply cannot do it in a different fashion yet. Most children take years to learn to read silently, during which they go through a number of phases including whispering and lip movement. Several studies have shown that, for novice readers, reading aloud supports comprehension. This should not come as a surprise, sometimes when reading difficult English words, I still read aloud! While we do not know exactly how reading aloud helps, the fact that it does is often attributed to the fact that reading aloud focuses your attention to the text, and thus makes it less likely that you will skip letters or words.

This made us wonder, why do we not practice to read code aloud? In the same way that reading text aloud helps to understand meaning, so could reading source code! We call this idea code phonology. Settling on a phonology could be challenging than you think, even for simple statements. For example, how should we pronounce an assignment statement like

$x = 5$? Is it “ x is 5”? Or “set x to 5”? Or “ x gets 5”? And what about an equality check? Is it “if x is is 5”? Or “if x is 5”? Or “is x is equal to 5”? As you can see, this could lead to tantalizing discussions.

3.12 Sensing and First Data

Matthew C. Jadud (Bates College – Lewiston, US)

License © Creative Commons BY 3.0 Unported license
© Matthew C. Jadud

My work broadly explores the behavior of novice programmers and tools to support them in their learning. Recently, my students and I have been developing custom hardware for environmental sensing and extensions to Microsoft’s MakeCode online programming environment to support novices in quickly logging data captured from the world around them. Our goal is to enable the capture of small, personally relevant data sets that beginners can use when first learning to use code to work with data.

3.13 Giving Feedback and Hints in (Haskell/Java/...) Programming Tutors Based on Comparing Model Solutions to Student Solutions

Johan Jeuring (Utrecht University, NL)

License © Creative Commons BY 3.0 Unported license
© Johan Jeuring

Ask-Elle is a tutor for learning the higher-order, strongly-typed functional programming language Haskell. It supports the stepwise development of Haskell programs by verifying the correctness of incomplete programs, and by providing hints. Teachers can add programming exercises to Ask-Elle by providing a task description for the exercise, one or more model solutions, and properties that a solution should satisfy. A teacher can annotate properties and model solutions with feedback messages, and can specify the amount of flexibility allowed in student solutions. We calculate feedback using a variant of higher-order unification, extended such that it can deal with several more pragmatic aspects, such as the order of arguments of a function, or the order of declarations in a let expression.

3.14 Philosophical Concept Analysis in PL or SE or CSE or ...


Antti-Juhani Kaijanaho (University of Jyväskylä, FI)

License © Creative Commons BY 3.0 Unported license
© Antti-Juhani Kaijanaho

Concepts such as notional machines create a lot of confusion, because people use the same term in multiple different ways, and sometimes this difference in definitions is hard to spot. I argue that it is necessary to foreground the debate on concepts by having people explicitly state and defend their analyses (definitions), and for others who disagree to provide thoughtful counter-arguments. The goal might be a precise definition of the concept (in the classical style), or the replacement of a concept with a better one (in a Carnapian style). The result might be an agreed definition, but it also could be an agreement that the concept is incoherent, or actually is multiple concepts that need to be differentiated from each other.

3.15 Towards Algorithm Comprehension


Philipp Kather (Universität Münster, DE) and Jan Vahrenhold (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license
© Philipp Kather and Jan Vahrenhold

Comprehending and developing algorithms are very common activities in computer science and other studies. But what does it mean to comprehend an algorithm? Why are students creating flawed algorithms with correct proofs? We presented the current progress of a grounded theory study concerning algorithm comprehension to discuss this topic from a notional machines perspective.

3.16 Code and Cognition Lab


A. J. Ko (University of Washington – Seattle, US)

License  Creative Commons BY 3.0 Unported license
© A. J. Ko

My work contributes to the fields of computing education, human-computer interaction, and software engineering. My lab has recently focused on programming language learning, API learning, programming problem solving, machine learning literacy, and design literacy, as well as issues of diversity, equity, and inclusion in all of these topics. These discoveries building an evidence base for how to effectively and inclusively educate outstanding design- and data- literate programmers.

3.17 Language Levels

Shriram Krishnamurthi (Brown University – Providence, US)

License  Creative Commons BY 3.0 Unported license
© Shriram Krishnamurthi

Students don't program in one language; they program in several. Even through the course of a single book, ostensibly in a single language, the amount of the language they are exposed to keeps growing. This growth usually corresponds to an increase in complexity of the language's semantics. However, our IDEs rarely reflect this growth, presenting a monolithic language interface and leaving it to students to ensure they stay in the expected sublanguage. The DrRacket programming environment represents a rare exception, presenting a series of pedagogic languages, and including tools for building many more. The tools also vary with the language level – especially the Algebraic Stepper, which is a visualization of the notional machine. The corresponding book, *How to Design Programs*, also presents the notional machine also as a series of increasingly complex rules as the language grows; these rules are then manifest in the Stepper. In fact, different books choose to decompose the full language in different ways, and each can get the environment to reflect its chosen decomposition.

3.18 Explicit Programming Strategies

Thomas D. LaToza (George Mason University – Fairfax, US)

License © Creative Commons BY 3.0 Unported license
© Thomas D. LaToza

Software developers solve a diverse and wide range of problems, relying on programming strategies that they have learned. A programming strategy is a human-executable procedure for solving a programming task. We have developed a notation for writing strategies down explicitly in a program-like notation called Roboto. Using a strategy tracker tool, developers can follow a strategy step by step, as the computer keeps track of the next step and information they have collected executing the strategy. We've given explicit programming strategies to software developers and demonstrated that this representation enables developers to break their existing habits and work in new and more effective ways.

3.19 Conceptual Change & Knowledge in Pieces (KiP)

Colleen Lewis (Harvey Mudd College – Claremont, US) and Matthias Hauswirth (University of Lugano, CH)

License © Creative Commons BY 3.0 Unported license
© Colleen Lewis and Matthias Hauswirth

This workshop presented background information about theories within conceptual change, and particular details regarding the Knowledge in Pieces (KiP) perspective, created by Dr. Andrea A. diSessa (the PhD advisor of Dr. Lewis). The workshop presented three important aspects of KiP:

- (1) Typical uses and definitions of “mental model” ignore variations within a single student.
- (2) Learning involves learning to consistently use the “right” knowledge in different contexts.
- (3) Various knowledge fragments exist because they have been productive in some context.

For researchers new to the area, we also defined some frequently misunderstood terms: phenomenological-primitive (p-prim) and coordination class. The workshop concluded with a call to iteratively refine our understanding of CS learning by conducting research that takes into account students' moment-by-moment reasoning and is accountable to patterns of long-term conceptual change.

3.20 Concrete Notional Machines

Colleen Lewis (Harvey Mudd College – Claremont, US)

License © Creative Commons BY 3.0 Unported license
© Colleen Lewis

Abstraction is frequently mentioned as a core skill developed when learning programming, but computer science (CS) education rarely draws on education research focused on helping students build their understanding of abstraction. A particularly promising practice is known as concrete-to-representational-to-abstract, or CRA. Common practices for teaching addition are an example of CRA. CRA begins by introducing a physical (i.e., concrete) object. For example, this could be physical blocks that could be counted to add them together. Once

students are comfortable adding together sets of physical blocks, students could advance to solving the same problems given only a picture (i.e., representation) of the blocks. Once students are comfortable using only the pictures, students could advance to solving the same problems using only numbers (i.e., abstraction). If a student has trouble, they can always return to a previous representation. I have applied CRA to develop concrete memory models (i.e., the concrete), which then transition to drawn memory models (i.e., the representational), and ultimately Java code (i.e., the abstract).

3.21 Reference-point Errors: Slips? or Misconceptions of the Notional Machine?


Craig Miller (DePaul University – Chicago, US)

License  Creative Commons BY 3.0 Unported license
© Craig Miller

Novice programmers may mistakenly write code that references an object when the attribute of the object is intended, or vice versa. These errors are consistent with the use of metonymy, a type of figurative expression in human-to-human communication. Instead of misconceptions, the errors may be slips based on well-practiced habits of figurative communication

3.22 Notional Machines for Everyday Life

Andreas Mühling (Universität Kiel, DE)

License  Creative Commons BY 3.0 Unported license
© Andreas Mühling

Notional machines are typically seen as a way to allow learners to predict how a given program will execute. In the current discussion about “digital literacy” as a necessary qualification for all current and future citizens, the question arises how to explain how digital artefacts work without necessarily delving into programming. To this end, the idea of broadening the concept of notional machines to explain everyday phenomena (in particular also when considering communicating digitale devices) has been developed. The current state of the project was presented as a series of increasingly detailed abstractions of how digital devices work starting from a complete black box and ending somewhere above the von Neumann system of a machine. Each new level is introduced by a phenomenon that cannot be explained with the current level of detail. Future research will help in identifying the educational value of this concept and the optimal progression and level of granularity.

3.23 Making a Causal Diagram for Learning Programming

Greg Nelson (University of Washington – Seattle, US)

License © Creative Commons BY 3.0 Unported license
© Greg Nelson

Greg invited everyone to expand a simple causal diagram for learning programming, using sticky notes. The big idea behind making a causal diagram is that the feedback loops are the main determinants of system behavior (i.e. learning outcomes). Thanks to Ben du Boulay, Titus Barik, Rodrigo Duran, Kathryn Cunningham, Thomas LaToza, and Tom Ball for participating.

3.24 Pointer Concepts in C

Andrew Petersen (University of Toronto, CA)

License © Creative Commons BY 3.0 Unported license
© Andrew Petersen

Many introductory computing courses at the University of Toronto are built around frequent practice, supported by an online system that delivers online exercises and provides feedback on student submissions. In an early example of the type of analysis that can be performed from this data, we investigated student use (and mis-use) of pointers in their first week of exposure to pointer types in C. We defined a set of core pointer concepts to be covered in that week and then developed a pre- and post- assessment to identify which topics were most frequently mis-applied by students. We use the results of these questions to roughly order the concepts by difficulty. Additionally, we analyze student submissions to coding exercises, revealing inefficient behaviours students use to solve pointer problems and identifying the most common errors committed.

3.25 Semantics Tutorial

Joseph Gibbs Politz (UC – San Diego, US)

License © Creative Commons BY 3.0 Unported license
© Joseph Gibbs Politz

Semanticists use formalisms like grammars, relations, and trees to model the behavior of programs and programming languages. As they are a description of how programs evaluate, semantics are closely related to notional machines. This tutorial motivates semantics for programming languages with an example of syntactic scope in Python and substitution in Racket. It then goes on to show a worked example of a small-step, substitution-based semantics with evaluation contexts (in the style of Felleisen and Hieb) for a subset of Python.

3.26 Activity Theory

R. Benjamin Shapiro (University of Colorado Boulder, US)

License  Creative Commons BY 3.0 Unported license
© R. Benjamin Shapiro

Computing education research typically draws on theories from educational psychology and cognitive psychology. While useful, these theories, and their applications in computing education research, often fail to account for the situated, embodied, culturally-constructed, historically anchored, social, and materially-mediated nature of learning. I describe how activity theory can help us to attend to the practice of computing education in more nuanced and expansive ways. Here, practice refers to systems of teaching and learning, including how tools (like programming languages) are used within that practice, are designed with particular sets of values and practices in mind, and are also adopted based on sets of practices that may or may not be shared by researchers and designers operating in this area. I then draw on questions posed by Engestrom to challenge the audience to consider

- (a) how the history of computing, computing education, and educational institutions shapes our present practice,
- (b) what tools and signs (e.g. programming languages or assessments) are available to different participants in the networks of computing education practice, and how they are used to construct the objects of our activity,
- (c) what contradictions exist within our activity systems, and
- (d) to thoughtfully consider what can and should be done to construct better systems of practice.

3.27 Stuff We Wish We Knew (About Notional Machines)

Juha Sorva (Aalto University, FI) and Otto Seppälä (Aalto University, FI)

License  Creative Commons BY 3.0 Unported license
© Juha Sorva and Otto Seppälä

We identified four areas with open questions related to notional machines. Dynamic visualizations of science concepts work best when designed for specific roles within a pedagogical approach and when students are taught representational competencies for reading the visualization, but there is insufficient research on these topics in notional-machine visualization. Little is known about how various increasingly common programming-language features – such as higher-order functions, type systems, type inference, and anonymous functions – should be attended to in notional machines. Studies of programming knowledge could be better informed by knowledge-in-pieces theories of conceptual change, which suggest a role for notional machines in integrating fragmented knowledge. The concept of notional machine requires further analysis and clarification, and it is unclear whether the term is helpful when disseminating research-based practices to teachers.

3.28 Revisiting Two Past Publications through the Lens of Notional Machines

J. Ángel Velázquez Iturbide (Universidad Rey Juan Carlos – Madrid, ES)

License © Creative Commons BY 3.0 Unported license
© J. Ángel Velázquez Iturbide

In this poster, I presented two past publications [1,2] in a slightly different way, through the lens of notional machines. In one work [1], three instantiations of recursion (namely recursion in grammars, in functional programs, and in procedural programs) were analyzed to understand varying difficulties of students in understanding them. The analysis identified their different representations of information and operational models, hypothesizing increasing complexity of their, let us say, notional machines. In a second work [2], we presented and evaluated a novel approach to enhancing students' understanding of recursion. We presented removal of linear recursion into equivalent, iterative code by means of a transformation scheme. In retrospect, we were explaining a part of a procedural notional machine (namely, recursion) in terms of another part of the same notional machine (iteration).

References

- 1 J.Á. Velázquez-Iturbide. "Recursion in gradual steps (is recursion really that difficult?)". Proc. SIGCSE 2000, 310-314, DOI 10.1145/330908.331876
- 2 J.Á. Velázquez-Iturbide, M.E. Castellanos & R. Hijón-Neira. "Recursion removal as an instructional method to enhance the understanding of recursion tracing". IEEE Trans. Education, 59(3):161-168, August 2016, DOI 10.1109/TE.2015.2468682

4 Working groups

4.1 Breakout Group 1

Geoffrey L. Herman (University of Illinois – Urbana Champaign, US) and Philipp Kather (Universität Münster, DE)

License © Creative Commons BY 3.0 Unported license
© Geoffrey L. Herman and Philipp Kather

This breakout group suggested a perspective on notional machines as a model between the source code and the actual machine. Models such as those in physics are only useful within certain bounds. They are half truths, excluding some aspects to be more useful in some contexts. Research questions related to the bounds of notional machines in various learning contexts, such as teaching multiple languages interleaved were developed. The need for instruments measuring the quality of students understanding of notional machines and the relevance of considering non-cognitive factors was also highlighted.

4.2 Breakout Group 2


Craig Miller (DePaul University – Chicago, US) and Franziska Carstens (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license
© Craig Miller and Franziska Carstens

This breakout session was shaped by the question “What do we need to know about notional machines that we don’t know yet?” The objective was to brainstorm on possible research questions, record them and discuss about appropriate study designs. During the discussion, the group pointed on the questions: what is the relation between notional machines and formal semantics, what is the language we use to describe a notional machine, and what is instructors practice on notional machines? To answer these questions, four different approaches were discussed. The first suggestion was to look at experts and collect data on explanations given by instructors. A second idea was to focus on the students execution of different possible notional machines and come up with an experimental control group design. For a third approach the group discussed about examining textbooks to identify presented notional machines and at least, they thought about taking a look at other areas (e. g. electrical engineering) and get an inside if and how notinal machines are used and presented there. During the whole session, the participants reflected on a domain sensitivity and asked if we would need different notional machines for different domains or if it is more a question of highlighting parts of one notional machine. In the course of the session, the participants increasingly used the term notional machine synonymous with lie. This perspective led to further considerations such as the impact of lies in a notional machine on students learning and the question of how much practice is required so that a simplified notional machine or lie becomes obsolete. In the end, the group agreed on taking a further look on instructors’ perspectives and formulated the research question “How much are instructors willing to lie to their students?” Under the assumption that every instructor has preferences and beliefs, the group thought about a survey-study to identify preferred ‘lies’ of instructors to their students and collect information on influences that may lead to possible preferences.

4.3 Breakout Group 3

Brett A. Becker (University College Dublin, IE), Neil C. C. Brown (King’s College London, GB), Paul Denny (University of Auckland, NZ), Rodrigo Duran (Aalto University, FI), Robert L. Goldstone (Indiana University – Bloomington, US), Antti-Juhani Kaijanaho (University of Jyväskylä, FI), Greg Nelson (University of Washington – Seattle, US), Carsten Schulte (Universität Paderborn, DE), Otto Seppälä (Aalto University, FI), and Steven A. Wolfman (University of British Columbia – Vancouver, CA)

License  Creative Commons BY 3.0 Unported license
© Brett A. Becker, Neil C. C. Brown, Paul Denny, Rodrigo Duran, Robert L. Goldstone, Antti-Juhani Kaijanaho, Greg Nelson, Carsten Schulte, Otto Seppälä, and Steven A. Wolfman

Breakout group number three reflected and generated research questions about previous approaches that in some manner used or were built on the concept of a notional machine and what instructional design was used to achieve the goal of such approach. The discussion topics generally fell into 3 categories:

- (a) How do we elicit learner mental models?

- (b) How do we change learner mental models?
- (c) How do we achieve near-term and long-term pedagogical goals?

These topics included: how mental models are used in practice; how to elicit students mental models and how they are connected to a notional machine presented by the instructor; how notional machines along with a concrete-to-formal continuum impact learning outcomes for different audiences; what are the perspectives of the CSEd community regarding notional machines; How to sequence notional machines and how to achieve transfer between programming languages, and which presentation form best suits novices. Greg's report back slides are here and are a quick synthesis of the clusters of RQs our group generated, and also include pictures of original ideation materials.

4.4 Breakout Group 4

Kathi Fisler (Brown University – Providence, US) and Kathryn Cunningham (University of Michigan – Ann Arbor, US)

License © Creative Commons BY 3.0 Unported license
© Kathi Fisler and Kathryn Cunningham

We considered the scope of notional machines. We agreed that a notional machine is a pedagogical tool, and it must explain the execution of programs. For systems that don't have programs (e.g. a cell phone in standard use), notional machines don't apply, and findings from HCI are likely more applicable. What is a minimal notional machine? IFTTT (If This Then That) is a platform where people can program different systems to interact with each other using very simple rules in "if-then" format. The notional machine to understand IFTTT seems quite limited, although the applications are powerful. This balance is possible since so much of the functionality is black-boxed. We believe a notional machine is a mediating artifact that attempts to reconcile the mental model of a student and a teacher. From this perspective, there are research questions about the way teachers interact with notional machines, the way students interact with notional machines, and the way the context of a topic or learning environment interacts with notional machines. We decided that one of the foundational research questions is how instructors use notional machines in practice. We proposed a collection of notional machine examples from a variety of instructors, to examine the different ways that instructors describe program execution to students.

4.5 Categorizing Notional Machines and their Representation or Visualization

Franziska Carstens (Universität Münster, DE)


License © Creative Commons BY 3.0 Unported license
© Franziska Carstens

The aim of this group was to develop a better understanding of the characteristics of notional machines. At the beginning of the session, the group had a short introduction about pattern language, given by Sally Fincher. During this talk, the participants got a small inside into The Engineer's Sketch-Book (<https://archive.org/details/engineerssketchb00barb/page/n17>) and discussed structuring principles and their importance. Afterwards, the group brainstormed

characteristics of notional machines and refined the result with concrete examples that originate directly from teaching practice. The group stayed with the plan to conduct an interview study after the seminar to collect further simplifying examples from teaching practice that are used to help students understand program execution or program state.

4.6 Instructional Design for Notional Machines

Barbara Ericson (University of Michigan – Ann Arbor, US), Robert L. Goldstone (Indiana University – Bloomington, US), Matthias Hauswirth (University of Lugano, CH), Antti-Juhani Kaijanaho (University of Jyväskylä, FI), Greg Nelson (University of Washington – Seattle, US), André L. Santos (University Institute of Lisbon, PT), and Anya Tafliovich (University of Toronto, CA)

License  Creative Commons BY 3.0 Unported license
© Barbara Ericson, Robert L. Goldstone, Matthias Hauswirth, Antti-Juhani Kaijanaho, Greg Nelson, André L. Santos, and Anya Tafliovich

This breakout group generated a set of instructional design guidelines, as well as a library of examples of instructional designs.

4.7 Concept Analysis for Notional Machines

Antti-Juhani Kaijanaho (University of Jyväskylä, FI), Thomas Ball (Microsoft Research – Redmond, US), Markus Müller-Olm (Universität Münster, DE), and Juha Sorva (Aalto University, FI)

License  Creative Commons BY 3.0 Unported license
© Antti-Juhani Kaijanaho, Thomas Ball, Markus Müller-Olm, and Juha Sorva

This ad-hoc breakout group discussed the need for a tradition of deliberate argumentation in support or against particular concept analyses, taking as a starting point Kaijanaho's Onward 2017 essay "Concept analysis in programming language research: Done well it is all right". We concluded that if someone believes another to be wrong, it is their obligation to respond with a counter-argument. It is, however, difficult to publish such arguments in academic forums, and we believe this needs to change. Right now, we can start by collecting current understanding of notional machines and by encouraging people to write position papers with deliberate argumentation.

4.8 Notional Machines for Everything

Joseph Gibbs Politz (UC – San Diego, US), Mark Guzdial (University of Michigan – Ann Arbor, US), and Philipp Kather (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license
© Joseph Gibbs Politz, Mark Guzdial, and Philipp Kather

This breakout group discussed aspects of notional machines for parallel computing, reactive programming, reading mathematical proofs and javascript. Considering the audience the notional machine is taught to was the most important aspect when one develops a notional machine. A mental model of a notional machine, e.g. for serial computing, might be already developed, when a student engages in parallel computing. Events triggering the need for a

new notional machine were discussed as opportunities for introducing a new notional machine or extending the existing one. However, many languages used in industry allow behavior that is convenient for experienced programmers but too complex for a student to grasp at once. This is why the development of sub-languages and defining their notional machines in a way, that a student would be able to explain all behavior, would be an appropriate approach to teach those languages.

4.9 Notional Machines for Scratch and Python

Otto Seppälä (Aalto University, FI), Thomas Ball (Microsoft Research – Redmond, US), Titus Barik (Microsoft Research – Redmond, US), Brett A. Becker (University College Dublin, IE), Paul Denny (University of Auckland, NZ), Rodrigo Duran (Aalto University, FI), Juha Sorva (Aalto University, FI), and J. Ángel Velázquez Iturbide (Universidad Rey Juan Carlos – Madrid, ES)


License © Creative Commons BY 3.0 Unported license

© Otto Seppälä, Thomas Ball, Titus Barik, Brett A. Becker, Paul Denny, Rodrigo Duran, Juha Sorva, and J. Ángel Velázquez Iturbide

We originally set ourselves a goal to study and contrast the notional machines for Python and Scratch. A blog post by Greg Wilson “Is this a notional machine for Python?” (<http://third-bit.com/2018/04/12/notional-machine-for-python.html>) was used as a starting point for the discussion and a possible reference to create and contrast a Scratch version with. A notional machine can be suited for a specific audience and written to target a specific part of the programming language and the execution environment. Studying Wilson’s suggested notional machine for Python, we found it in many cases to be more generic in nature and describe features common to imperative languages in general – characteristics such as memory management and call stack behavior. This led to discussions about notional machines that could be used for a family of languages. Scratch, being a language that has been intentionally simplified for a younger audience, however was not found to have all the traits described in Wilson’s notional machine. Recreating, for example, a recursive version of a function calculating a factorial using Scratch was not possible as Scratch only allows user-made procedures without return values. (The function studied in the first meeting can be seen here: <https://tinyurl.com/dags-nm>) In our second meeting our goal was to start from the most minimalistic Python(-compatible) program imaginable, to consider the minimal language required to reason about the program and to come up with a sound and complete notional machine for this specific language. Our first program consisted only of a single assignment. The notional machine for this language had three rules explaining variables, values and assignment. We then augmented our program three times, each time using a new language feature and tried to find the minimal addition to the existing rules. This eventually led to a progression of notional machines each building on the previous iteration. One key outcome from the exercise was the additive design process itself.

4.10 Notional Machines and Simulation

Steven A. Wolfman (University of British Columbia – Vancouver, CA), Kathryn Cunningham (University of Michigan – Ann Arbor, US), Kathi Fisler (Brown University – Providence, US), Greg Nelson (University of Washington – Seattle, US), and Jan Vahrenhold (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license

© Steven A. Wolfman, Kathryn Cunningham, Kathi Fisler, Greg Nelson, and Jan Vahrenhold

This breakout group elaborated the use of notional machines as pedagogical tools during code writing, tracing and debugging and discussed how students develop a coherent mental model of those. There are clues that notional machine knowledge is not immediately helpful for code writing. If notional machine knowledge could somehow be integrated with higher-level, more abstract knowledge—such as programming plans and goals—notional machine knowledge and code writing knowledge may be brought together.

The breakout group suspects that the most fruitful goal for notional machines in students' code creation process is to foster a habit-of-mind of meta-cognition about the behaviour of the code they generate. Research questions going forward consider goals of instructors teaching notional machines, students' and experts' application of notional machines, pedagogical practices encouraging a productive habit-of-mind of simulating/verifying code and the use of documentation of abstraction created by students or instructors to elaborate their mental models of notional machines.

Participants

- Thomas Ball
Microsoft Research –
Redmond, US
- Titus Barik
Microsoft Research –
Redmond, US
- Brett A. Becker
University College Dublin, IE
- Neil C. C. Brown
King's College London, GB
- Franziska Carstens
Universität Münster, DE
- Luke Church
University of Cambridge, GB
- Kathryn Cunningham
University of Michigan –
Ann Arbor, US
- Paul Denny
University of Auckland, NZ
- Brian Dorn
University of Nebraska –
Omaha, US
- Benedict du Boulay
University of Sussex –
Brighton, GB
- Rodrigo Duran
Aalto University, FI
- Barbara Ericson
University of Michigan –
Ann Arbor, US
- Sally Fincher
University of Kent –
Canterbury, GB
- Kathi Fisler
Brown University –
Providence, US
- Robert L. Goldstone
Indiana University –
Bloomington, US
- Mark Guzdial
University of Michigan –
Ann Arbor, US
- Reiner Hähnle
TU Darmstadt, DE
- Matthias Hauswirth
University of Lugano, CH
- Arto Hellas
University of Helsinki, FI
- Geoffrey L. Herman
University of Illinois –
Urbana Champaign, US
- Felienne Hermans
Leiden University, NL
- Matthew C. Jadud
Bates College – Lewiston, US
- Johan Jeuring
Utrecht University, NL
- Antti-Juhani Kaijanaho
University of Jyväskylä, FI
- Philipp Kather
Universität Münster, DE
- A. J. Ko
University of Washington –
Seattle, US
- Shriram Krishnamurthi
Brown University –
Providence, US
- Thomas D. LaToza
George Mason University –
Fairfax, US
- Colleen Lewis
Harvey Mudd College –
Claremont, US
- Elena Machkasova
University of Minnesota –
Morris, US
- Craig Miller
DePaul University – Chicago, US
- Andreas Mühling
Universität Kiel, DE
- Markus Müller-Olm
Universität Münster, DE
- Greg Nelson
University of Washington –
Seattle, US
- Andrew Petersen
University of Toronto, CA
- Joseph Gibbs Politz
UC – San Diego, US
- André L. Santos
University Institute of
Lisbon, PT
- Carsten Schulte
Universität Paderborn, DE
- Otto Seppälä
Aalto University, FI
- R. Benjamin Shapiro
University of Colorado –
Boulder, US
- Juha Sorva
Aalto University, FI
- Anya Tafilovich
University of Toronto, CA
- Jan Vahrenhold
Universität Münster, DE
- J. Ángel Velázquez Iturbide
Universidad Rey Juan Carlos –
Madrid, ES
- Eugene Wallingford
University of Northern Iowa –
Cedar Falls, US
- David Weintrop
University of Maryland –
College Park, US
- Steven A. Wolfman
University of British Columbia –
Vancouver, CA



Data Series Management

Edited by

Anthony Bagnall¹, Richard L. Cole², Themis Palpanas³, and
Kostas Zoumpatianos⁴

1 University of East Anglia – Norwich, GB, anthony.bagnall@uea.ac.uk

2 Tableau Software – Palo Alto, US, ricole@tableau.com

3 University of Paris, FR, themis@mi.parisdescartes.fr

4 Harvard University – Cambridge, US, kostas@seas.harvard.edu

Abstract

We now witness a very strong interest by users across different domains on data series (a.k.a. time series) management. It is not unusual for industrial applications that produce data series to involve numbers of sequences (or subsequences) in the order of billions (i.e., multiple TBs). As a result, analysts are unable to handle the vast amounts of data series that they have to manage and process. The goal of this seminar is to enable researchers and practitioners to exchange ideas and foster collaborations in the topic of data series management and identify the corresponding open research directions. The main questions answered are the following: i) What are the data series management needs across various domains and what are the shortcomings of current systems, ii) How can we use machine learning to optimize our current data systems, and how can these systems help in machine learning pipelines? iii) How can visual analytics assist the process of analyzing big data series collections? The seminar focuses on the following key topics related to data series management: 1) Data series storage and access patterns, 2) Query optimization, 3) Machine learning and data mining for data series, 4) Visualization for data series exploration, 5) Applications in multiple domains.

Seminar July 7–12, 2019 – <http://www.dagstuhl.de/19282>

2012 ACM Subject Classification Information systems → Data management systems

Keywords and phrases data series; time series; sequences; management; indexing; analytics; machine learning; mining; visualization

Digital Object Identifier 10.4230/DagRep.9.7.24


1 Executive Summary

Anthony Bagnall (University of East Anglia, GB)

Richard L. Cole (Tableau Software, US)

Themis Palpanas (Paris Descartes University, FR)

Kostas Zoumpatianos (Harvard University, US)

License  Creative Commons BY 3.0 Unported license

© Anthony Bagnall, Richard L. Cole, Themis Palpanas, Kostas Zoumpatianos

We now witness a very strong interest by users across different domains on data series¹ (a.k.a. time series) management systems. It is not unusual for industrial applications that produce data series to involve numbers of sequences (or subsequences) in the order of billions. As

¹ A data series, or data sequence, is an ordered set of data points.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Data Series Management, *Dagstuhl Reports*, Vol. 9, Issue 7, pp. 24–39

Editors: Anthony Bagnall, Richard L. Cole, Themis Palpanas, and Konstantinos Zoumpatianos



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

a result, analysts are unable to handle the vast amounts of data series that they have to filter and process. Consider for instance that in the health industry, for several of their analysis tasks, neuroscientists are reducing each of their 3,000 point long sequences to just the global average, because they cannot handle the size of the full sequences. Moreover, in the quest towards personalized medicine, scientists are expected to collect around 2-40 ExaBytes of DNA sequence data by 2025. In engineering, there is an abundance of sequential data. Consider for example that each engine of a Boeing Jet generates 10 TeraBytes of data every 30 minutes, while domains such as energy (i.e., wind turbine monitoring, etc.), data center, and network monitoring continuously produce measurements, forcing organizations to develop their custom solutions (i.e., Facebook Gorilla).

The goal of this seminar was to enable researchers and practitioners to exchange ideas in the topic of data series management, towards the definition of the principles necessary for the design of a big sequence management system, and the corresponding open research directions.

The seminar focused on the following key topics related to data series management:

Applications in multiple domains: We examined applications and requirements originating from various fields, including astrophysics, neuroscience, engineering, and operations management. The goal was to allow scientists and practitioners to exchange ideas, foster collaborations, and develop a common terminology.

Data series storage and access patterns: We described some of the existing (academic and commercial) systems for managing data series, examined their differences, and commented on their evolution over time. We identified their shortcomings, debated on the best ways to lay out data series on disk and in memory in order to optimize data series queries, and examined how to integrate domain specific summarizations/indexes and compression schemes in existing systems.

Query optimization: One of the most important open problems in data series management is that of query optimization. However, there has been no work on estimating the hardness/selectivity of data series similarity search queries. This is of paramount importance for effective access path selection. During the seminar we discussed the current work in the topic, and identified promising future research directions.

Machine learning and data mining for data series: Recent developments in deep neural network architectures have also caused an intense interest in examining the interactions between machine learning algorithms and data series management. We discussed machine learning from two perspectives. First, how machine learning techniques can be applied for data series analysis tasks, as well as for tuning data series management systems. Second, we how data series management systems can contribute towards the scalability of machine learning pipelines.

Visualization for data series exploration: There are several research problems in the intersection of visualization and data series management. Existing data series visualization and human interaction techniques only consider very small datasets, yet, they can play a significant role in the tasks of similarity search, analysis, and exploration of very large data series collections. We discussed open research problems along these directions, related to both the frontend and the backend.

2 Table of Contents

Executive Summary

Anthony Bagnall, Richard L. Cole, Themis Palpanas Kostas Zoumpatianos 24

Overview of Talks


Interaction Metaphors for Time Series Analysis <i>Azza Abouzied</i>	28
Mini Tutorial on Time Series Data Mining Top of Form <i>Anthony Bagnall</i>	28
Visualizing Large Time Series (a brief overview) <i>Anastasia Bezerianos</i>	29
Anomaly Detection in Large Data Series <i>Paul Boniol</i>	29
Data Series Management and Query Processing in Tableau <i>Richard L. Cole</i>	29
Location Intelligence <i>Michele Dallachiesa</i>	30
Data Series Similarity Search: Where Do We Stand Today? And Where Are We Headed? <i>Karima Echihabi</i>	30
Progressive PCA for Time-Series Visualization <i>Jean-Daniel Fekete</i>	30
Deep Learning for Time Series Classification, and Applications in Surgical Data Science <i>Germain Forestier</i>	31
Seismic Time Series: Introduction and Applications <i>Pierre Gaillard</i>	32
Progressive Similarity Search in Large Data Series Collections <i>Anna Gogolou</i>	32
Model-Based Management of Correlated Dimensional Time Series <i>Søren Kejser Jensen</i>	33
Time Series Recovery <i>Mourad Khayati</i>	33
Adaptive and fractal time series analysis: methodology and applications <i>Alessandro Longo</i>	33
Helicopters Time Series Management & Analysis <i>Ammar Mechouche</i>	34
Socio-temporal Data Mining <i>Abdullah Mueen</i>	34
Data Series Mining and Applications <i>Rodica Neamtu</i>	34

Fulfilling the Need for Big Sequence Analytics	
<i>Themis Palpanas</i>	35
Accelerating IoT Data Analytics through Time-Series Representation Learning	
<i>John Paparrizos</i>	36
Contradictory Goals of Classification, Accuracy, Scalability and Earliness	
<i>Patrick Schäfer</i>	36
More Reliable Machine Learning through Refusals	
<i>Dennis Shasha</i>	37
Systems and Tools for Time Series Analytics	
<i>Nesime Tatbul</i>	37
Data Series Similarity Search	
<i>Peng Wang</i>	37
Tableau for Data Series	
<i>Richard Wesley</i>	38
Managing and Mining Large Data Series Collections	
<i>Konstantinos Zoumpatianos</i>	38
Participants	39

3 Overview of Talks

3.1 Interaction Metaphors for Time Series Analysis

Azza Abouzied (New York University – Abu Dhabi, AE)

License  Creative Commons BY 3.0 Unported license
© Azza Abouzied

Through Qetch, I describe how a simple canvas metaphor can afford an intuitive and powerful querying language by allowing users to sketch patterns of interest, annotate them, as well as apply regular expression operations to search for repeated patterns or anomalies. The canvas metaphor also affords powerful multi-series querying functionality through the relative positioning of sketches. Through revisiting fundamental interaction metaphors, we can uncover elegant mechanisms for other complex time series analysis tasks.

3.2 Mini Tutorial on Time Series Data Mining Top of Form

Anthony Bagnall (University of East Anglia – Norwich, GB)

License  Creative Commons BY 3.0 Unported license
© Anthony Bagnall

TSDM is a research area that involves developing algorithms for tasks relating to time series. These can be grouped into two families of tasks:

1. Specializations of generic machine learning tasks: classification, regression, clustering, rule discovery and query problems, and all variants thereof, such as semi-supervised/active learning, attribute selection, reinforcement learning, etc.
2. Time series specific tasks:
 - a. Forecasting/panel forecasting;
 - b. Time to event modelling/survival analysis;
 - c. Annotation, such as segmentation, anomaly detection, motif discovery, discretization, imputation.

Problems can move from one task to another through a reduction strategy. For example, a regression task can be transformed to a classification task through discretizing the response variable, and forecasting can be reduced to regression through applying a sliding window. The challenges for TSDM include promoting reproducibility through open source code and improving evaluation strategies through better use of public data repositories and dealing with the challenges of large data so that algorithms can balance scalability vs accuracy. This becomes hugely important when dealing with streaming data, in particular with IoT applications involving widespread sensor nets where decisions need to be made about what data to store.

3.3 Visualizing Large Time Series (a brief overview)

Anastasia Bezerianos (*INRIA Saclay – Orsay, FR*)

License © Creative Commons BY 3.0 Unported license
© Anastasia Bezerianos

Visually representing in a meaningful way large timeseries remains a research challenge for the visualization community. We present examples of existing approaches that attack the problem using different solutions, such as representing visual aggregations, illustrating representative patterns in the data, or creating novel compact visual representations. One key aspect in deciding what to visualize and how, is to understand why the timeseries needs to be visualized – i.e., what tasks the viewer needs to perform. This influences both what type of visual representation is more appropriate to use, but also what interactions need to be supported to help visual analysis. We conclude with general challenges (and new directions) in visualizing and iteratively interacting with large amounts of data in real time.

3.4 Anomaly Detection in Large Data Series

Paul Boniol (*Paris Descartes University, FR*)

License © Creative Commons BY 3.0 Unported license
© Paul Boniol
Main reference Paul Boniol, Michele Linardi, Federico Roncallo, Themis Palpanas: “Automated Anomaly Detection in Large Sequences”, ICDE, 2020.

Subsequence anomaly (or outlier) detection in long sequences is an important problem with applications in a wide range of domains. However, the approaches that have been proposed so far in the literature have limitations: they either require prior domain knowledge, or become cumbersome and expensive to use in situations with recurrent anomalies of the same type. We briefly discuss these problems in this talk.

3.5 Data Series Management and Query Processing in Tableau

Richard L. Cole (*Tableau Software – Palo Alto, US*)

License © Creative Commons BY 3.0 Unported license
© Richard L. Cole

Tableau supports operations on time series, such as formatting, filters, calcs, date parts, date parse, and time zones. This talk is about Tableau’s aspirations to support query processing of exceptionally large data series, including complex data mining analytics, such as similarity search. Query processing may be divided into query compilation and query execution. New query compiler language elements and query execution operators will be needed. Additionally, support for data series data sources, i.e., time focused database systems, and federated query processing for data series in general will be desirable.

3.6 Location Intelligence

Michele Dallachiesa (Minodes GmbH – Berlin, DE)

License © Creative Commons BY 3.0 Unported license
© Michele Dallachiesa

For the first time in human’s history, the position of more than three-quarters of the world’s population is recorded at a fine-grained spatiotemporal resolution. This massive data source provides a unique view for infrastructure planning, retail development, and demographic research. In this talk, I overview two important localization strategies based on cellular and WiFi networks. In addition, the correct handling of missing or imprecise data points is presented as one of the major challenges in providing actionable insights with quality guarantees.

3.7 Data Series Similarity Search: Where Do We Stand Today? And Where Are We Headed?

Karima Echihabi (ENSIAS-Mohammed V University – Rabat, MA)

License © Creative Commons BY 3.0 Unported license
© Karima Echihabi

Main reference Karima Echihabi, Kostas Zoumpatianos, Themis Palpanas, Houda Benbrahim: “The Lernaean Hydra of Data Series Similarity Search: An Experimental Evaluation of the State of the Art”, Proc. VLDB Endow., Vol. 12(2), pp. 112–127, VLDB Endowment, 2018.

URL <https://doi.org/10.14778/3282495.3282498>

Main reference Karima Echihabi, Kostas Zoumpatianos, Themis Palpanas, Houda Benbrahim: “Return of the Lernaean Hydra: Experimental Evaluation of Data Series Approximate Similarity Search”, PVLDB, Vol. 13(3), pp. 403–420, 2019.

URL <http://dx.doi.org/10.14778/3368289.3368303>

Increasingly large data series collections are becoming commonplace across many different domains and applications. A key operation in the analysis of data series collections is similarity search, which has attracted lots of attention and effort over the past two decades. We presented the results of two extensive experimental evaluations. The three main lessons learned are as follows: 1) choosing the best approach is an optimization problem that depends on several factors (hardware, data characteristics, summarization quality and clustering efficacy); 2) exact search is slow; 3) approximate search can be fast and accurate (our extensions to exact techniques outperform the state-of-the-art on disk). We also outlined our future research directions: 1) building a new index that outperforms the state-of-the-art in-memory and on-disk; and 2) exploring query optimization for data series.

3.8 Progressive PCA for Time-Series Visualization

Jean-Daniel Fekete (INRIA Saclay – Orsay, FR)

License © Creative Commons BY 3.0 Unported license
© Jean-Daniel Fekete

With EDF, we are interested in the visual sensitivity analysis for ensemble simulation. EDF uses simulation software for forecasting the evolution of rivers and sea levels in the next 100 years. Their simulation system produces a large number of results, called “ensemble simulations”, that are plausible evolutions for a river, such as the level every month for the

next 100 years. These time series are then analyzed to find out if the results cluster around one value, or spread over multiple possible “regimes” or “modes”. This analysis is usually performed by clustering of the results but should be supervised and interpreted by analysts. Therefore, we use a dimensionality reduction algorithm to project the resulting time-series using Principal Component Analysis, explore the results, cluster it, and allow experts to write reports on the possible outcomes of the simulation. We are adapting PCA to cope with a large number of time series. First, PCA has seen recently a surge of new results related to its use online for out-of-core datasets. Second, iterative PCA computations have also been recently improved recently to boost its convergence using momentum. We are exploring these new algorithms as well as multi-resolution computation to reach interactive rates for computing PCA over a large number of time-series.

3.9 Deep Learning for Time Series Classification, and Applications in Surgical Data Science

Germain Forestier (University of Mulhouse, FR)

License  Creative Commons BY 3.0 Unported license
© Germain Forestier

In recent years, deep learning approaches have demonstrated a tremendous success in multiple domains like image processing, computer vision or speech recognition. In this talk, I reviewed recent advances in deep learning for univariate and multivariate time series classification. I presented experimental results obtained with the principal architectures proposed in the literature. I also discussed the main challenges linked with the use of deep learning like transfer learning, data augmentation, ensembling and adversarial attacks. Moreover, I presented some applications in the field of Surgical Data Science which is an emerging field with the objective of improving the quality of interventional healthcare through capturing, organizing, analyzing and modeling of data. Finally, I discussed an application of the above in Surgical Data Science. The need for automatic surgical skills assessment is increasing, especially because manual feedback from senior surgeons observing junior surgeons is prone to subjectivity and time consuming. Thus, automating surgical skills evaluation is a very important step towards improving surgical practice. I presented how we used a Convolutional Neural Network (CNN) to evaluate surgeon skills by extracting patterns in the surgeon motions performed in robotic surgery. The proposed method has been validated on the JIGSAWS dataset and achieved very competitive results with 100% accuracy on the suturing and needle passing tasks. While we leveraged from the CNNs efficiency, we also managed to mitigate its black-box effect using class activation map. This feature allows our method to automatically highlight which parts of the surgical task influenced the skill prediction and can be used to explain the classification and to provide personalized feedback to the trainee.

3.10 Seismic Time Series: Introduction and Applications

Pierre Gaillard (CEA de Saclay – Gif-sur-Yvette, FR)

License © Creative Commons BY 3.0 Unported license
© Pierre Gaillard

Seismometers, also called seismic stations, are sensitive instruments located all over the world, that allow to record continuously the smallest displacements of the ground. The given data take the form of discrete time series that are the basis of various studies: seismic risk analysis, seismic wave propagation, tomography of the Earth and seismic monitoring. This presentation is focused on seismic monitoring, and we present a standard pipeline dedicated to detect and characterize seismic event. To perform this task automatically, fast and reliable processing is required to extract as much information as possible from all the available time series. Such processing includes quality control, detection of event, measurement of features (amplitude, direction of arrival, polarity...), clustering or classification (e.g. anthropic versus natural events). All this information is then used by seismologists and are controlled, improved, shared or stored. This user intervention is usually performed through interactive software that need to manage and display large collection of time-series, as well as the associated data (detections, events, features...). Due to the increase of data available to perform seismic monitoring, we emphasize in the conclusion of this presentation, the need of new management system as well as new processing techniques based on machine learning in order to improve the analysis pipeline.

3.11 Progressive Similarity Search in Large Data Series Collections

Anna Gogolou (INRIA Saclay – Orsay, FR)

License © Creative Commons BY 3.0 Unported license
© Anna Gogolou


Main reference Anna Gogolou, Theophanis Tsandilas, Themis Palpanas, Anastasia Bezerianos: “Progressive Similarity Search on Time Series Data”, in Proc. of the Workshops of the EDBT/ICDT 2019 Joint Conference, EDBT/ICDT 2019, Lisbon, Portugal, March 26, 2019., CEUR Workshop Proceedings, Vol. 2322, CEUR-WS.org, 2019.

URL http://ceur-ws.org/Vol-2322/BigVis_5.pdf

Time series data are increasing at a dramatic rate, yet their analysis remains highly relevant in a wide range of human activities. Due to their volume, existing systems dealing with time series data cannot guarantee interactive response times, even for fundamental tasks such as similarity search. Therefore, in this talk, we present our vision to develop analytic approaches that support exploration and decision making by providing progressive results, before the final and exact ones have been computed. We demonstrate through experiments that providing first approximate and then progressive answers is useful (and necessary) for similarity search queries on very large time series data. Our findings indicate that there is a gap between the time the most similar answer is found and the time when the search algorithm terminates, resulting in inflated waiting times without any improvement. We present preliminary ideas on computing probabilistic estimates of the final results that could help users decide when to stop the search process, i.e., deciding when improvement in the final answer is unlikely, thus eliminating waiting time. Finally, we discuss two additional challenges: how to compute efficiently these probabilistic estimates, and how to communicate them to users.

3.12 Model-Based Management of Correlated Dimensional Time Series


Søren Kejser Jensen (Aalborg University, DK)

License  Creative Commons BY 3.0 Unported license
© Søren Kejser Jensen

Owners and manufacturers of wind turbines would like to collect and store large quantities of high-quality sensor data. However, the amount of storage required makes this infeasible and only simple aggregates are stored. This removes outliers and hides fluctuations that could indicate problems with the wind turbines. As a remedy, these high-quality regular time series can instead be stored as models which reduces the amount of storage required by approximating the time series within a user-defined error bound (possibly 0%). As time series change over time, each time series should be represented using multiple different model types, and as a data set often contains multiple similar time series, correlation should be exploited to further reduce the amount of storage required. ModelarDB is a time series management system that stores time series as models and takes all the above factors into account. There are still open-questions related to model-based storage of time series. How do we assist users with selecting a good set of model types to use for a particular data set? Can similarity search be performed directly on models instead of on data points reconstructed from the models? Can models be fitted at the turbines without significantly increasing the latency and/or the amount of data being transferred? And can the error bound be inferred from the user's query workload?

3.13 Time Series Recovery


Mourad Khayati (University of Fribourg, CH)

License  Creative Commons BY 3.0 Unported license
© Mourad Khayati

Recording sensor data is seldom a perfect process. Missing values often occur as blocks in time series data due to multiple reasons, e.g., sensor failure, server transmission, etc. In my talk, I introduced the problem of missing values in real-world time series data. Then, I introduced our solution to recover missing blocks in time series with mixed correlation. Finally, I summarized the main open research problems in the field.

3.14 Adaptive and fractal time series analysis: methodology and applications

Alessandro Longo (University of Rome III, IT)

License  Creative Commons BY 3.0 Unported license
© Alessandro Longo

A methodology for adaptive and fractal time series analysis, based on Empirical Mode Decomposition, time-varying filter EMD and Detrended Fluctuation Analysis has been applied to characterize time series data from different physical systems. It has been applied to seismometer data from sensors monitoring the Virgo interferometer and to data of activity concentration of cosmogenic beryllium-7, sampled worldwide by the International Monitoring

System of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). In the first case, seismometer was recording during an acoustic noise injection performed for detector characterization purposes. Using adaptive and fractal algorithms, the seismic perturbation due to acoustic noise performed in the room can be separated from the underlying nonlinear nonstationary noise affecting the seismometer. Furthermore, applying Hilbert Spectral analysis provided with a high-resolution time frequency representation, even though the data length is short due to the low sampling frequency. In the second case, extracting the yearly oscillatory mode of beryllium-7, sampled by a worldwide distributed network, allowed to characterize its shift in time of occurrence in term of patterns of large scale atmospheric dynamics, namely in term of the seasonal shift of the Hadley cell.

3.15 Helicopters Time Series Management & Analysis

Ammar Mechouche (Airbus Helicopters – Marignane, FR)

License  Creative Commons BY 3.0 Unported license
© Ammar Mechouche

Massive time series data are collected within the aerospace domain, making its management and analysis a challenging task. This presentation offers a return of experience regarding AIRBUS Helicopter flight data management and analysis. It illustrates first the big amount of time series data collected at every flight. Then, it shows how this data is managed and exploited using latest big data technologies. After that, illustrative examples that show the benefits from the analysis of this data are provided. Finally, some challenging use cases are presented, highlighting some limitations of existing tools and analysis methods.

3.16 Socio-temporal Data Mining


Abdullah Mueen (University of New Mexico, US)

License  Creative Commons BY 3.0 Unported license
© Abdullah Mueen

Modern social media produce timestamped events that can be considered as a data series to mine for patterns. We consider a large number of tweets from millions of Twitter users in a streaming manner for several years and mine clusters, motifs and cluster-dynamics. Emerged patterns represent automated user behavior, host curation events, and political events. We also show an application of pattern mining to extract hidden seismic events.

3.17 Data Series Mining and Applications

Rodica Neamtu (Worcester Polytechnic Institute, US)

License  Creative Commons BY 3.0 Unported license
© Rodica Neamtu

My research interests are at the crossroads of theoretical computer science and Big Data analytics. In this light, my work reveals that domain-specific distances preferred by analysts for exploring similarities among time series tend to be “point-to-point” distances. Unfortunately, this point-wise nature limits their ability to perform meaningful comparisons between

sequences of different lengths and with temporal mis-alignments. Analysts instead need “elastic” alignment tools such as Dynamic Time Warping (DTW) to perform such flexible comparisons. However, the existing alignment tools are limited in that they do not incorporate diverse distances. To address this shortcoming, our work introduces the first conceptual framework called Generalized Dynamic Time Warping (GDTW) that supports now alignment (warping) of a large array of domain-specific distances in a uniform manner. We further use these warped distances to explore data in diverse application domains including neuroscience, finance, healthcare and to facilitate communication for people with disabilities. My talk discusses briefly three projects in these areas and highlights the common denominator represented by the omnipresence of data series. First I showcase our work incorporating machine learning to support an Augmentative Alternative Communication app for people with several verbal and motor-skill challenges. Our LIVOX application incorporates artificial intelligence algorithms to reduce the so-called “reciprocity gap” that acts as a communication barrier between disabled people and their interlocutors, thus enabling people with disabilities, especially children, to participate in daily social and educational activities. Integrating them into the existing social structures is central to making the world a more inclusive place. Then I discuss our use of generalized warping distances to explore data for neuroadaptive technology. We show that our exploratory tool can use different similarity distances for robust identification of similar patterns in the brain data during complex tasks. This builds a foundation for interactive systems that are capable of identifying cognitive states and adapting system behavior to better support users. Lastly, I discuss a tool for automatic website content classification for the financial technology (Fintech) domain that facilitates the identification of promising startup fintec companies.

3.18 Fulfilling the Need for Big Sequence Analytics

Themis Palpanas (Paris Descartes University, FR)

License © Creative Commons BY 3.0 Unported license
© Themis Palpanas

Main reference Themis Palpanas: “Data Series Management: The Road to Big Sequence Analytics”, SIGMOD Record, Vol. 44(2), pp. 47–52, 2015.

URL <https://doi.org/10.1145/2814710.2814719>

Massive data sequence collections exist in virtually every scientific and social domain, and have to be analyzed to extract useful knowledge. However, no existing data management solution (such as relational databases, column stores, array databases, and time series management systems) can offer native support for sequences and the corresponding operators necessary for complex analytics. We argue for the need to study the theory and foundations for sequence management of big data sequences, and to build corresponding systems that will enable scalable management and analysis of very large sequence collections. To this effect, we need to develop novel techniques to efficiently support a wide range of sequence queries and mining operations, while leveraging modern hardware. The overall goal is to allow analysts across domains to tap in the goldmine of the massive and ever-growing sequence collections they (already) have.

3.19 Accelerating IoT Data Analytics through Time-Series Representation Learning

John Paparrizos (University of Chicago, US)

License © Creative Commons BY 3.0 Unported license

© John Paparrizos

Main reference John Paparrizos, Michael J. Franklin: “GRAIL: Efficient Time-series Representation Learning”, Proc. VLDB Endow., Vol. 12(11), pp. 1762–1777, VLDB Endowment, 2019.

URL <http://dx.doi.org/10.14778/3342263.3342648>

Main reference John Paparrizos, Luis Gravano: “Fast and Accurate Time-Series Clustering”, ACM Trans. Database Syst., Vol. 42(2), pp. 8:1–8:49, 2017.

URL <https://doi.org/10.1145/3044711>

The analysis of time series is becoming increasingly prevalent across scientific disciplines and industrial applications. The effectiveness and the scalability of time-series mining techniques critically depend on design choices for three components: (i) representation; (ii) comparison; and (iii) indexing. Unfortunately, these components have to date been investigated and developed independently, resulting in mutually incompatible methods. The lack of a unified approach has hindered progress towards fast and accurate analytics over massive time-series collections. To address this major drawback, we present GRAIL, a generic framework to learn in linear time and space compact time-series representations that preserve the properties of a user-specified comparison function. Given the comparison function, GRAIL (i) extracts landmark time series using clustering; (ii) optimizes necessary parameters; and (iii) exploits approximations for kernel methods to construct representations by expressing time series as linear combination of the landmark time series. We build GRAIL on top of Apache Spark to facilitate analytics over large-scale settings and we extensively evaluate GRAIL’s representations for querying, classification, clustering, sampling, and visualization of time series. For these tasks, methods leveraging GRAIL’s compact representations are significantly faster and at least as accurate as state-of-the-art methods operating over the raw high-dimensional time series. GRAIL shows promise as a new primitive for highly accurate, yet scalable, time-series analysis.

3.20 Contradictory Goals of Classification, Accuracy, Scalability and Earliness

Patrick Schäfer (HU Berlin, DE)

License © Creative Commons BY 3.0 Unported license

© Patrick Schäfer

Time series classification (TSC) tries to mimic the human understanding of similarity. Classification approaches can be divided into 6 areas: whole series, Shapelets, Dictionary and Interval, Ensembles and Deep Learning. Our research focusses on three contradictory goals of TSC, namely accuracy, scalability and earliness. Much research has gone into improving the accuracy of TSC. When it comes to long or larger time series datasets, these state-of-the-art classifiers reach their limits because of high training or prediction times. To improve scalability, a classifier has to sacrifice on accuracy. In contrast, early time series classification (eTSC) is the problem of classifying a time series after seeing as few measurements as possible with the highest possible accuracy. The most critical issue is to decide when enough data of a time series has been seen to take a decision: Waiting for more data points usually makes the classification problem easier but delays the time in which a classification is made.

3.21 More Reliable Machine Learning through Refusals

Dennis Shasha (New York University, US)

License © Creative Commons BY 3.0 Unported license
© Dennis Shasha

SafePredict is a meta-algorithm that sits on top of one or more machine learning algorithms. It takes each prediction from these algorithms (which may be weighted) and decides whether to accept or refuse to accept that prediction. Suppose that a user sets an error threshold E . SafePredict will endeavor to guarantee that among all accepted predictions the fraction of errors doesn't exceed E . Under very general assumptions, SafePredict can guarantee this. When the data points are i.i.d. (independent and identically distributed), SafePredict does even better.

3.22 Systems and Tools for Time Series Analytics

Nesime Tatbul (Intel Labs & MIT – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Nesime Tatbul
Main reference Metronome Project
URL <http://metronome.cs.brown.edu/>

From autonomous driving to industrial IoT, the age of billions of intelligent devices generating time-varying data is here. There is a growing need to ingest and analyze time series data accurately and efficiently to look for interesting patterns at scale. Our key goal in the Metronome Project is to build novel data management, machine learning, and interactive visualization techniques for supporting the development and deployment of predictive time series analytics applications, such as anomaly detection [1]. In this talk, I give three example tools that we have recently built for time series anomaly detection: (i) a customizable scoring model for evaluating accuracy, which extends the classical precision/recall model to range-based data; (ii) a zero-positive learning paradigm, which enables training anomaly detectors in absence of labeled datasets; and (iii) a visual tool for interactively analyzing time series anomalies.

3.23 Data Series Similarity Search

Peng Wang (Fudan University – Shanghai, CN)

License © Creative Commons BY 3.0 Unported license
© Peng Wang

Similarity search is a fundamental task for data series mining. In this talk, I introduce our works for both whole matching problem and subsequence matching problem, DSTree and KV-match. Also, some ongoing works and open problems are discussed.

3.24 Tableau for Data Series

Richard Wesley (Tableau Software – Seattle, US)

License © Creative Commons BY 3.0 Unported license
© Richard Wesley

Tableau is an interface for converting visual specifications into queries. To enable this, it uses a unified data model that interfaces to a large number of query engines. This model has many advantages for simplifying the user experience and integrating data, but it also leads to a lowest common denominator approach that restricts analysis to a small number of data types and makes it hard to integrate complex data types like data series and the associated query operations.

3.25 Managing and Mining Large Data Series Collections

Konstantinos Zoumpatianos (Harvard University – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Konstantinos Zoumpatianos

Main reference Kostas Zoumpatianos, Themis Palpanas: “Data Series Management: Fulfilling the Need for Big Sequence Analytics”, in Proc. of the 34th IEEE International Conference on Data Engineering, ICDE 2018, Paris, France, April 16-19, 2018, pp. 1677–1678, IEEE Computer Society, 2018.

URL <https://doi.org/10.1109/ICDE.2018.00211>

Data series management has recently gathered a great amount of attention. This is mainly driven by the large amount of sequential information that analysts both in science as well as in industry need to be able to monitor and analyze. In this talk we will look at how we can manage large collections of data series, the types of data analysis tasks that are commonly performed, and how they can be efficiently performed from a data systems perspective. Specifically, we will look at an overview of the most commonly found query templates and data mining tasks (clustering, classification, deviation detection, frequent pattern mining), specialized index structures for efficiently answering such queries, as well as pinpoint the open problems and research directions.

Participants

- Azza Abouzied
New York University –
Abu Dhabi, AE
- Anthony Bagnall
University of East Anglia –
Norwich, GB
- Anastasia Bezerianos
INRIA Saclay – Orsay, FR
- Paul Boniol
Paris Descartes University, FR
- Richard L. Cole
Tableau Software – Palo Alto, US
- Michele Dallachiesa
Minodes GmbH – Berlin, DE
- Karima Echihabi
ENSIAS-Mohammed V
University – Rabat, MA
- Jean-Daniel Fekete
INRIA Saclay – Orsay, FR
- Germain Forestier
University of Mulhouse, FR
- Pierre Gaillard
CEA de Saclay –
Gif-sur-Yvette, FR
- Anna Gogolou
INRIA Saclay – Orsay, FR
- Søren Kejser Jensen
Aalborg University, DK
- Mourad Khayati
University of Fribourg, CH
- Alessandro Longo
University of Rome III, IT
- Ammar Mechouche
Airbus Helicopters –
Marignane, FR
- Abdullah Mueen
University of New Mexico, US
- Rodica Neamtu
Worcester Polytechnic
Institute, US
- Themis Palpanas
Paris Descartes University, FR
- John Paparrizos
University of Chicago, US
- Patrick Schäfer
HU Berlin, DE
- Dennis Shasha
New York University, US
- Nesime Tatbul
Intel Labs & MIT –
Cambridge, US
- Peng Wang
Fudan University – Shanghai, CN
- Richard Wesley
Tableau Software – Seattle, US
- Konstantinos Zoumpatianos
Harvard University –
Cambridge, US



Values in Computing

Edited by

Christoph Becker¹, Gregor Engels², Andrew Feenberg³,
Maria Angela Ferrario⁴, and Geraldine Fitzpatrick⁵

1 University of Toronto, CA, christoph.becker@utoronto.ca

2 Universität Paderborn, DE, engels@uni-paderborn.de

3 Simon Fraser University – Burnaby, CA, feenberg@sfu.ca

4 Lancaster University, GB, m.a.ferrario@lancaster.ac.uk

5 TU Wien, AT, geraldine.fitzpatrick@tuwien.ac.at

Abstract

Values are deeply held principles guiding decisions of individuals, groups and organizations. Computing technologies are inevitably affected by values: through their design, values become embodied and enacted. However, some values are easier to quantify and articulate than others; for example, the financial value of a software product is easier to measure than its ‘fairness’. As a result, less measurable values are often dismissed in decision making processes as lacking evidence.

This is particularly problematic since research shows that less measurable values tend to be more strongly associated with sustainable practices than easier to quantify ones; it also indicates that the systems we design are likely to be inadequate for tackling long-term complex societal problems such as environmental change and health-related challenges that so often computing technologies are asked to address.

This seminar aims to examine the complex relations between values, computing technologies and society. It does so by bringing together practitioners and researchers from several areas within and beyond computer science, including human computer interaction, software engineering, computer ethics, moral philosophy, philosophy of technology, data science and critical data studies. The outcomes include concrete cases examined through diverse disciplinary perspectives and guidelines for values in computing research, development and education, which are expressed in this report.

Seminar July 14–19, 2019 – <http://www.dagstuhl.de/19291>

2012 ACM Subject Classification Human-centered computing → Collaborative and social computing, Social and professional topics → Computing / technology policy, Applied computing → Law, social and behavioral sciences, Social and professional topics → Computing education, Social and professional topics → Computing profession, Software and its engineering → Software creation and management, Human-centered computing

Keywords and phrases computing in society, responsible innovation, sustainability informatics
computer ethics, philosophy of technology and moral philosophy

Digital Object Identifier 10.4230/DagRep.9.7.40

Edited in cooperation with Klementina Josifovska



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Values in Computing, *Dagstuhl Reports*, Vol. 9, Issue 7, pp. 40–77

Editors: Christoph Becker, Gregor Engels, Andrew Feenberg, Maria Angela Ferrario, and Geraldine Fitzpatrick



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Maria Angela Ferrario (Lancaster University, GB)

Christoph Becker (University of Toronto, CA)

Gregor Engels (Universität Paderborn, DE)

Andrew Feenberg (Simon Fraser University – Burnaby, CA)

Geraldine Fitzpatrick (TU Wien, AT)

License © Creative Commons BY 3.0 Unported license

© Maria Angela Ferrario, Christoph Becker, Gregor Engels, Andrew Feenberg, and Geraldine Fitzpatrick

The purpose of Dagstuhl Seminar 1929 ‘Values in Computing’ was to bring together practitioners and researchers with expertise stretching beyond computer science, to include sociology, ethics, and philosophy to examine the complex relations between human values, computing technologies and society. In so doing, the seminar invited an inter-disciplinary community to share their challenges, illustrate their approaches through concrete case studies, and distil lessons learned into actionable guidelines for research and education with tangible implications for policies and industry.

The seminar was motivated by the growing urgency for computing research and industry to answer questions about the role that digital technology plays in society. The greater the scale and reach of digital technology systems, the greater their impact, both intended and unintended. Mainstream media, popular science, and the general public have only started grappling with the scale of these consequences. Many are calling institutions, professionals, and scientists to act [3]. Recent years have seen an increasing number of high-profile software scandals and malpractices in which individual privacy and democracy have been undermined (Cambridge Analytica’s use of Facebook data), the environmental impact of air pollutants disregarded (the Volkswagen’s diesel emission scandal), and human lives lost (the Boeing 737 Max anti-stall software disasters).

These events are the constant reminders that human values are indeed “the facts of the future” [1], as Feenberg argues. Values are not the opposite of facts, they become facts: the more weight we give to certain values (e.g. wealth, political influence, power), the bigger the ‘blind spots’ of the existing values become (e.g. environmental sustainability, equality and social justice). There is a pressing need then to understand how human values operate and to build on this understanding to consider how research and education might contribute to a more socially responsible computing industry.

To this end, the seminar brought together disciplines with a long tradition of critical thinking and human-centred approaches to computing with those that, such as Software Engineering, have been traditionally considered, albeit increasingly controversially, as ‘values neutral’. The breadth and depth of the interdisciplinary debate, one of the key distinguishing features and strengths of this five-day seminar, was also, and intentionally so, one of its main challenges. This was particularly evident when the need to unpack the multifaceted and often abstract notion of human values was met by the demand for the discussion to be of concrete relevance to computing education and practice. Within this context, one of the key objectives of the seminar was to facilitate both the exploration of broad themes and the identification of specific topics that would require meaningful cross-disciplinary effort. To this end, a two-pronged approach was designed to encourage both divergence and converge of viewpoints.

Thematic divergence was encouraged through six short *Seed Talks*, ten open-floor *Lightning Talks*, and a *Soap Box* session where participants would pitch high-level challenges to provoke discussion. Convergence was facilitated by *World Café* style group discussions around six emergent themes. Over the last two days, these themes were then distilled into four topics with one working group assigned to each (*Action*, *Education*, *Research*, and *Response*). Seed Talks were invited 20-minute talks designed to be informative and provocative. Thematically, they were structured around the original seminar proposal scoping areas: theory and methodologies (Feenberg and Mainzer), professional practice (Spiekermann and Whittle); and educational pathways (Nathan and Patitsas). Participants offered Lightning Talks on a variety of topics of their own choosing. For instance, Easterbrook focused on the environmental crisis and called for urgent action; Walker, McCord and Lievrouw shared their experiences of socially responsible digital activism; Frauenberger provided concrete examples on how different ways of thinking informatics in education [2]; Winter outlined the tools and techniques used to study values in software production [4]; and Jensen-Ferreira shared her approach to software industry research. Finally, four teams worked on a specific *Values in Computing* topic, each identifying a possible course of action:

1. *Action* – This group worked under the premise that the professional knowledge and critical insight of computer and social scientists should be mobilized as an active force in public education and policy-making concerning the design, implementation and regulation of information technology. With a view to these three lines of action, the group proposed the penning and wide distribution of a document, tentatively entitled “*The Dagstuhl Declaration*” here included.
2. *Research* – The Research group pursued a threefold-goal: understand the state-of-the-art of the research and highlight under-explored research areas; discuss methods and tools that have been or can be used, and identify future research directions.
3. *Education* – The Education group discussed the implications for undergraduate and graduate computing education by conducting a brief but focused exploration of existing university-level courses, methods and tools and their mapping of curriculum cross-cutting learning objectives.
4. *Response* – This group worked on the intersection between climate emergency and the future of computing and centred its activity on gathering resources about this intersection and writing an opinion piece to address it.

References

- 1 Andrew Feenberg. Ten paradoxes of technology. *Techné: Research in Philosophy and Technology*, 14(1):3–15, 2010.
- 2 Christopher Frauenberger and Peter Purgathofer. 2019. Ways of thinking in informatics. *Commun. ACM* 62, 7 (June 2019), 58-64.
- 3 Leon J. Osterweil. Be prepared. *SIGSOFT Softw. Eng. Notes*, 41(5):4–5, November 2016.
- 4 Emily Winter, Stephen Forshaw, Lucy Hunt, and Maria Angela Ferrario. 2019. Towards a systematic study of values in SE: tools for industry and education. In *Proceedings of the 41st International Conference on Software Engineering: ICSE-NIER '19*. IEEE Press, Piscataway, NJ, USA, 61-64.

2 Table of Contents

Executive Summary

<i>Maria Angela Ferrario, Christoph Becker, Gregor Engels, Andrew Feenberg, and Geraldine Fitzpatrick</i>	41
---	----

Seed Talks

Values are the Facts of the Future <i>Andrew Feenberg</i>	45
Towards Value-Based Computing Challenges in Software Engineering and AI <i>Klaus Mainzer</i>	45
Value-based System Engineering for Ethics by Design <i>Sarah Spiekermann-Hoff</i>	46
Values in the Software Industry <i>Jon Whittle</i>	47
Holding onto Disruption <i>Lisa P. Nathan</i>	47
Education and Values in Computing <i>Elizabeth Patitsas</i>	49

Lightning Talks

Technology and Neighbourhood Values <i>Ann Light</i>	49
In Search for PANDORA <i>Peter Reichl</i>	49
The Immorality of Artificial Emotions <i>Blay R. Whitby</i>	50
Ways of Thinking in Informatics <i>Christopher Frauenberger</i>	50
Tackling Digital Resignation <i>Irina Shklovski</i>	51
Mobilization and Solidarity <i>Curtis McCord and Dawn Walker</i>	51
Deconstructing Values in Computing <i>Doris Allhutter</i>	52
Theoretical and Methodological Approach to Studying the Role of Human Values <i>Emily Winter</i>	53
People Involvement in the AI System Development Life-Cycle <i>Juliana Soares Jansen Ferreira</i>	53
The Discontinuous Future <i>Steve Easterbrook</i>	54

World Café's

World Café Report: Understanding Values in Computing <i>Austen W. Rainer</i>	55
---	----

World Café Report: Research Challenges <i>Christoph Becker</i>	58
World Café Report: Values in Computing in Education <i>David Hendry</i>	60
World Café Report: From Principles to Software Industry Practice <i>Jon Whittle</i>	62
World Café Report: On Politics <i>Christopher Frauenberger</i>	64
World Café Report: Values Activism, Outreach, Mobilization, and Narratives – Learning from CPSR <i>Leah Lievrouw</i>	65
Working Groups	
Values in Computing – <i>Action</i> <i>Maria Bakardjieva, Doris Allhutter, Stefanie Betz, Gregor Engels, Andrew Feenberg, Peter Reichl, and Blay R. Whitby</i>	68
Values in Computing – <i>Education</i> <i>David Hendry, Christoph Becker, Marta Cecchinato, Teresa Cerratto-Pargman, Geraldine Fitzpatrick, Leah Lievrouw, Austen W. Rainer, Irina Shklovski, and Jon Whittle</i>	71
Values in Computing – <i>Research</i> <i>Juliana Soares Jansen Ferreira, Clarisse Sieckenius de Souza, Klementina Josifovska, Selma Lamprecht, Daniel Pargman, Barbara Russo, and Emily Winter</i>	72
Values in Computing – <i>Response</i> <i>Dawn Walker, Christoph Becker, Steve Easterbrook, Christopher Frauenberger, Ann Light, Curtis McCord, Lisa P. Nathan, Elizabeth Patitsas, and Irina Shklovski</i> . . .	76
Participants	77

3 Seed Talks

3.1 Values are the Facts of the Future

Andrew Feenberg (Simon Fraser University – Burnaby, CA)

License © Creative Commons BY 3.0 Unported license
© Andrew Feenberg

Technology is always technical and cultural. In introducing key ideas from critical constructivism [1], Prof. Andrew Feenberg placed two questions center stage: Why is technology clearly rational but simultaneously value-laden, and how can values and technical rationality co-exist? Critical theory does not reject rationality, but it rejects the idea of technical neutrality. In a famous article by Langdon Winner, he used Robert Moses' bridges in New York to articulate how "artifacts have politics" [2] – in other words, they embody and enact political and social values. In critical constructivism, the concept of underdetermination highlights that technological choices are always more than technical, taken relative to a context responding to a social world. The adaptation of technology to its environment introduces bias from society. Critical constructivism distinguishes substantive bias from formal bias – the latter is not incompatible with rationality but simply a particular form of rationality. The technical and cultural aspects of technology manifest in affordances – physical or perceived properties of an object that determine or indicate how that object can be used within the experienced, meaningful and socially shared universe of concepts, objects and actions.

In critical constructivism, technical elements are the words of the language of technology, and the technical code is the grammar that governs how larger sentences and arguments can be formed out of these elements. Through technical codes, cultural values are embodied in technical artifacts, but culture and its values appear so obvious to the actors in that process that their influence often remains overlooked. As a result, values in computing in its current capitalist context often express excluded needs of marginalized stakeholders. The public resistance to aspects of technology we encounter today highlights the possibilities of alternate futures in which those values that are now marginalized could become facts.

References

- 1 Feenberg, A. (2017). *Technosystem: The social life of reason*. Cambridge, Massachusetts: Harvard University Press.
- 2 Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109(1), 121–136.

3.2 Towards Value-Based Computing Challenges in Software Engineering and AI

Klaus Mainzer (TU München, DE)

License © Creative Commons BY 3.0 Unported license
© Klaus Mainzer

Main reference Klaus Mainzer: "Künstliche Intelligenz – Wann übernehmen die Maschinen?", Springer, Berlin, 2nd edition 2019.

URL <https://doi.org/10.1007/978-3-662-58046-2>

In the world of computing and software engineering, machine learning with learning algorithms becomes more and more powerful with exponentially growing computing capacity. Machine learning algorithms are not only applied in science and technology, but dominate business

strategies and the industrial Internet. They control the processes of a networked world in the Internet of Things. But, the state of the art in machine learning is mainly based on statistical learning and reasoning with an exploding number of parameters. These „black boxes“ need more explainability and accountability w.r.t. safety-critical systems and societal infrastructures. Obviously, with increasing complexity, the challenges of security and responsibility come to the fore. Safety-critical software demonstrates that values in computing must not be restricted to ethical values only. In practice, we must also consider the costs of testing which depend on value-based decisions. Rigorous proofs of complex software with mathematical accuracy need an immense amount of time and man-power. On the other side, it is risky to rely only in ad-hoc testing and empirical testing in the case of safety-critical systems. For certification of AI-programs, we must aim at increasing accuracy, security, and trust in software in spite of increasing complexity of civil and industrial applications, but with respect to the costs of testing (e.g., utility functions for trade-off time of delivery vs. market value, cost/effectiveness ratio of availability). There is no free lunch for the demands of safety and security. Responsible AI must find fair and sustainable degrees of certification. The author is engaged in the steering board for AI-certification (DIN/ISO) of the German government. Behind this is the fundamental insight that computing technology does not work independently of societal and civil values. At that point, humanities and social sciences come in. Without considering societal structures and processes, hardly any innovation in software engineering and AI can be successful. But, vice versa, without better understanding and explainability of computing technology, societies cannot be governed. The Internet of Things should be transformed into an Internet of Values. Therefore, questions of humanities and social sciences must be addressed right from the start in the design of computing and software technology and not only in a subsequent “add-on” that comes into play when the technology has already created facts. This talk is a plea for “Technikgestaltung” which means more than “shaping” and “governance” of technology. In short, we aim at a value-based roadmap which is no innovation killer, but the breakthrough to a responsible and sustainable, and therefore, better technology.

3.3 Value-based System Engineering for Ethics by Design

Sarah Spiekermann-Hoff (Wirtschaftsuniversität Wien, AT)

License  Creative Commons BY 3.0 Unported license
© Sarah Spiekermann-Hoff

In this talk I will first give an overview of what ethics is and what roles values play for ethics by design. I will explain what values are according to 100 years of research in material value ethics (phenomenology) and how they can be identified with the help of classical ethical theories (like utilitarianism, virtue ethics and duty ethics). Against this background my talk will include a cautious note on why “lists” of values or value principles (transparency, bias, accountability, et.) are nice to have, but incomplete to create a sustainable system design in practice. I will then outline how ethics by design can be achieved through value-based engineering. I show how innovation teams can identify relevant values for a system’s design and then ensure that value dispositions effectively end up in the technical and organizational concepts. The methods described are largely corresponding to the current work status of the IEEE P7000 group, which I co-chair and co-initiated. IEEE P7000 is planned to become IEEE’s standard for a model process for ethical system engineering¹.

¹ <https://standards.ieee.org/project/7000.html>

3.4 Values in the Software Industry

Jon Whittle (Monash University – Clayton, AU)

License © Creative Commons BY 3.0 Unported license
© Jon Whittle

Joint work of Waqar Hussain, Davoud Mougoui, Rifat Shams, Harsha Perrera, Arif Nurwidyantoro, Gillian Oliver
Main reference <https://www.slideshare.net/jonwhittle9/values-in-the-software-industry>

This talk reported on how the software industry thinks about and implements human values such as integrity, diversity, inclusion and social responsibility. The history of software engineering is a focus on developing software with particular functionality at affordable cost and that is safe, secure and reliable. This talk argues for a shift to consider also broader human values. Results were presented on two case studies with industry as to how they translate corporate values into software products. The main findings are that companies care about human values but currently have limited methods and tools for ensuring that software respects the values they care about. For more information, see <http://www.ovislab.net>.

3.5 Holding onto Disruption

Lisa P. Nathan (University of British Columbia – Vancouver, CA)

License © Creative Commons BY 3.0 Unported license
© Lisa P. Nathan

I was asked by the organizers of this seminar to address the topic “disruption”. I was encouraged to “be bold”. The boldness of this talk is not because I introduce an innovative values-in-computing design theory, methodology, or framework. Rather, the boldness is in extending an invitation. I invite you—computer scientists and academics whose research and teaching are entwined with computing technologies—to join me in facing our climate crisis. Frankly, a more appropriate term for what we face is ecocide, the destruction of our environment.

Across the globe environmental conditions are deteriorating faster than anticipated, causing suffering across species and ecosystems [3]. To gather at this prestigious scientific venue and continue to ignore the myriad connections between computing and our climate would be a stunning act of denial. I am not suggesting that this dire situation can be solved by ‘values in computing’, the topic of our seminar. Rather, in this talk I discuss how the norms of academia, computing in particular, are implicated by climate change. The accepted ways we go about research and teaching—our approach to computer science research, computer science education, applying for grants, reviewing papers, and traveling to conferences—are entangled with the shifting state of our world. Each of these activities is wrapped up with our climate. As professionals caught up in the rhetoric of digital technologies, each day we perpetuate and benefit from the forces that have led to this crisis. I say this not to assign blame, certainly not to suggest that we are individually responsible for the state of things, but to point to the incredible difficulties of extricating ourselves from the norms and expectations of our professional lives.

Scientists are unequivocal that each year will set heat records, larger and more violent storms, and rising, acidifying seas [7, 4]. Countless communities are experiencing the effects of climate disruptions, flooding, food crops failing, devastating fires, mass human migrations, etc. We are learning that early climate forecasts suffered from overstated optimism and understated threat levels [6], yet there continues to be a reluctance to appear alarmist in the

face of this disaster. Scientists' fears, my fear, of professional and public censure are part of why the issue slips from the mind [1, 2, 5]. In professional conversations I am wary of talking about it, because it makes people so uncomfortable. The topic is disruptive. Yet, the alternative is to ignore the research of climate scientists and others, which calls into question the purpose of the academic enterprise. Why is it deemed acceptable to ignore findings we find distressing?

This seed presentation is an attempt to make space within our discussions to consider the climate crisis and how it influences our work. I ask whether we can step away from discussions grounded by computing logics that privilege values of efficiency, speed, utilitarianism, control, and unlimited growth? Even if we are able to sustain our attention, should we address this emergency through the same ways of thinking, the same normative structures and systems, tropes and stories that created them? Changing dominant stories is hard, but we have models... many found in stories. I draw upon such unsettling stories in this talk. I encourage us to consider ways to face climate disruption together as the complicated, brilliant, flawed, arrogant, and creative professionals, academics, and human beings that we are. The forces that have caused this trouble, are a part of us, just as we are a part of the natural world. I argue to hold onto this knowledge of ourselves as we strive to change our ways of being in the world.

Or will we carry on as "normal", letting the research and our climate emergency slide out of our minds once again?

References

- 1 Keynyn Brysse, Naomi Oreskes, Jessica O'Reilly, and Michael Oppenheimer. 2013. Climate change prediction: Erring on the side of least drama? *Global Environmental Change* 23, 1 (February 2013), 327–337. DOI:<https://doi.org/10.1016/j.gloenvcha.2012.10.008>
- 2 J. E. Hansen. 2007. Scientific reticence and sea level rise. *Environ. Res. Lett.* 2, 2 (April 2007), 024002. DOI:<https://doi.org/10.1088/1748-9326/2/2/024002>
- 3 IPCC, 2018: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, Maycock, M. Tignor, and T. Waterfield (eds.)]. World Meteorological Organization, Geneva, Switzerland, 32 pp.
- 4 David Malmquist. 2018. Researchers issue first-annual sea-level report cards. *Phys.org/news*. Retrieved June 19, 2019 from <https://phys.org/news/2018-03-issue-first-annual-sea-level-cards.html>
- 5 George Marshall. 2015. Don't even think about it: why our brains are wired to ignore climate change (Paperback edition ed.). Bloomsbury, London Oxford New York New Delhi Sydney.
- 6 Reuters. 2019. Scientists shocked by Arctic permafrost thawing 70 years sooner than predicted. *The Guardian*. Retrieved June 18, 2019 from <https://www.theguardian.com/environment/2019/jun/18/arctic-permafrost-canada-science-climate-crisis>
- 7 M.M. Vogel, J. Zscheischler, R. Wartenburger, D. Dee, and S.I. Seneviratne. 2019. Concurrent 2018 hot extremes across Northern Hemisphere due to human-induced climate change. *Earth's Future* (June 2019), 2019EF001189. DOI:<https://doi.org/10.1029/2019EF001189>

3.6 Education and Values in Computing

Elizabeth Patitsas (McGill University – Montreal, CA)

License © Creative Commons BY 3.0 Unported license
© Elizabeth Patitsas

Education is an important means through which values in computing are spread, communicated, and encoded. In this seed talk I will be sharing and discussing two key concepts from the sociology of education: the hidden curriculum and Freire’s banking model of education. I will also be discussing Sam Breslin’s ethnography of CS education and how it teaches students to “render the world technical”, and what this means for values in computing. Finally I will encourage some critical reflection on our own values here in this Dagstuhl and the structures herein.

4 Lightning Talks

4.1 Technology and Neighbourhood Values

Ann Light (University of Sussex – Brighton, GB)

License © Creative Commons BY 3.0 Unported license
© Ann Light

I contrasted the sustainability of managing idle capacity through sharing economy tools with the merits of collective local agency bred by caring-based sharing in a locality. I described how ‘relational assets’ form and build up over time in a neighbourhood to act as local socio-technical infrastructure to sustain alternative economies and different models of trust. I proposed digital networks of support for local solidarity and resourcefulness and suggested how technologies can be pro- or anti-futures in their characteristics.

4.2 In Search for PANDORA

Peter Reichl (Universität Wien, AT)


License © Creative Commons BY 3.0 Unported license
© Peter Reichl

Recently, the necessity of an ethical consideration of the digital change has become evident also within the community of computer scientists as the responsible driving force, especially in the context of the ongoing discussion about autonomous vehicles and the ethical dilemmata they may be facing. In this way, however, only one of Kant’s notorious three questions is addressed, i.e.: What can I know? What ought I to do? What may I hope? Similarly, it seems not enough to only consider “the good”, but also other transcendentia like “the true” and “the beautiful”. Hence, I wonder whether ethics is indeed the appropriate philosophical field to deal with the issue of values in computing. Instead I believe that, eventually, this is a question of philosophical anthropology. This is in line with Günter Anders and his concept of the Promethean slope, i.e. the ever increasing gap between technological advances and human imperfection. The resulting search for PANDORA – a Philosophical Anthropology between Next-generation internet, Digito-Ontological Revolution

and Anticopernican turn (see www.homodigitalis.at) finally aims at answering also Kant's fourth and final question: what is the human being? What is the human being with respect to digital change, and what is the world we are currently building for him and her?

4.3 The Immorality of Artificial Emotions

Blay R. Whitby (University of Sussex – Brighton, GB)

License  Creative Commons BY 3.0 Unported license
© Blay R. Whitby

Research into the simulation of human emotions is a major research theme in robotics, artificial intelligence, and cognitive science. Unfortunately, there is a very high probability that this research will be used – and indeed already is sometimes being used – in ways that are clearly unethical and dangerous. Artificial emotions (AE) research will facilitate the development of technology that can obviously be used to manipulate, exploit, and abuse humans. It is also an area in which there currently exist almost no legal or ethical restrictions. At present, work, on A.E. is at the level of crude simulation of emotions BUT it has been shown to be very effective at producing emotional responses in humans even when they know that it is merely a trick. This has been demonstrated empirically for example by Cynthia Breazeal (Breazeal, C., and Brooks, R. 2005) and Briggs and Scheutz (Briggs, G and Scheutz, M. ,2102). The opportunities for exploitation and manipulation are so great that there ought to be controls, if not a complete ban, on this technology. On balance therefore, it is time to scrutinize this area of work on moral grounds and abandon the assumption that it is always beneficial.

4.4 Ways of Thinking in Informatics

Christopher Frauenberger (TU Wien, AT)

License  Creative Commons BY 3.0 Unported license
© Christopher Frauenberger

Main reference Christopher Frauenberger, Peter Purgathofer: “Ways of thinking in informatics”, Commun. ACM, Vol. 62(7), pp. 58–64, 2019.

URL <https://doi.org/10.1145/3329674>

In this short talk, I provide an overview of a University course “Ways of Thinking in Informatics”. It is a 6 ECTS university course that is mandatory for all first-year students of Informatics bachelor studies at TU Wien. It was conceptualised by Chris Frauenberger and Peter Purgathofer in 2015, and is part of the degree programs since winter semester 2017. It was inspired by “The first five computer science principles pilots”, re-interpreted through the lens of European scientific traditions. Chapters include scientific thinking, computational thinking, design thinking, critical thinking, economical thinking and responsible thinking. The aim of the course is to equip students with a range of perspectives that allows them to think about computing in different ways, enabling them to critically reflect on their education, research and practice.

4.5 Tackling Digital Resignation

Irina Shklovski (IT University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license
© Irina Shklovski

The goal of this workshop is to consider the issues of values in computing and how to not only speak about these but also to intervene into the infrastructures and practices that need to become more reflexive and aware if we are to have a better digital future than the relatively apocalyptic one that seems to be on the horizon. Yet the problems that we are tackling are so big, so looming, so distressing and so all-encompassing that at times it becomes too devastating to grasp it all. The other day one of my students asked: *How do you avoid getting really depressed when working on this?* Indeed, that is a good question. So along with considering the 'big' issues of values in design I want to raise a smaller concern that I believe is foundational as well. How do we teach about values in computing and the necessity of these considerations in ways that are not paralyzing? In efforts to intervene, education is one such intervention and I want to call for attention to approaches to teaching that can introduce critical issues in more pragmatic and practical terms that can promote more effective action from our students when they go on into the world.

4.6 Mobilization and Solidarity

Curtis McCord (University of Toronto, CA) and Dawn Walker (University of Toronto, CA)

License © Creative Commons BY 3.0 Unported license
© Curtis McCord and Dawn Walker

During the first days of Dagvic, several participants expressed an interest in the history of activism both within the computer science profession (notably that of Computing Professionals for Social Responsibility) and around activism around computational technologies more generally. Given that many at Dagvic were keen to take more action in advocating for social and environmental justice causes, we wanted to make the case that working in solidarity with existing causes and groups outside of academic computer science can be effective, as well as theoretically rich. Our lightning talk showcased three key areas of value-driven technology development and activism, and to provide some examples.

A lot of important work is being done to develop Community and Mesh Networks, with the intention of creating equitable, empowering, and resilient telecommunications infrastructures. Examples of work in this area include:

- LibreRouter ² and AlterMundi ³
- Telecomunicaciones Indígenas Comunitarias (TIC AC) ⁴ and Rhizomatica ⁵ as well as other community networking organizations.

Owing to the proliferation of ICT-mediated employment, movements advocating for digital labour rights are emerging and exploring new ways of using computational systems to organize. Examples include:

² <https://librerouter.org/>

³ <https://www.altermundi.net/>

⁴ <https://www.tic-ac.org/>

⁵ <https://www.rhizomatica.org/>

- Platform Cooperatives ⁶
- justice4couriers ⁷ and justice4foodora ⁸
- The Tech Workers Coalition ⁹ and the technwontbuildit campaign.

Work continues to be done to involve communities as active and credible participants in software and systems development work, to build local capacity, and to create information commons. Examples include:

- The Bristol Approach ¹⁰ to technology development and intervention, now being employed across Europe
- Digital Democracy ¹¹, which uses digital technologies, such as mapping systems, to help marginalize communities advocate for their rights.

4.7 Deconstructing Values in Computing

Doris Allhutter (OEAW – Wien, AT)

License  Creative Commons BY 3.0 Unported license
© Doris Allhutter

Main reference Doris Allhutter: “Mind Scripting: A Method for Deconstructive Design”, *Science, Technology, & Human Values*, Vol. 37, pp. 684–707, 2012.

URL <http://dx.doi.org/10.2307/23474485>

This talk asked what we can gain from connecting the research on values in computing with research on the normativity of computational methods and concepts. In computing, we encounter different kinds of values: epistemic values and social values, stated values or reflected values, explicit and implicit values, but often there are implicit or invisible implications of either of these values. Values are inscribed to and at the same time enacted by computational methods, concepts, and ways of thinking in multi-layered ways. They are entangled with norms and mundane beliefs and the social power relations that are coproduced through practices of computing.

This raises the question of what different methodologies we need to deal with different sorts of values and their complex entanglement with computing. I suggest that DECONSTRUCTION opens up perspectives that go beyond reflecting on values and on how to translate them to systems. It contributes to understanding the implicit entanglements of values and to reflecting and redefining concepts and methods applied in computing.

For example, recently research on bias and discrimination in machine learning and AI has emphasized the need for multi- or interdisciplinary approaches to get a grip on the complex intertwining of social power relations and technical norms and practices. Clearly, this multi- and interdisciplinary research includes different normative frameworks and ways of thinking that need to be negotiated. This is complicated by the fact that these frameworks are not fully transparent and ready for reflection. We need to ask: how do we (computer scientists, developers, inter- and transdisciplinary teams) mobilize values, norms and implicit assumptions in our practices (research and development practices)?

⁶ <https://platform.coop/>

⁷ <https://www.justice4couriers.fi/>

⁸ <https://www.foodstersunited.ca/>

⁹ <https://techworkerscoalition.org/>

¹⁰ <https://www.bristolapproach.org/>

¹¹ <https://www.digital-democracy.org/>

In my research, I use deconstruction to trace the implicit normativity of computing practices. I am currently organizing a number of workshops using a method called ‘mind scripting’, a deconstruction method based in theories of discourse, ideology, memory and affect that uncovers and negotiates the implicit assumptions in practices of computing and how they are entangled with values, norms and mundane beliefs [1, 2, 3]. In these workshops a group of six to ten participants starts either 1) from a computational problem to explore its normativity, or 2) from a value question. It also works great in teaching. If you are interested, please get in touch: dallhutt@oeaw.ac.at

References

- 1 Allhutter, D., Berendt, B., et al. forthcoming. Deconstructing Practices of ‘Debiasing in Machine Learning’, in preparation.
- 2 Allhutter, D. 2012. Mind Scripting: A Method for Deconstructive Design. In *Science, Technology & Human Values* 37(6), 684-707.
- 3 Allhutter D. & Hofmann, R. 2010. Deconstructive Design as an Approach to Opening Trading Zones. In J. Vallverdú (Ed.), *Thinking Machines and the Philosophy of Computer Science: Concepts and Principles*, Hershey/New York: IGI Global, 175-192.

4.8 Theoretical and Methodological Approach to Studying the Role of Human Values

Emily Winter (Lancaster University, GB)

License © Creative Commons BY 3.0 Unported license
© Emily Winter

This talk presented a theoretical and methodological approach to studying the role of human values in software production. Values are too often dismissed as something ‘fuzzy’. We wanted to explore how to study values in a way that could speak effectively to the software engineering community. To do this, we turned to psychological values theory (particularly the work of Schwartz and Maio) and the Q-Sort, an established card-ranking method that produces both qualitative and quantitative data. The Q-Sort exercise – as a systematic task – was appealing for industry-based software engineers, and the data allowed us to analyse the role of values at the three levels identified by Maio: the system level (how values relate to each other); the personal level (how individuals interpret values); and the instantiation level (how values are manifested through behaviours).

4.9 People Involvement in the AI System Development Life-Cycle

Juliana Soares Jansen Ferreira (IBM Brazil Research Laboratory – Rio de Janeiro, BR)

License © Creative Commons BY 3.0 Unported license
© Juliana Soares Jansen Ferreira

Joint work of Juliana Soares Jansen Ferreira, Clarisse Sickenius de Souza

Main reference Rafael Brandão, Joel Carbonera, Clarisse S. de Souza, Juliana Jansen Ferreira, Bernardo Gonçalves, Carla Faria Leitão: “Mediation Challenges and Socio-Technical Gaps for Explainable Deep Learning Applications”, CoRR, Vol. abs/1907.07178, 2019.

URL <https://arxiv.org/abs/1907.07178>

In my lighting talk, I briefly talk about my current research topic, which is related to investigate different people involved in the AI system development life-cycle. I am particular interested in professionals that have a lot of experience on designing and developing systems


that now need to adapt, create or even “forget” practices, tools, models they are used to adopt to develop system in the AI paradigm. Some research findings related to this research are found in the paper “Mediation Challenges and Socio-Technical Gaps for Explainable Deep Learning Applications”¹². I was also the author of a book that presents the SigniFYI Suite, which is a resourceful tool for the research I am doing. The SigniFYI Suite consists of a set of conceptual, methodological, and technical tools that aim to support the study of meaning-making and meaning-taking processes in software design, development and use. See details at the book “Software Developers as Users: Semiotic Investigations in Human-Centered Software Development”¹³.

References

- 1 Brandão, R., Carbonera, J., de Souza, C., Ferreira, J., Gonçalves, B., & Leitão, C. (2019). Mediation Challenges and Socio-Technical Gaps for Explainable Deep Learning Applications. arXiv preprint arXiv:1907.07178.
- 2 De Souza, C. S., Cerqueira, R. D. G., Afonso, L. M., Brandão, R. D. M., & Ferreira, J. S. J. (2016). Software Developers as Users. Cham: Springer International Publishing.

4.10 The Discontinuous Future

Steve Easterbrook (University of Toronto, CA)

License  Creative Commons BY 3.0 Unported license
© Steve Easterbrook

We speak of research agendas, curriculum change, working with practitioners, etc, as if the future world (over the next decade or so) will be like the present world. It won't be.

The next decade will be marked by a struggle for rapid transformational change throughout society, and the outcome of that struggle will determine the future of human civilization. Yet everything we've mapped out speaks of incremental change. It's a gradualist agenda that talks about working with existing companies, existing curricula, existing research labs, nudging them to take human values a little more seriously in their work.

But if you take seriously the confluence of (at least) three serious and urgent crises, it's clear we don't have time for an incrementalist approach:

- 1) The climate crisis, in which digital technology is deeply implicated. The carbon footprint of computing is growing dramatically, because we're putting the internet in everything, and it's amplifying all the worst trends of our disposable, consumerist society. Silicon valley's model of innovation (“move fast, break things, and leave others to clear up the mess”) has focussed for so long on finding new ways to monetize our data that we've forgotten what innovation really looks like. A reminder: over the next decade or so, we need to completely transform our energy infrastructure to reach net zero global emissions. We can't do this while silicon valley continues to hoover up all the available investment capital.
- 2) Automation and AI, which threatens to destroy any notion of a stable job for vast sectors of society, and which replaces human empathy for the cold, impenetrable injustice of algorithmic regulation (How do we just say “no” as a society to such technologies?).

¹² <https://arxiv.org/abs/1907.07178>

¹³ <https://www.springer.com/gp/book/9783319428291>

- 3) The dismantling of democracy, through the use of ubiquitous digital surveillance by autocrats and corporatists, and the exploitation of (addictive) social media as a vector for extremist propaganda designed to pit us against one another.

So we should be striving for a much more radical agenda that envisages the wholesale end to the technological solutionism of Silicon valley, turning it into a humble enterprise that places human dignity first. We need to dismantle the stranglehold of the big five tech corporations, break the relationship between digital technology and consumerism, and give ourselves the power to ban some technologies completely. We should not put activism in a box. As academics, activism should infuse all of our teaching, all our research, all our community engagement. If we're not working for transformational change, we're reinforcing the status quo.

Put simply, we need to recognize the unique historical moment we find ourselves in, and the role computing has played in our current existential crises.

5 World Café's

5.1 World Café Report: Understanding Values in Computing

Austen W. Rainer (Queen's University of Belfast, GB)

License © Creative Commons BY 3.0 Unported license
© Austen W. Rainer

Joint work of All participants of the seminar

The group discussions comprised five different groups discussing the prompt statement, "Understanding of Values in Computing". Each group talked for about 20 minutes. To encourage divergence, each group made notes on a separate sheet of paper, and did so without seeing the sheets completed by other groups. There were different dynamics for each group, though all groups were constructive and thought-provoking. There were a variety of contrasting ideas raised by each group. There were wide ranging discussions, consistent with a divergent approach.

Potential themes emerging from the discussions were:

1. The definition of values: What are values? How are values distinct from norms, principles etc.? (This has already been explored in a paper published by Ferrario and her colleagues [1]. Does it make sense even to try to define values precisely?)
2. Values-in-relation: How do values relate to other values, to emotions, to motivation, to promises, even to identity? How do values relate to actions, decision-making, outcomes and impact?
3. Negotiation of values: How can we negotiate, compare, align, contrast, balance values?
4. Technology (broadly defined): How can we develop new tools, methods etc (or reuse existing tools, methods etc.) to work with values, e.g., to discover them, compare them, prioritise them?
5. Example: using privacy as an example, is privacy a value? To whom is privacy a value?
6. One potential next step is to converge the divergent ideas, so that they coherently relate to computing, as per the theme of the week, Values in Computing.

Content of individual discussions:**Group 1**

1. What interests us – the Dagstuhl attendees – about values?
2. Why are values valuable to us?
3. How do we distinguish values from x, where x is from the set ethics, norms, principles, ...?
4. How do values relate to Ways of Thinking?
5. What research questions might emerge from these discussions?
6. How do assumptions and perspectives of different roles relate to values? For example, an engineer is motivated to ‘solve’ a ‘problem’ whilst a scientist is motivated to understand the world: does this suggest different values?
7. How do personal values relate to professional values? What clashes or conflicts arise between personal and professional values? What internal conflicts arise from balancing multiple personal values, or from balancing multiple professional values? Does this give rise to ambivalence toward values?
8. How do we negotiate values?
9. As an example, what values arise with regards to marriage vs values with regards to a civil partnership?
10. What tools and methods are there, or do we need, to help us work with values e.g. to discover values?
11. Other comments (as bullet points): Explicit values, Context, Relational, Political, Norms, World knowledge, Reflected values, Revealing value / discovering values.

Group 2

1. How do we understand the phrase, “Understanding of values (in computing)”? What does the phrase mean?
2. How do we align values between communities, e.g., between software developers and end users? What contrasts are there between awareness of values, negotiation of values, and alignment of values?
3. Can we – should we – talk about a universality of values? Or a plurality of values? What about a dialect of values?
4. How can we ‘stand’ in tension with values? Assuming that it is not possible to arrive at a consensus of values, how do we live with tension (conflict) between values (e.g., from different stake-holders)?
5. Unresolved tension (conflict) may itself be valuable in the sense that it is fruitful for encouraging thinking, reflection etc. In other words, tension can be positive. We can use value tension as some kind of resource or catalyst.
6. In contrast to resolving tensions, or optimising values, it may be helpful to think in terms of satisfying values.
7. Do we want to define value? Defining value risks ‘freezing’ the concept. As a contrasting example, laws sometimes do not define a word/concept (e.g., “reasonable”) leaving it to ‘practitioners’ (and lawyers) to define it in practice.
8. Assuming we cannot – or should not – define the concept of value, can we pin down the ‘corners’ of the concept enough that we can have a conversation?
9. To what degree are values embodied? We talk about understanding values but they are understood by someone... specific situated lifeworld.
10. To what degree are values situated, and contextual?
11. Who is the “us” that has a value?
12. Would it be more insightful to think in terms of valuing rather than value?
13. What is the relationship between values and promises?

Group 3

1. How do we make a general discussion about values more specific to computing, i.e., values in computing? Values can be defined simply as: what is important, when, for whom, and under what conditions.
2. Is “values” the appropriate word (concept) to use? What about values in business, or values in systems?
3. Why are we talking about values in computing?
4. Why are we motivated to be an engineer, or a scientist, or an artists? Does this motivation convey something about our values?
5. What is “values”?
6. Whose values?
7. Does the situation affect whether a person even thinks about values? Everyday situations may not require, or trigger, values. How do values relate to actions and to decision-making?
8. How do values relate to decision theory?
9. Evidence Based Software Engineering (EBSE) talks about integrating best evidence from research with practical experience and human values: is there the opportunity to connect Values in Computing with EBSE?

Group 4

1. What does the phrase “Understanding of values” mean?
2. How do values relate to requirements? A possible model might be something like: (knowledge, context, experience) → values → decision → action → outcome → impact.
3. What is the relationship of emotions to values?
4. What are the barriers or facilitators of decisions? How do these relate to, or affect, values?
5. Values are relational, with the well-known Schwarz model of values.
6. Are values multidimensional, e.g., individual – team, company – industry, society, users.
7. Is there a temporality of values?
8. What perspective ‘frames’ a value?
9. Can we develop tools that allow us to map the values of stakeholders as software progresses through the lifecycle? For example, what are the values of the project managers during the requirements phase? How do the values of developers compare with those of end-users during the acceptance testing phase? Do the values change over the course of the project to develop the software?
10. What about the decommissioning of software: what values are ‘active’ at that point? The software industry has the well-known ‘iron triangle’ of: cost, time (schedule) and quality. This iron triangle conveys three values. Another example is: information, energy, time (citation: Daniel Spreng)
11. Are values goals that are important, valid, legitimate?
12. Is there an ordering to products and values e.g., values first then products, or products first then values?
13. Is there a hierarchy of values?
14. In a conflict of values, what value/s get sacrificed?

Group 5

1. Understanding is not measurable.
2. Values are not universal (in the sense of not being static)
3. Are values what drive you to do something? In other words, values motivate?

4. Values may drive (motivate) an individual or a group.
5. Can culture be understood as values?
6. Values are not inherently positive or negative, but may lead to positive or negative outcomes or consequences.
7. Perhaps we should think in terms of valuing rather than values.
8. Given privacy as an example, then: Is privacy a value? This may depend on whether privacy is something that is important to someone. Different cultures place more or less importance on privacy (e.g., USA vs former USSR-bloc countries) suggesting that privacy is a value for some cultures and not (much of) a value for other cultures. Different contexts might also entail different relations to privacy. There are multiple conceptions of privacy. Rather than a value, privacy may be understood as a state.
9. How does privacy contrast to secrecy?
10. How does individual privacy contrast with community / collectivism?
11. There may be levels of abstraction, for example: I have an awareness of something, e.g., privacy I then decide on whether that something is valuable Technology has supported the 'deployment' of a set of values into other cultures, e.g., the values of corporate America are 'deployed' into other countries through user agreements, GDPR (do the degree that GDPR represents or contains values) has implications for non-EU countries cultures, e.g., the values of corporate America are 'deployed' into other countries through user agreements, GDPR (do the degree that GDPR represents or contains values) has implications for non-EU countries.
12. Would it be more helpful to think in terms of value systems rather than values?
13. What is the relationship of values to (personal, community) identity?
14. What is the relationship of values to emotions?

References

- 1 Winter, E., Forshaw, S., Hunt, L., & Ferrario, M. A. (2019). Towards a systematic study of values in SE: tools for industry and education. In Proceedings of the 41st International Conference on Software Engineering: New Ideas and Emerging Results (pp. 61-64). IEEE Press.

5.2 World Café Report: Research Challenges

Christoph Becker (University of Toronto, CA)

License  Creative Commons BY 3.0 Unported license
© Christoph Becker

Joint work of All participants of the seminar

5.2.1 Overview

The five rounds of discussions on research challenges were intense and broad-sweeping. Most rounds centered around one particular cluster of themes. That also means there is much unexplored territory.

A few key themes stood out.

1. **Empirical approaches** to eliciting, representing, reasoning about values from multiple angles
2. **Interdisciplinarity:** Challenges, specific modes of thinking about it, and research designs
3. **Education** research
4. The importance of **framing**, perspective, and historical understandings.

5.2.2 Empirical approaches

Given a project (recent, current, or commencing)... how can we identify, uncover, make visible, represent the explicit and the implicit values that 'really' drive (guide, shape, constrain) the project? (Q1)

We used Q1 to explore how different research designs would approach this question; what disciplinary and inter-disciplinary perspectives they would center; what questions of research design would surface; what we could learn; and what research challenges are waiting there. An interesting set of rough designs emerged quickly, covering action research, design-oriented research, project-focused case study research, mixed-methods analyses of specific phenomena, and designs focused on psychology/sociology tools for elicitation; as well as retrospection and post-mortem analysis of a project as part of reflective practice. We then used Q1 to explore what methods and research tools this type of research would require. An example (by Maria Bakardjieva) is the inscription of subject positions in platforms supporting democratic processes. One method may see technology as text to be critically deconstructed. A complementary method would perform empirical studies of projects.

5.2.3 Challenges of interdisciplinarity

We aimed to steer the discussions clear of exhaustive elaboration of all the challenges, including the barriers built into current academic disciplinary systems, but to focus instead on some aspects that seemed productive.

Ways of understanding technology-people relationships *across* disciplinary perspectives were discussed. We observe the spectrum of disciplines. Traditionally, scholars on each end of a spectrum from computing/natural science to social construction and science and technology studies have struggled to see the other end. Interestingly, new generations of scholars are more comfortable in bridging – (e.g. people using Machine Learning in Critical Data Studies; using ethnography in HCI, etc). How can this be supported? What are practical ways of bridging and of supporting the bridging?

This requires a better and more systematic understanding of the nature of interdisciplinary work: What kinds of interdisciplinary ‘regimes’, as Clarisse de Souza calls them, are more or less effective in the space of *Values in Computing*? Interdisciplinary regimes are configurations of ontology/epistemology/methodology/axiology relations. Typical patterns include the application of one discipline’s framework to address a question in another discipline’s domain; the transfer of one method to another domain; but there are many others. ‘Patterns of regimes’ for *Values in Computing* will be a very valuable resource, providing reusable research design knowledge and contextual reasoning around how what is effective.

What skills are needed to engage in this type of research? Two are identified here: (1) the ability to transcend one’s own worldview and understand the role of worldviews across disciplines; (2) the ability to navigate variations in scale of time/space of interest to disciplines.

What are the values embedded in the research methods we use and the research we do?

5.2.4 Educational research questions

An expected learning outcome we agree on is for students to become aware of the normative implications of computing.

Since there is a field ‘public understanding of science’, there should also be a field ‘public understanding of technology’. Professionalization and relates to the issues of hyper-specialization and exceeding divisions of labor, which have often brought up as a factor in the lack of understanding of computing and other technological work as political – since everyone only sees only a tiny piece of the overall systemic work, the politics of the whole remain invisible.

We would like to see solid empirical research exploring longitudinally the effectiveness of pedagogical methods and practices of getting CS students to relate their values to their technical work. This could involve comparisons across institutional settings, countries, disciplines or cultural aspects.

How can education help future professionals in escaping/transcending operationalist conceptions of “values” in computing?

5.2.5 Framing

Different framings are brought up at different points:

1. **Value as a verb.** We speak about ‘values’ as nouns a lot. What about ‘value’ as a verb? *Valuing in Computing* brings the action of the subject to the foreground.
2. **An axiology of computing?** One way to frame our work is this: We want an axiology of computing (as in ontology, epistemology, methodology, axiology).
3. Values as the frame for other issues: We seem to begin many conversations looking for the values in something – whether computing in general, or systems, or groups, or organizations... Lisa Nathan reframes it: How do values help us to understand X? (a project, a system, an issue (such as privacy, cf. VSD work). In this framing, values provide a generative way of thinking with values. For example, Nissenbaum’s work focuses on the context and situation.

5.3 World Café Report: Values in Computing in Education

David Hendry (University of Washington – Seattle, US)

License © Creative Commons BY 3.0 Unported license
© David Hendry

Joint work of All participants of the seminar

Computer science holds potential for great benefit. And, harm too, for example: Fueling consumerism through psychological manipulation of individuals, including children; Undermining democracy; Unsustainable energy and resource demands. Hence, computer science can no longer ignore such values as human dignity, human well-being, and environmental sustainability. Moving beyond engineering values – performance, scale, reliability, and correctness – computer science needs new criteria for judging the quality of systems and for holding engineers accountable.

Discerning the potential benefits and harms of the discoveries and inventions of computer science is not simply a technical matter. Nor is it a matter that is readily placed within the social sciences, law, or politics. Instead it is a matter that requires knowledge and skills for engineering, together with the social sciences and humanities.

The key question: How can students in engineering be positioned to be responsible engineers? One response is to draw on the agenda of Values in Computing to include new learning objectives that will enable students to account for human values in a principled and systematic manner throughout the design process.

Pointing the way are recent efforts by the IEEE and other professional organizations. See for example:

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems Actions*, First Edition. IEEE. ¹⁴

Like *Ethically Aligned Design*, *Values in Computing* is action- and process-oriented. By taking a design stance rather than solely an analytic, critical stance, *Values in Computing* goes well beyond “computer ethics” and “professional ethics.”

The adoption of *Values in Computing* in higher education might be achieved by considering the following areas of change:

1. Culture: Cultural change within departments of computer science. The *Values in Computing* educational agenda will require commitment from departmental leadership and professional organizations toward a systemic cultural change.
2. Curricula: New learning objectives. New learning objectives that cut across the degree program will need to be developed. Not all courses will address all new learning objectives.
3. Pedagogy: New tools for learning. Individual instructors will need support in developing and appropriating new educational methods. Note: Teaching practices in computer science are currently being questioned and seem poised to undergo a major transformation in the next 5 years.

To address these three areas of change, one conceptual approach would be to develop a *model curriculum*. The model curriculum would offer an idealized target, including recommendations on culture (commitments to a future), curricula (what to learn), and pedagogy (methods that lead to successful learning). To assist departments in adopting the model curriculum, a set of *appropriation strategies* will be needed, acknowledging that different departments will have different capacities and interests for adopting the model curriculum. When considering the international context, appropriating strategies become particularly important.

Looking to the future, the following elements might be developed to fill in the areas of culture, curricula, and pedagogy. Activist student groups. Support the development of student groups that mobilize with activist agendas. (e.g., Environmental sustainability and engineering.)

Culture: Cultural change within departments of computer science. The *Values in Computing* agenda will require cultural change within the department, catalyzed by such commitments as (examples):

1. Diversity, Inclusion, and Equity. Engaging anti-oppression strategies related to student mental health, ethnicity, gender, family conditions, physical abilities, and so forth.
2. Responsible Innovation. Students will be responsible engineers and innovators.
3. Technology is Non-Neutral. Politics matter. A critical constructivism is necessary.

¹⁴ <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>

4. Beyond Engineering Values. Engineering must move beyond efficiency, scale, reliability, correctness and other engineering values to human values such as human well-being, human dignity, and environmental sustainability.

Employer needs: Employers will seek students with skills and knowledge for Values in Computing because such students will have distinctive knowledge and skills that will give a competitive advantage. Institutions. Different institutions will have different goals, needs, and aspirations.

Learning objectives (across the degree program). Examples:

1. Demonstrate that computer programs, algorithms, and data sets are non-neutral
2. Using a rubric of social elements, write a critical reflection on the possible upsides and downsides (benefits and harms) of a proposed technology
3. Conduct a threat analysis of a software system from the perspective of vulnerable stakeholders
4. Apply methods for addressing value tensions and develop value-based rationale for design choices
5. Design a participatory workshop and demonstrate skills for moderating a workshop
6. Describe and give brief definitions to 10 human values
7. Describe 10 technical features/mechanisms that support these values
8. Apply the precautionary principle to develop a rationale for stopping a project
9. Conduct a requirements analysis that identifies values analysis that identifies values
10. Develop implementation requirements for operationalizing a small set of human values.

Students – Different students at different levels will engage the cross-cutting learning objectives to different degrees and different levels of depth.

Professional Identity- Engineering and computer science students will develop a professional identities that includes engineering competency along with socially critical mindset. Reflective Practice provides an intellectual basis [1].

Questions:


1. In technical design and engineering, what do students need to know about social science, policy, law, and governance?
2. How do we teach students to recognize the limits of their expertise?
3. What kind of language must students learn in order to be able to identify the right domain expertise they need to engage as well as how to speak to these domain experts effectively?

References

- 1 Schön, D. A. (1938). The reflective practitioner. New York, 1083.

5.4 World Café Report: From Principles to Software Industry Practice

Jon Whittle (Monash University – Clayton, AU)

License  Creative Commons BY 3.0 Unported license
© Jon Whittle

Joint work of All participants of the seminar

The group “Translation into Practice” considered how the great ideas/concepts/thoughts from the Dagstuhl seminar so far on “values in computing” could be moved into professional practice. That is, how can our community influence the software industry and other practitioner communities to take up the cause of values?

At a high level, we are aiming towards a translation from a conceptual space to an activity space: **Conceptual Space** → **Activity Space**

That is, how do we translate high level concepts around values to concrete activities that software practitioners can carry out? The answer depends on whether the values are externally imposed (e.g., by society, legislation) versus internally imposed (e.g., corporate values).

The groups came up with the following key considerations/themes.

1. **Study existing cases to identify “best” or “worst” practices**

Values are embedded in all design decisions, so we have a lot of data on which to draw on! By undertaking case studies, we can elicit examples of how (good or bad) values have been implemented in products. For “values disasters” such as Cambridge Analytica, Molly Russell, VW emissions scandal, we can reverse engineer what went wrong and come up with mitigation strategies. For exemplars, we can derive best practice guidelines.

In particular, it is important to study a range of cases from different communities: e. g., corporate versus not-for-profit versus activist developer versus open source communities. Each deal with values differently.

2. **Speak the language of the practitioner communities**

We are ultimately reliant on software practitioners to implement any values-related ideas/concepts. To be successful, therefore, we need to speak their language. Simplicity is key. Academic language and terminology will not be understood or welcomed by practitioners. We need to find a way to distil complex, nuanced notions into simple language and provide supporting guidelines, methods and tools that are super lightweight and fit into existing practices: evolution of practice rather than revolution. One example that has proven to work well in the past is the notion of a “Roadmap” that outlines the future global directions and constraints for practitioner communities to respond/fit into (cf. Industry 4.0 roadmaps). Such roadmaps may be defined at state level or at sector level.

3. **Work within the realities of industry**

Values are complex, infinite things. But different communities work within very real constraints that values work needs to consider:

- budget realities (values relevant work is often the first to be cut in tough financial times)
- timeline realities (industry cannot afford to take a long time over “properly” considering values)
- simplicity realities (see point 2 above)
- finitary nature realities (dealing properly with values is never-ending but industry must “finish” products)

4. **Influence key practices to include more reflection**

We need to influence/change key practices in management, requirements and design, learning. A key concept is self or group reflection: how to change activities of these actors to instil natural reflection on values?

- Management: values are actually very prevalent in management culture right now with lots of training provided to leaders on culture and values. How can our community improve or influence this?
- Requirements and Design: how can we adapt software engineering practices and methods to include values?

- Learning: software professionals are continual learners (by necessity, since technology changes quickly); they take both formal and informal learning opportunities. We should provide easy-to-digest learning opportunities on values (e.g., microcredentials, MOOCs).

One approach is to build communities of practice that support these. Another is to influence political channels such as the Informatics Europe Board, ACM, IEEE, BCS or The Chartered Institute for IT.

5.5 World Café Report: On Politics

Christopher Frauenberger

License  Creative Commons BY 3.0 Unported license
© Christopher Frauenberger

Joint work of All participants of the seminar

The starting point for the theme at this group was: Creating technology is an inherently political activity. How do we get computer scientists, interaction designers and technologists to recognise and engage with their political role? Saying, “I just did the tech” will not be good enough.

In five rounds, all participants of the seminar spent around 20 minutes at this table. The work by previous groups was summarised for each new round, before initiating discussions by asking participants where and when they experience their work as being political? Where and when they bring their politics to work and if they experience any barriers to do so.

The following is a thematic summary of the discussions at this table:

Identifying Political Arenas and Their Power Dynamics:

- The market (the power of the consumer/unions etc)
- Policy making (through institutions, consultation, lobbying) – steering funding, future research agendas, policy around techconsultation, lobbying)
- Teaching – (hidden) curricula, what do future technologists need?
- Public outreach – making our opinions heard in the general public
- Academia itself is a political arena (policing disciplinary borders, distribution of resources...)
- The politics of code and platforms (open source, commercial lock in)
- Politics in the making of artefacts/services

Modes and Conditions for Participation in Politics:

- Who has a voice? Who is invited and who is marginalised, not heard? Who has access to making change?
- Power structures determine leverage
- Knowledge, a forum and a standing is necessary to participate
- Infrastructures empower/marginalise participation

As academics we are in a privileged position to speak up and that come with a responsibility to speak up? However not all academics are so preiviliged, there are academic and pragmatic pressures that limit the perceived agency to bring our own politics to the research we are doing (or to our teaching). Pragmatic choices need to be made in order to safeguard one’s career/PhD. Often academics do not recognize the politics inherent in the kinds of research questions they ask or research methods they leverage and this is more common to junior scholars. This recognition is key to understanding how to position their work more effectively.

Examples in (project) work in which politics became visible: The notion of a *Moratorium*: When have we ever stopped doing a project because we thought it was a bad idea? How much personal politics can you bring to your work? It depends on the agency to speak up. A common dilemma is working with the military.

Example of challenges to bringing politics to work:

- Undermining legitimisation (you are not a scientist – but there are many other scientists that say something else, what has this to do with the science etc).
- Legal challenges when calling out unethical practices – are we protected as political actors?

Politics in Teaching: There is always a hidden curricula / agenda. Do political opinions belong into the classroom? Management often says no (which means, often political teachers seek to fly under the radar). On the other hand Universities are often places / spaces for resistance in oppressive regimes – do we feel we can/should revive this tradition?

Broadening Participation: ...both in the actual design and creation of technology and the big question of what we want (negotiating future alternatives)

A useful theoretical lens may be to think about Matters of Concern (Latour). The question may be a) what is a matter of concern and b) how do we configure people around matters of concern. Different levels of participation are possible (access to change from above and the Arnstein ladder of participation)

Politics and Expertise: In decision processes, often there is participation in a high-level political arena. However, then experts are taking over to finish the job – this is where a lot of politics happens, but it is hidden. Therefore, we need to think about *Re-politicising the experts (and thus computing experts)*

Related narratives / arguments:

Slavoj Žižek points out that when Romano Prodi was installed as head of an expert government, it was sold as neutral technocrats to the people of Italy while Prodi was a Goldman Sachs manager.

Google programmers protested and have successfully moved the company to abandon their weapons program – it needs a few millions users to shift a company's position, but maybe only a few hundred engineers

Filling the Political Vacuum: Very fundamental decisions for our society are now being made very explicitly outside the political realm (Facebook, Google, Microsoft...) How can we get them to operate in the political arena?

5.6 World Café Report: Values Activism, Outreach, Mobilization, and Narratives – Learning from CPSR

Leah Lievrouw (University of California at Los Angeles, US)

License  Creative Commons BY 3.0 Unported license
© Leah Lievrouw

Joint work of All participants of the seminar

In the first days of the Seminar, several conversations were sparked among participants who recalled the outreach efforts and effectiveness of the organization Computer Professionals for Social Responsibility (CPSR). Although CPSR was founded in the early 1980s and

officially disbanded in 2013, its influence in setting a value-driven agenda for computing practice, particularly in the U.S., was considerable. It was the launching point for spin-off advocacy organizations like the Electronic Privacy Information Center (EPIC)¹⁵ and the Electronic Frontier Foundation (EFF)¹⁶, both of which continue their work today as highly visible and effective public voices in computing-related social controversies. CPSR also launched the highly-regarded Participatory Design Conferences (PDC, now run under the aegis of the Design Research Society¹⁷, which are still held biannually, and the Directions and Implications of Advanced Computing (DIAC) seminars, both of which have published proceedings. CPSR also sponsored the prestigious Norbert Wiener Award for Social and Professional Responsibility, whose recipients included Douglas Engelbart, Joseph Weizenbaum, Barbara Simons, and Mitch Kapor. The Dagstuhl participants wondered what, if any, aspects of CPSR might serve as a model, or whether new strategies, approaches, and techniques might be needed to launch a sustainable values-in-computing movement, encourage values activism, and influence professional practice in today's cultural and technical environment dominated by social media and new "leaderless" modes of movement organizing. Based on preliminary seminar discussions that identified values activism and the need for narratives to help communicate a values-in-computing agenda as a key theme for exploration, a world café exercise was conducted on the second afternoon of the seminar. Five different groups of participants (about 25 people in all) engaged in brainstorming sessions to articulate the essential aspects of values activism that might be incorporated into the Seminar's outcomes and products. Predictably, the discussions ranged widely, from the highly theoretical to on-the-ground pragmatism. However, three broad clusters of ideas/concerns emerged, which can be characterized as "the who," "the how," and "the what" of values activism.

Who plays a role or is a relevant stakeholder in values activism? A principal concern of the world café participants, perhaps unsurprising given the crossdisciplinary quality of the groups, was identifying potential key players in movement organizing and collective action around values in computing. The first question in this respect was whether computing professionals should be the primary constituents (or perhaps "vanguard," in the language of social movement studies), or whether a looser coalition structure among different groups with allied aims, for example data activism or civic hacking movements, would be more effective. CPSR was conceived and operated very much as a professional vanguard, setting agendas and raising visibility for values and ethical issues in computing, but arose in response to an academic and private-sector work environment dominated by defense funding and firms, and specifically U.S. President Ronald Reagan's Strategic Defense Initiative (SDI, popularly criticized as "Star Wars"). The situation today, with widespread public engagement with computing and the domination of a small number of highly concentrated commercial technology firms and platforms, may require a different approach.

1. Should any activist efforts be led by credentialed, "legitimate" or organized professionals only? e.g., ACM Computers & Society SIG; Partnership for AI; EUSSET.org; LCA, SLCA; relevant government or regulatory agencies; etc.
2. Or would it be more useful to form alliances or coalitions with grass-roots, "amateur" or volunteer activists, or solidarity with "worker" movements. For example, Tech Workers Coalition, right to repair movement, FemPower Tech or Extinction Rebellion. Furthermore, in relation to the public sectors – librarians, civil service, teachers and cultural workers.

¹⁵ <https://www.epic.org>

¹⁶ <https://www.eff.org>

¹⁷ <https://www.designresearchsociety.org/events/participatory-design-conference>

Next, the gig workers, hackers & bug bounty hunters, makers, technology service workers, data activists and civic hackers, open source or FLOSS communities, Fridays for the Future can also be considered.

3. Enlistment of private-sector technology firms with strong commitments to corporate social responsibility (CSR) or maintaining public goodwill might also be useful – e.g., participants in “Partnership for AI” joining industry and advocacy organizations.
4. Movement/action repertoires may be generational; What worked for CPSR may no longer be sufficient to mobilize across stakeholder groups.

Participants also identified potential problems with the tension between a more centralized vanguard and broader coalition structures, namely problems with maintaining solidarity across groups with different aims, and the risk of diluting a clear values agenda, message or narrative.

How to launch, organize, and mobilize collective actors and action to maximize engagement with relevant stakeholders/publics? What possible action strategies & tactics or activist “repertoires” might be employed?

Another major theme of the world café exercise on values activism was what activist interventions, tactics and methods might achieve, which movement objectives. A range of possibilities were articulated, for example:

1. Direct engagement with and education of user publics (e.g., cryptoparties, teaching people how to stay safe or preserve their privacy online, how to understand user agreements or complex financial technologies, designing curriculum or teaching plans for schools, art practice/exhibitions, producing online media such as a YouTube channel or full-length documentaries like Terms and Conditions May Apply).
2. Building on the current public sense of outrage or powerlessness with respect to digital technologies, for example by participation in public meetings/town halls/demonstrations, lobbying campaigns to inform and influence politicians & regulators, creating media packages or writing op-ed pieces, etc.
3. Develop a values agenda and teaching tools for professional education in relevant disciplines (another major theme of the Seminar).
4. Organizing conferences, symposia, opportunities for sharing experiences and “what works”.
5. Identifying funding sources for organizing: public, foundation, relevant industry sources

What in fact is the values agenda? What needs to be changed? What futures do we envision? The third and most difficult aspect of the values activism world café exercise turned out to be the articulation of the values-in-computing agenda itself, given that no specific values or perspectives had yet been formulated by the Seminar in plenary discussions. As one participant put it, “Aren’t we supposed to be answering this in the Seminar this week?”

1. The spectrum from “What is to be done?” (Lenin) to the ironic “What is our one demand?” (Occupy) is wide: where does a “values in computing” movement lie on the spectrum?
2. Two major matters of concern in public discussion now are (1) Tech industry concentration and monopoly behavior, and consequences; (2) Pervasive capture and exploitation of personal data, especially to steer political power and outcomes.
3. What theories or values frameworks might be the basis on which to build a values agenda? Human rights? Care ethics? Social justice? Economic equity/political economy? Environmental justice? Place/groundedness?

4. Values perspective from the legacy of participatory design (PD), originally linked to unions and labor control over technical work: PD has diffused differently in different cultural, economic and national contexts but we might identify what values embedded in PD that were once “radical” have now become normative. Doug Shuler’s leadership of the Computers & Society SIG may be instructive.
5. Possibilities include the right to repair; transparency; fairness/equity; “post-growth,” cooperative action not competition; justice; voice; awareness of place or groundedness; sustainability; responsibility; the German Academy of Sciences focus on “responsibility,” etc. Each may “mobilize different parts of the world”.

6 Working Groups

6.1 Values in Computing – Action

Maria Bakardjieva (University of Calgary, CA), Doris Allhutter (OEAW – Wien, AT), Stefanie Betz (HFU – Furtwangen, DE), Gregor Engels (Universität Paderborn, DE), Andrew Feenberg (Simon Fraser University – Burnaby, CA), Peter Reichl (Universität Wien, AT), and Blay R. Whitby (University of Sussex – Brighton, GB)

License © Creative Commons BY 3.0 Unported license
 © Maria Bakardjieva, Doris Allhutter, Stefanie Betz, Gregor Engels, Andrew Feenberg, Peter Reichl, and Blay R. Whitby

Premise: This group worked under the premise that the professional knowledge and critical insight of computer and social scientists should be mobilized as an active force in public education and policy-making concerning the design, implementation and regulation of information technology. Professionals working in these fields are key players in the shaping of computer systems and applications. Therefore, their stance and their voices are able to make a decisive difference. In our view, there are several main avenues to achieve that outside of our professional activities as researchers, innovators and educators.

1. Public education: Active participation in public education including talks, media publications, events and collective organizing oriented toward critical assessment of the social effects of information technologies and the social practices arising around them.
2. Activist mobilization of professional communities: Taking a leading role in raising awareness and social accountability in the professional communities of information technology developers, analysts and practitioners with regard to the social consequences of new and existing technologies and systems.
3. Political pressure: Organizing and taking steps to exert collective pressure on political institutions and industry players in the direction of recognizing, counteracting and preventing the negative social effects of information technologies.

With a view to these three lines of action, the group proposed the penning and wide distribution of a document, tentatively entitled “The Dagstuhl Declaration.” The purpose of the Declaration is to succinctly outline the key concerns shared by critically-minded computer and social scientists regarding the undesirable social impacts of current trends in information technology systems, practices and policies (diagnosis); to propose directions for change aligned with progressive social values such as human dignity, well-being, equity and sustainability (prognosis); to formulate compelling reasons why political institutions need to be pressured by the professional communities working in these areas and the public at

large (motivation); and to serve as an anchor for the emergence of collective identification and agency – a “we” – including the variety of actors who are willing and capable of exerting such pressure.

The overall structure of the Declaration and its main points were drafted by the group as follows:

DAGSTUHL DECLARATION: Content Outline

We, the participants in the Dagstuhl Seminar on Values in Computing representing researchers and professionals from disciplines such as computer science, system and software engineering, ethics, philosophy, education, science and technology studies, hereby state our deep concern with the present status of the information technology industry and its influence on society. Another digital future is possible!

There are numerous reasons to question whether information technology is a force for good in contemporary society. The structure and design of the most widely information technology systems, and the regime of their ownership and control, have proven to generate major social problems. Users and consumers are subjected to new modes of pervasive exploitation and surreptitious control. Vulnerable populations are being hurt. Inequalities of access and power are growing. The lack of accountability of major corporations is unprecedented. Governments and regulatory agencies are falling behind, unable to steer the course of technological systems and practices in the public interest. Data have emerged as a major source of economic and political power. Algorithms have demonstrated their capacity as powerful tools. When left unchecked, power of this sort corrupts absolutely. Unsurprisingly, we have seen numerous demonstrations of blatant abuse of data and algorithm power.

We, the signatories of this declaration, believe that urgent measures are needed for bringing data and algorithm power under public control. To avoid the further regress of our liberal-democratic societies, human and democratic values need to be infused into information technology design and application by way of policy, regulation and explicit codes of ethical conduct on the part of information technology creators and users. Ethics should be an essential dimension of computing practice and standard development. Computer professionals, corporations and researchers should be held accountable not only to their employers and investors for the profitability of their products, but to the public at large for the ethical, social and cultural repercussions of their work. Business models relying on the manipulation of users should be challenged and dismantled. In computing, business values should be balanced with the values of human dignity, social well-being, equity and sustainability. Human dignity means respect for persons as opposed to reducing them to manipulable things. Social well-being means increasing individual and collective capacities for freedom, recognition, happiness and fulfillment. Equity refers to ensuring equal access to choices and opportunities across social and cultural groups. Sustainability refers to cultivating modes of production and consumption compatible with the preservation of the natural environment and the happiness and prosperity of other human beings.

With a view to the plethora of examples of corporate irresponsibility, negligence and disregard for these values on the one hand, and the subservience and docility of public administrations and politicians, we believe that grassroots political mobilization is needed to defend and fight for securing their dominant position in information technology design and application worldwide. We are convinced that the communities of computer professionals and social scientists can play the role of the leading agent of such a mobilization. We can be that agent. It is our moral responsibility to speak up and act now.

The working group proposes the following key demands to be included in the Declaration:

- Make manipulative targeting of users illegal.
- Raise critical public awareness; engage civil society; mobilize resistance to misuse of computing power.
- Demand accountability & transparency from ICT & data processing companies: what affects the public should be open to public scrutiny.
- Legally institute data rights and tools for their enforcement.
- Break up the monopolies so as to favour competition and creativity.
- Provide resources for democratizing socially responsible and environmentally sustainable innovation;
- Encourage and fund alternative ICT projects;
- Institute oversight in technology applications to ensure social justice and equality and protect the vulnerable. The vulnerable R us.
- Institute forms of broad democratic participation in computer systems design and in policy-making related to information technology.
- Demand carbon neutral technologies and processes in the IT industry.

DAGSTUHL DECLARATION: Addressees

The finalized, approved and signed text of the Declaration should be formally submitted on behalf of all co-signatories to the following types of addressees demanding formal response:

- Professional organizations
- Government agencies
- Intergovernmental bodies
- Civil society organizations, groups, movements
- EU bodies
- Industrial bodies
- Corporations
- Labour organizations
- Student unions and other organizations

Organizations on the following list should be approached with priority:

- ACM
- IEEE
- Informatics Europe
- Acatech
- BCS
- CRA
- CEPIS
- IFIP
- Corporate leaders and employee organizations at Google, Facebook, Apple, Microsoft, Amazon, IBM, Huawei

DAGSTUHL DECLARATION: Circulation

The group recommends that individual participants personally present the Declaration to the professional and civic organizations they belong to as well as to governmental and administrative contacts they can access firsthand.

To achieve wide public circulation of the Declaration, the group recommends the following steps:

- Publish and comment it in the mass media
- Share it on existing social media sites dedicated to similar causes
- Create social media sites for this purpose
- Web site
- Facebook page;
- Twitter hashtag
- LinkedIn designation
- Create a video – YouTube
- Address the existing alternative social media platforms
- Address the profiles of individual politicians; parliaments
- Consider an online petition

Practical steps:

- Collectively write and edit a rhetorically compelling document.
- Share this brief with all participants in the seminar, invite input, invite suggestions to be entered in a collaborative document using track-changes
- Distribute it to the organizations we belong to, relate to, or know people in . . .
- Look for additional organizations that are potential allies
- Look for organizations that have ethics or ethical in their title in relation to information technology
- Mobilize our social & professional networks
- Crowdsourcing the work among the Dagstuhl participants

6.2 Values in Computing – Education

David Hendry (University of Washington – Seattle, US), Christoph Becker (University of Toronto, CA), Marta Cecchinato (University of Northumbria – Newcastle upon Tyne, GB), Teresa Cerratto-Pargman (Stockholm University, SE), Geraldine Fitzpatrick (TU Wien, AT), Leah Lievrouw (University of California at Los Angeles, US), Austen W. Rainer (Queen's University of Belfast, GB), Irina Shklovski (IT University of Copenhagen, DK), and Jon Whittle (Monash University – Clayton, AU)

License © Creative Commons BY 3.0 Unported license

© David Hendry, Christoph Becker, Marta Cecchinato, Teresa Cerratto-Pargman, Geraldine Fitzpatrick, Leah Lievrouw, Austen W. Rainer, Irina Shklovski, and Jon Whittle

The Values in Computing Education committee met for about 8 hours to discuss the implications of Values in Computing for undergraduate and graduate education. The committee members divided into four sub-committees. Each committee considered a different topic and wrote a report, presented below.

1. Brief Exploration of University-level Courses. Sub-report 1¹⁸ presents a review of existing courses that address values in computing. One key finding is that many or most existing courses concern “ethics” rather than “human values.” The report concludes with four questions intended to be used in a detail analysis of the courses that cover “ethics,” “human values,” and related topics. This work makes progress on a comprehensive review of how values are being engaged in computer science and technical education.

¹⁸ <https://pads.c3w.at/file/#!/2/file/vo8vEOUK0V5JKwlgngH7nZy5/>

2. Methods, Practices, and Tools. Sub-report 2¹⁹ presents a list of over 30 pedagogical approaches for engaging ethics and values in education. The approaches are divided into three categories: (1) Awareness raising; (2) Understanding design and development through a values lens; and (3) Group work and reflective practitioners. For instructors looking for straightforward methods for engaging with values in their existing courses, the list of 30 approaches is an excellent starting point.
3. Framing Considerations for Learning Objectives. Sub-report 3²⁰ presents a list of 7 general pedagogical aims. Highlighting the need for a systemic shift in curriculum design and student support, these aims are intended to shape the development of specific learning objectives in a Bachelor's degree in data science that draws on the values in computing agenda. Examples include:
 - A main goal – Ensuring that students develop competencies in translations between societal concerns and information systems;
 - A universal focus – A focus on a small number of universal values;
 - Responsible innovation – Considering how the language of responsible innovation can be brought into teaching data science;
4. Curriculum Cross-cutting Learning Objectives. Sub-report 4²¹ proposes 13 specific learning objectives, divided into six categories (Create, Evaluate, Analyze, Apply, Understand, Remember). Examples include: (1) Be able to redesign existing systems to better manifest social considerations (Create); (2) Be able to evaluate competing design decisions against social considerations (Evaluate); and (3) Understand that technology design choices inevitably influence how we live (Understand). This work demonstrates how a set of learning objectives can cut across a computer science curriculum.

6.3 Values in Computing – Research

Juliana Soares Jansen Ferreira (IBM Brazil Research Laboratory – Rio de Janeiro, BR), Clarisse Sieckenius de Souza (PUC – Rio de Janeiro, BR), Klementina Josifovska (Universität Paderborn, DE), Selma Lamprecht (Fraunhofer FOKUS – Berlin, DE), Daniel Pargman (KTH Royal Institute of Technology – Stockholm, SE), Barbara Russo (Free University of Bozen-Bolzano, IT), and Emily Winter (Lancaster University, GB)

License © Creative Commons BY 3.0 Unported license

© Juliana Soares Jansen Ferreira, Clarisse Sieckenius de Souza, Klementina Josifovska, Selma Lamprecht, Daniel Pargman, Barbara Russo, and Emily Winter

Goal: The goal of the workshop is threefold: 1) to understand the state-of-the-art of the research and possibly highlight missing areas of exploration, 2) to discuss methods and tools that have been used or can be used by research in the area and 3) to suggest future directions for research in the area. Given the timeframe of the workshop, all workshop outcomes can only be considered to be starting points for future research. There is definitely more work to be done in the area; our proposals are to be seen as first steps. The working group was heterogeneous and consisted of senior and junior researchers in Human–Computer Interaction, Sustainability and Software Engineering as well as representatives from industry.

¹⁹ <https://pads.c3w.at/file/2/file/Hf2ZVPd0uTE7hIJZgiMZFahg/>

²⁰ <https://pads.c3w.at/file/2/file/2p3nQ6tvnUtwNk9D7NRnPXqH/>

²¹ <https://pads.c3w.at/file/2/file/5zcaJeT3yWsCwypeqkY-++6S/>

Work method: The work took as its starting point a formulation from the organisers' Dagstuhl application: "What tools and techniques (e.g. values maps, questionnaires, users' stories, and case studies) can support the representation, articulation, negotiation, and observation of values from the very early stages of requirements elicitation to the final appropriation of a technology?" After an initial round of presentations and brainstorming, the work was driven by two major research questions (see below). Then examples from research and industry drove the discussion these and associated questions that arose. The work was finally summarized (see below) by a set of outputs including emerging results/burning questions/areas of exploration.

Research questions:

RQ1. What are values in computing?

RQ2. What research frameworks and methods exist for doing research on values in general and in computing?

Discussing the research questions: RQ1. What are values in computing?

The members believe that this question can be answered in a pragmatic way. A few different proposals were discussed: values can originate in beliefs as well as in biases [6]; generic frameworks for analyzing values and values relations exist in literature and can be used as a starting point for discussions. Such frameworks can, if necessary, be adapted and customized to the computing field, or existing frameworks can be applied to scope and define values in computing. As such, further questions need to be addressed such as: are generic values appropriate for computing? Are known relations (still) valid when speaking about technology and technology adoption? Do values emerge or evolve with the evolution of the computing knowledge and technology? The working group suggests that future research develops a systematic literature review on values in computing.

RQ2. What research frameworks and methods exist for doing research on values in general and in computing?

Of general values frameworks, the group is aware of the Schwartz's values circumplex that has been relatively widely applied to computing [11, 12, 19, 18, 19], and Values Sensitive Design (VSD), developed specifically for computing [7, 8, 9]. The group concludes that there is a need to identify existing general frameworks and tools that have been adapted to computing (e.g., Schwartz's values circumplex), and frameworks and tools that have been specifically developed in/for computing (e.g. VSD).

The discussion then developed in two directions:

1. How to detect / identify values?
2. What are the major aspects the research must take into account?

How to detect/identify values? Research must explore different sources of information like documents (e.g. software requirements, design artifacts) developed with a technology as well as people's practices (e.g., code) to elicit implicit knowledge about values in computing. By comparing different sources of information at different stages of technology development and maintenance, research can surface gaps between attitude and practice. To identify values from different unstructured sources of information, research may apply tools like the economy of conventions [5]. What are the major aspects the research must take into account? The group engaged in lengthy discussion of this question. The interdisciplinarity and heterogeneity of the group favored a rich discussion needed to align different research culture and vocabulary. The group proposes five interrelated aspects that future research must consider: Process, People, Computing System, Context and Environment. The research must also be performed at different levels of people interactions: individual/community/organization/society ... etc.

Recommendations of Research Future Directions:

1. Perform a systematic literature review on values and values in computing to understand the state-of-the-art.
2. Explore roles and relations between “beliefs,” “values,” and “bias”. In the process of understanding values in computing, one approach is to start from stakeholders’ beliefs which can be collected with interviews, surveys or a literature review. We hope that by answering the question *how are values built?*, we can eventually understand what are values in computing. Research questions can motivate the research in this respect: Can values be grounded on beliefs? Can beliefs evolve into values? For example, one of the major beliefs of software developers is “developing software is a creative activity.” Can this belief evolve into values related for instance to the freedom of developing software (e.g., Free / Libre Software and the Free Software Foundation)? What is the role of bias in relation with beliefs and values? In the above example, can creativity be biased by some sort of overestimation of the actual work of a software developer? Finally, if any relation between beliefs, bias and values exists, what is the gap (qualification) between beliefs / bias and the related values? Can this gap by any means be associated to the renowned “attitude – behavior gap” [13]? An example of attitude – behavior gap is when a user can have an attitude towards specific sustainable technologies, but they may purchase the ones that are more convenient or affordable. In general, this aspect is looking for other concepts connected with values. Apart from beliefs and ethics, this also includes the notion of social conventions. By analyzing justifications, the economy of conventions [2, 5] could offer an interesting framework to consider the construction and weighting of values in practices [4].
3. Perform research on five key dimensions: Process, People, Computing System, Context and Environment. The working group believes that these are the major aspects that need to be taken into account. A computing system is a system that is conceived, developed, orchestrated, operated and maintained by computing technologies in a modern environment and for which technology is the essential building block (i.e., without technology the system cannot perform any of the above mentioned activities). Process refers to the whole set of inter-related activities that lead to the production, deployment, operation and maintenance of a computing system. People refers to any person that has any interest in or may be impacted by the Computing System, that is system developers, users, stakeholders or even people that do not even acknowledge that the system exists. Context is the circumstances (e.g., purpose, use or behaviour) in which the computing systems operates and under which is observed (e.g., domain of application). A specific context also defined the meaning of the system itself (e.g., an ERP system can operate differently in different context of use). Environment is the ambient that interacts with the computing system. Both context and environment can evolve. For example, in the case of a low-energy website ²², the environment comprises all settings that enable the system to interact with the external components (e.g., connection to solar cells).
4. Study the evolution of values:
 - Values in computing evolve and emerge from practice. Technology triggers evolution of values and redefines values on the go. Research must acknowledge the dynamic aspect of values. Is it technology that motivates evolution or is it societal evolution that requires the evolution in computing? Case studies in both directions are needed. What are the effects of a technology on the society and its values? What is the role of practice in this relation? Values can emerge or adapt from conflicts and confrontations due to

²² <https://www.lowtechmagazine.com/2018/09/how-to-build-a-lowtech-website.html>

daily practice [10, 1]. What are the circumstances in which conflicts and practice have any effect on values in computing? An example of values conflicts and confrontations can be identified once software developers realize that their “coding decisions” might have a real impact on other people’s lives and that they are not used to consider the “big picture” of the software they produce [3].

- Technology can be used to other aims than the original ones. Research must consider that evolution, use, people, context and environment may change the original nature of a computing system. For example, the Open Source Software (OSS) initiative was born with the intention to offer a European economic strategy of technology production and distribution that can help differentiate the software market at the time dominated by the US production. To enforce such strategy the EU government has for long supported the OSS development [14]. Thus, the OSS initiative was born with a clear commercial intent. Today, instead, in many official speeches, OSS is presented as a technology that increases IT literacy and culture and frees users from legacy systems and vendors (lock-in effect) [15, 16].

The group found this change of the nature and value of a technology often happens in the so called “social discourse” where technology is presented for the rights and wrongs of society.

References

- 1 Berenbach B. & Broy M.. (2009) Professional and Ethical Dilemmas in Software Engineering, Computer, pp. 74-80
- 2 Boltanski L. & Thevenot L. (1991) De la justification: Les economies de la grandeur (2006) On Justification: The Economies of Worth, NRF Essais, Gallimard
- 3 Brandão R., et al. Mediation Challenges and Socio-Technical Gaps for Explainable Deep Learning Applications: <https://arxiv.org/abs/1907.07178>
- 4 Cappel V. & Kappler K. E. (2019) Plurality of values in mHealth: Conventions and ethical dilemmas. in: The futures of EHealth – Social, Ethical and Legal Challenges, Ed.: Bächle/Wernick. Berlin: HIIG
- 5 Eymard-Duvernay, François, et al. (2005) Pluralist integration in the economic and social sciences: The economy of conventions.“ Post-autistic economics review 34.30, 22-40.
- 6 Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3), 330-347
- 7 Friedman, B. (Ed.). (1997). *Human values and the design of computer technology* (No. 72). Cambridge University Press.
- 8 Friedman, B., Kahn, P. H., & Borning, A. (2008). Value sensitive design and information systems. *The handbook of information and computer ethics*, 69-101.
- 9 Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. Mit Press.
- 10 Fleischmann K. & Wallace W.. (2010) Value Conflicts in Computational Modeling. *Computer* 43, 7, 57-63.
- 11 Knowles, B. (2013). Re-imagining persuasion: designing for self-transcendence. In *CHI’13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2713-2718). ACM.
- 12 Knowles, B., Blair, L., Walker, S., Coulton, P., Thomas, L., & Mullagh, L. (2014, June). Patterns of persuasion for sustainability. In *Proceedings of the 2014 conference on Designing interactive systems* (pp. 1035-1044). ACM.
- 13 Rogers, E.M. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.
- 14 Rossi B., Russo B., Succi G.. (2006) COSPA (consortium for studying, evaluating, and supporting the introduction of open source software and open data standards in the public administration), DG.O 2006: 153-154.

- 15 Rossi B., Russo B., Succi G.. (2007) Open Source Software and Open Data Standards as a form of Technology Adoption: a Case Study. *OSS 2007*: 325-330.
- 16 Rossi B., Russo B., Succi G.. (2012) Adoption of Free/Libre Open Source Software in Public Organizations: Factors of Impact. *IT & People* 25(2): 156-187.
- 17 Winter, E., Forshaw, S., & Ferrario, M. A. (2018). Measuring human values in software engineering. In *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (p. 48). ACM.
- 18 Winter, E., Forshaw, S., Hunt, L., & Ferrario, M. A. (2019). Towards a systematic study of values in SE: tools for industry and education. In *Proceedings of the 41st International Conference on Software Engineering: New Ideas and Emerging Results* (pp. 61-64). IEEE Press.
- 19 Winter, E., Forshaw, S., Hunt, L., & Ferrario, M. A. (2019). Advancing the study of human values in software engineering. In *Proceedings of the 12th International Workshop on Cooperative and Human Aspects of Software Engineering* (pp. 19-26). IEEE Press.

6.4 Values in Computing – Response

Dawn Walker (University of Toronto, CA), Christoph Becker (University of Toronto, CA), Steve Easterbrook (University of Toronto, CA), Christopher Frauenberger (TU Wien, AT), Ann Light (University of Sussex – Brighton, GB), Curtis McCord (University of Toronto, CA), Lisa P. Nathan (University of British Columbia – Vancouver, CA), Elizabeth Patitsas (McGill University – Montreal, CA), and Irina Shklovski (IT University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license

© Dawn Walker, Christoph Becker, Steve Easterbrook, Christopher Frauenberger, Ann Light, Curtis McCord, Lisa P. Nathan, Elizabeth Patitsas, and Irina Shklovski

When the group forming round the theme of “Action” grew larger, a number of people peeled off to form a second action group to work on the intersection between climate emergency and the future of computing.

This intersection of themes is more significant than it might appear – two related challenges that put the stable futures and wellbeing of billions of people at risk. So, while the Dagstuhl group as a whole commented on the need for greater consideration of values in computing, this group, reflecting their deeply held concerns, addressed the intersection of two global challenges: a capitalist computer industry and rampant climate change. This group centred activity on gathering resources about this intersection and writing an opinion piece to address it. Our argument is, in a nutshell, that the computer industry is using up a significant part of the world’s resources, creating a major carbon footprint and designing products and systems that actually reduce the capacity of societies to respond adequately to climate change, despite controlling tools that could lead to global collaborative action on this and other crises. Therefore, this group worked on an appeal for a coordinated joined-up response to the existential threat facing us. The resultant appeal:

1. Criticizes the abuse of eco-social capabilities that technology is amplifying – the manipulation of politics, individualisation of cultures and transactionalisation/quantification of relations;
2. Demonstrates how grassroot actions are already pointing the way towards an ecologically responsible green computing;
3. Calls for a computing industry that works to support social cohesion, curb growth for its own sake, promote fair work and adopt sustainable sources of production.

Participants

- Doris Allhutter
OEAW – Wien, AT
- Maria Bakardjieva
University of Calgary, CA
- Christoph Becker
University of Toronto, CA
- Stefanie Betz
HFU – Furtwangen, DE
- Marta Cecchinato
University of Northumbria –
Newcastle upon Tyne, GB
- Teresa Cerratto-Pargman
Stockholm University, SE
- Clarisse Sieckenius de Souza
PUC – Rio de Janeiro, BR
- Steve Easterbrook
University of Toronto, CA
- Gregor Engels
Universität Paderborn, DE
- Andrew Feenberg
Simon Fraser University –
Burnaby, CA
- Maria Angela Ferrario
Lancaster University, GB
- Geraldine Fitzpatrick
TU Wien, AT
- Christopher Frauenberger
TU Wien, AT
- David Hendry
University of Washington –
Seattle, US
- Klementina Josifovska
Universität Paderborn, DE
- Selma Lamprecht
Fraunhofer FOKUS – Berlin, DE
- Leah Lievrouw
University of California at Los
Angeles, US
- Ann Light
University of Sussex –
Brighton, GB
- Klaus Mainzer
TU München, DE
- Curtis McCord
University of Toronto, CA
- Lisa P. Nathan
University of British Columbia –
Vancouver, CA
- Daniel Pargman
KTH Royal Institute of
Technology – Stockholm, SE
- Elizabeth Patitsas
McGill University –
Montreal, CA
- Austen W. Rainer
Queen's University of
Belfast, GB
- Peter Reichl
Universität Wien, AT
- Barbara Russo
Free University of Bozen-
Bolzano, IT
- Irina Shklovski
IT University of
Copenhagen, DK
- Juliana Soares Jansen Ferreira
IBM Brazil Research Laboratory
– Rio de Janeiro, BR
- Sarah Spiekermann-Hoff
Wirtschaftsuniversität Wien, AT
- Dawn Walker
University of Toronto, CA
- Blay R. Whitby
University of Sussex –
Brighton, GB
- Jon Whittle
Monash University –
Clayton, AU
- Emily Winter
Lancaster University, GB



Mobile Data Visualization

Edited by

Eun Kyoung Choe¹, Raimund Dachzelt², Petra Isenberg³, and
Bongshin Lee⁴

1 University of Maryland – College Park, US, choe@umd.edu

2 Technische Universität Dresden, DE, raimund.dachzelt@tu-dresden.de

3 Inria Saclay – Orsay, FR, petra.isenberg@inria.fr

4 Microsoft Research – Redmond, US, bongshin@microsoft.com

Abstract

Mobile visualization is becoming more prevalent, and new mobile device form factors and hardware capabilities will continually emerge in the coming years. Therefore, it is timely to reflect on what has been discovered to date and to look into the future. This Dagstuhl seminar brought together both established and junior researchers, designers, and practitioners from relevant application and research fields, including visualization, ubiquitous computing, human-computer interaction, and health informatics. Five demos and five tutorials gave participants an opportunity to share their experiences and research, and learn skills relevant to mobile data visualization. Through brainstorming and discussion in break-out sessions, along with short report back presentations, participants identified challenges and opportunities for future research on mobile data visualization.

Seminar July 14–19, 2019 – <http://www.dagstuhl.de/19292>

2012 ACM Subject Classification Human-centered computing → Human computer interaction (HCI), Human-centered computing → Ubiquitous and mobile computing, Human-centered computing → Visualization, Human-centered computing → Interaction design

Keywords and phrases Data visualization, Human-computer interaction, Information visualization, Mobile computing, Ubiquitous computing

Digital Object Identifier 10.4230/DagRep.9.7.78

Edited in cooperation with Ricardo Langner and Tom Horak


1 Executive Summary

Eun Kyoung Choe (University of Maryland – College Park, US, choe@umd.edu)

Raimund Dachzelt (Technische Universität Dresden, DE, raimund.dachzelt@tu-dresden.de)

Petra Isenberg (Inria Saclay – Orsay, FR, petra.isenberg@inria.fr)

Bongshin Lee (Microsoft Research – Redmond, US, bongshin@microsoft.com)

License  Creative Commons BY 3.0 Unported license

© Eun Kyoung Choe, Petra Isenberg, Raimund Dachzelt, and Bongshin Lee

As pen- and/or touch-enabled mobile devices have become more powerful and ubiquitous, we see a growing demand for *mobile data visualization* to facilitate visual access to data on mobile devices (see Figure 1 for examples). Lay people increasingly access a wide range of data, including weather, finance, and personal health on their phone. Small business owners start to use business intelligence software equipped with data visualization on mobile devices to make better business decisions. In responding to these needs, practitioners have actively been designing mobile visualizations embedded in commercial systems. However,



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Mobile Data Visualization, *Dagstuhl Reports*, Vol. 9, Issue 7, pp. 78–93

Editors: Eun Kyoung Choe, Raimund Dachzelt, Petra Isenberg, and Bongshin Lee



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



Figure 1 Examples of mobile data visualizations: step count and sleep data visualization on Fitbit Ionic and mobile app (top left); a multiple coordinated views across two mobile devices in VisTiles (bottom left); and visual data exploration on a tablet leveraging pen and touch interaction in TouchPivot (right).

research communities, such as Human-Computer Interaction (HCI), Information Visualization (InfoVis), and Ubiquitous Computing (UbiComp) have not paid enough attention to mobile data visualization.

Over the past few decades, the visualization research community has conducted extensive research, designing and developing a large number of visualization techniques and systems mostly for a desktop environment. However, the accumulated knowledge may not be readily transferable to mobile devices due to their fundamental differences in their display size, interaction, and target audience, among others. The small display on mobile devices is more vulnerable to the scalability issue and poses a well-known challenge, the fat finger problem. Mouse-over interaction, which is prevalent in interactive visualization systems in the desktop environment, is not available on mobile devices. While traditional visualizations mainly target data-savvy groups of people such as scientists and researchers, visualizations on mobile devices should account for a broader range of target audience, including lay people who might have low data and visualization literacy.

This Dagstuhl seminar follows in the footsteps of the “Data Visualization on Mobile Devices” workshop at CHI 2018, our initial effort in establishing a community around mobile data visualization. We brought researchers and practitioners from relevant application and research fields, including InfoVis, UbiComp, mobile HCI, and interaction design to exchange information and experiences, to stimulate discussion, to make new connections, and to identify novel ideas and future directions around mobile data visualization.

Unlike the CHI workshop, this five-day Dagstuhl workshop enabled us to explore mobile data visualization in depth through speedy & intense research exchanges, interactive demos & tutorials, as well as active breakout group discussions.



■ **Figure 2** Exchange of research interest & background in a speed dating format.

The Week at a Glance

Monday. The seminar was kicked off by the organizers with an introduction to the topic of mobile data visualization and by providing organizational information. Afterwards, all participants introduced themselves and their expectations with a short two-minute slide presentation. This session was followed by a speedy research brainstorming activity (see Figure 2): In rapid five-minute sessions, two participants facing each other introduced their research activities and jointly sketched new ideas. By rotating half of the group, each session was repeated eleven times with new constellations of two people each time.

In the afternoon, five demo stations were set up and participants were split into groups to attend them in turn. Five researchers presented their latest mobile visualization demos in hands-on sessions (see Figure 3). These were:

- Tanja Blascheck: Smartwatch demo from a study comparing three representations—bar, donut, text (joint work with Lonni Besançon, Anastasia Bezerianos, Bongshin Lee, Petra Isenberg).
- Matthew Bremer: Tilting, brushing, & dialing for mobile vis (joint work with Bongshin Lee, Christopher Collins, Ken Hinckley).
- Tobias Isenberg: Personal home automation system with mobile data access and control.
- Alark Joshi: Visualization of off-screen data using summarization techniques (joint work with Martino Kuan, Alejandro Garcia, Sophie Engle).
- Jo Vermeulen: Product Fingerprints, a mobile visualization that allows people to compare nutritional information between food products (joint work with Carrie Mah, Kevin Ta, Samuel Huron, Richard Pusch, Jo Vermeulen, Lora Oehlberg, Sheelagh Carpendale).



■ **Figure 3** One of the mobile visualization demos presented to a small group of participants.

In a second activity, 14 participants presented a design critique of an existing mobile visualization, partly commercial products, partly research results (see Figure 4). Besides evoking the spirit of a good discussion, it helped getting a broad overview about currently available solutions.

In a followup activity, to arrive at a common understanding of the state of the art in mobile data visualization, we split attendees into three groups according to their main expertise. The three groups were:

- Information Visualization–Mobile Visualization Resources
- Visualization in Ubiquitous Computing Research
- Mobile Interaction and Human Computer Interaction

Each group was tasked to collect and discuss the state of the art, with an end goal of creating a short presentation to be given to the entire audience. As a result, the collected material and insights were presented to the plenum by each group.

Through these diverse activities during the first day, participants did not only gain a good understanding of each other's background and research interests, but also established a common ground and expertise in the field of mobile data visualization

Tuesday. The second day started with a lively brainstorming and discussion of challenges and important research questions in the field of mobile data visualization. From about ten larger topics we identified, four were chosen to form parallel breakout groups:

- Group 1: Evaluating Mobile Data Visualization
- Group 2: What is Mobile Vis?
- Group 3: Responsive Visualization
- Group 4: Vis for Good & Ethics



■ **Figure 4** Impressions from the Design Critique Session.

Using the impressive facilities of Dagstuhl in terms of rooms and places, space to think and coffee to drink, we had intense discussions within each group. We generated deeper research questions and challenges, and identified collaborative cross-disciplinary research opportunities and approaches. Section 4 provides more details on each of these working groups.

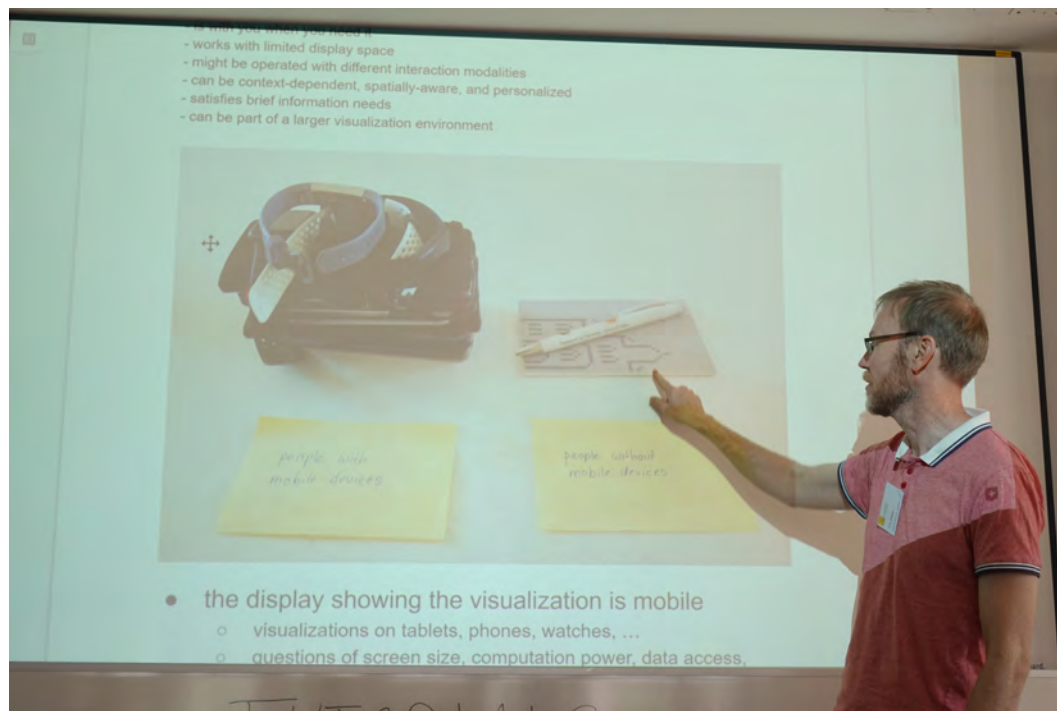
After lunch, groups reported back on what they had discussed (see Figure 5). The four groups decided to continue and deepen their discussions in the afternoon, this time focusing more on what could become a concrete research output.

Wednesday. Wednesday morning was devoted to the presentation of tutorials. Five participants had volunteered to give tutorials in two time slots, allowing other participants to attend two one-hour tutorials. Figure 6 shows the title slides of all informative and well-received tutorials, and Section 3 provides details on each of them.

Following the tradition of Dagstuhl Seminar, Wednesday afternoon was set aside for social activities. We took the bus to experience the famous Saarschleife high from the impressive treetop walk. Visiting Mettlach and having dinner in a brewery intensified personal conversations and fostered planning for joint research collaborations.

Thursday. Similar to Tuesday, the entire day was dedicated to group work (see Figure 7). The list of possible topics for breakout groups was revisited, and people assembled to form new groups on other challenging topics:

- Group 5: Starting Mobile Visualization from Scratch
- Group 6: Beyond Watch & Phone: From Mobile to Ubiquitous Visualization
- Group 7: (Discoverable) Interaction for Mobile Visualization
- Group 8: From Perception to Behavior Change: Designing and Evaluating Glanceable Mobile Vis
- Group 9: Mobile Vis for 3D Data / AR Vis



■ **Figure 5** Report back from Group 2 on "What is Mobile Vis?"

Again, both the morning and afternoon were used for intensely discussing challenges, defining design spaces, shaping the knowledge on the given topic, and identifying opportunities for joint research. Groups also reported back to the plenum, and results were discussed openly. Section 4 provides more details on each of these working groups and their outcomes.

Friday. After interesting and enriching days of joint discussions, which considerably broadened the horizon, time had come to wrap up the seminar on Mobile Data Visualization. Most importantly, a broad range of future collaborative activities were discussed: writing a state-of-the-art report, joint grant proposals, further workshop proposals, individual papers, editing a special journal issue, and writing a book on the topic. In the end, we agreed on a book as a possible major outcome (see Section 5). Organizational details were clarified, before the seminar was concluded with thanking all participants for their great contributions and commitment during the entire week.

2 Table of Contents

Executive Summary

Eun Kyoung Choe, Petra Isenberg, Raimund Dachzelt, and Bongshin Lee 78

Tutorials

Getting Started with Web-based Visualizations
Dominikus Baur 86

Designing Mobile Visualizations for Mass-Market Users
Frank Bentley 87

Crowdsourced Evaluation for Mobile Vis
Matthew Brehmer 87

The Immersive Analytics Toolkit – IATK
Tim Dwyer 87

Microcontroller Programming for Sensor Data Capture & Visualization
Tobias Isenberg 88

Working Groups

What is Mobile Vis?
Tim Dwyer, Lonni Besancon, Christopher Collins, Petra Isenberg, Tobias Isenberg, Ricardo Langner, Bongshin Lee, Charles Perin, Harald Reiterer, and Christian Tominski 88

Evaluating Mobile Data Visualization
Lena Mamykina, Frank Bentley, Eun Kyoung Choe, Pourang P. Irani, and John T. Stasko 89

Responsive Visualization
Wolfgang Aigner, Dominikus Baur, Matthew Brehmer, Tom Horak, Alark Joshi, Harald Reiterer, and Christian Tominski 89

Vis for Good & Ethics
Jo Vermeulen, Tanja Blascheck, Sheelagh Carpendale, Raimund Dachzelt, and Daniel Epstein 89

Starting Mobile Visualization from Scratch
Dominikus Baur, Sheelagh Carpendale, Daniel Epstein, Lena Mamykina, and Charles Perin 90

Beyond Watch/Phone: From Mobile to Ubiquitous Visualization
Christopher Collins, Raimund Dachzelt, Pourang P. Irani, Alark Joshi, Ricardo Langner, and Jo Vermeulen 90

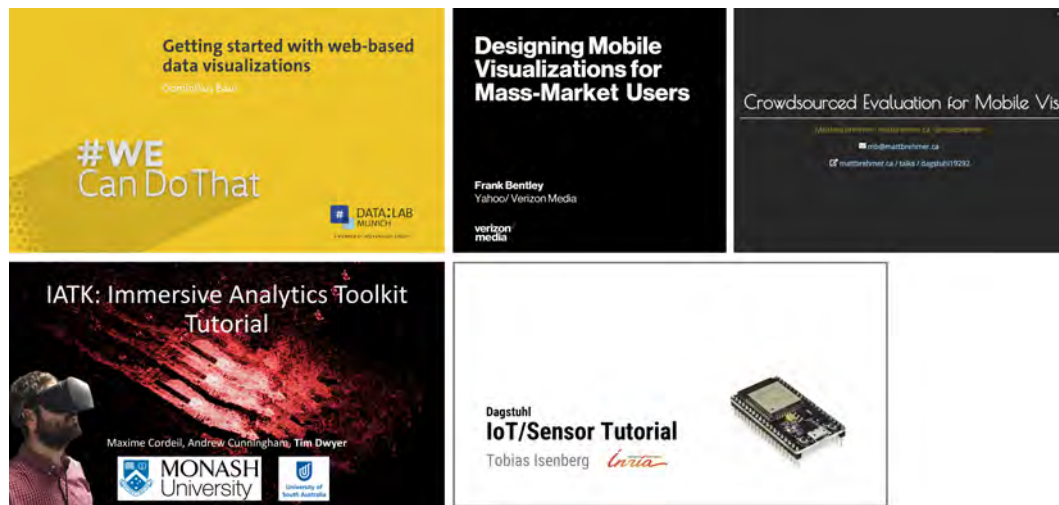
(Discoverable) Interaction for Mobile Visualization
Matthew Brehmer, Bongshin Lee, John T. Stasko, and Christian Tominski 91

From Perception to Behavior Change: Designing and Evaluating Glanceable Mobile Vis
Tanja Blascheck, Frank Bentley, Eun Kyoung Choe, Tom Horak, and Petra Isenberg 91

Mobile Vis for 3D Data / AR Vis
Tim Dwyer, Wolfgang Aigner, Lonni Besancon, Tobias Isenberg, and Harald Reiterer 92

Outlook and Conclusion 92

Participants 93



■ **Figure 6** Title slides of all five tutorials presented at the seminar.

3 Tutorials

Before the Dagstuhl seminar, we solicited volunteers to give tutorials and demos at the seminar. Five people gave tutorials to share their research and experiences relevant to mobile data visualization. In the following, we provide the abstracts of these tutorials.

3.1 Getting Started with Web-based Visualizations

Dominikus Baur (Volkswagen Data:Lab – München, DE)

License © Creative Commons BY 3.0 Unported license
© Dominikus Baur

In this hands-on tutorial, we learned about the basics of creating web-based data visualizations. Initially, we looked at the basic web technologies of HTML, CSS, and JavaScript, their basic syntax, and what they're used for. We played with the technologies in a browser-based development environment to receive instant feedback for our experiments. Next, we dove into d3.js, the JavaScript library commonly used for creating visualizations. We looked into the most important functions that d3.js provides to create simple visualizations. We also learned about the more idiosyncratic approaches that d3.js encompasses for mapping data to visual elements. Finally, we got into more of the real-world aspect of web development with an overview of the Node Package Manager (npm) and how bundlers like Parcel or rendering frameworks like React work.

3.2 Designing Mobile Visualizations for Mass-Market Users

Frank Bentley (Yahoo Labs – Sunnyvale, US)

License © Creative Commons BY 3.0 Unported license
© Frank Bentley

When designing visualizations, it's important to consider the broader population, and their ability to interpret what is shown. Properly recruiting users that match the broader population is extremely important. In this regard, getting 2/3 of participants without a college degree, diverse ages, incomes, and races while achieving a gender balance will allow you to understand how your interface would be perceived by a larger audience. Yet many visualizations are not evaluated this way. In addition, it has been shown that many users have trouble interpreting standard InfoVis techniques, such as time series graphs or maps. Alternatives to these visualizations, using text explanations, summaries, or other ways to simplify the data are critical if systems are to be adopted and understood by broader audiences. This talk highlighted some alternatives that have been tried over the past two decades in HCI and Ubicomp research.

3.3 Crowdsourced Evaluation for Mobile Vis

Matthew Brehmer (Vancouver, CA)

License © Creative Commons BY 3.0 Unported license
© Matthew Brehmer

In this tutorial, we reviewed the practical and methodological aspects of conducting crowdsourced experiments about visualization on mobile devices. This tutorial followed two recently conducted experiments of this sort (in collaboration with Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe). I described considerations for designing and developing mobile-only web apps for visualization experiments, as well as considerations for recruiting, piloting, and onboarding participants. I also described several ways to improve participant compliance and response quality. Finally, I pointed to other resources for crowdsourcing and mobile visualization design, and suggested some opportunities for future experimental work.

3.4 The Immersive Analytics Toolkit – IATK

Tim Dwyer (Monash University – Caulfield, AU)

License © Creative Commons BY 3.0 Unported license
© Tim Dwyer

Main reference Maxime Cordeil, Andrew Cunningham, Benjamin Bach, Christophe Hurter, Bruce H. Thomas, Kim Marriott, Tim Dwyer: "IATK: An Immersive Analytics Toolkit", in Proc. of the IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019, Osaka, Japan, March 23-27, 2019, pp. 200–209, IEEE, 2019.

URL <http://dx.doi.org/10.1109/VR.2019.8797978>

Immersive Analytics Toolkit (IATK) is a Unity project to help you build high quality, interactive and scalable data visualisations in Immersive Environments (Virtual/Augmented Reality). This tutorial allowed participants to learn how to use the Visualisation script to create data visualisations interactively in the editor, press play and view and interact with data in V/AR. Participants could also write simple code to use the IATK core graphics components to make own interactive visualisations programmatically.

3.5 Microcontroller Programming for Sensor Data Capture & Visualization

Tobias Isenberg (INRIA Saclay – Orsay, FR)

License  Creative Commons BY 3.0 Unported license
© Tobias Isenberg


The tutorial gave an overview of the ESP8266 and ESP32 microcontrollers and how to prototype sensor reading and data visualization with them. I showed the use of the Arduino IDE and talked about relevant electronics issues. I also covered how to manage battery operation of the microcontrollers. Finally, I covered how to collect the data using MQTT and how to prototype visualizations on e-ink displays.

4 Working Groups

The main seminar was dedicated to working in several breakout groups. From this activity, several research questions and challenging topics were identified. We first identified topics of interest by asking people to vote for the topics that they would like to discuss. All participants were part of at least two working groups, spending a day (Tuesday & Thursday) for each topic. In the following, we provide abstracts for all nine breakout groups.

4.1 What is Mobile Vis?

Tim Dwyer (Monash University – Caulfield, AU), Lonni Besancon (Linköping University, SE), Christopher Collins (Ontario Tech – Oshawa, CA), Petra Isenberg (INRIA Saclay – Orsay, FR), Tobias Isenberg (INRIA Saclay – Orsay, FR), Ricardo Langner (TU Dresden, DE), Bongshin Lee (Microsoft Research – Redmond, US), Charles Perin (University of Victoria, CA), Harald Reiterer (Universität Konstanz, DE), and Christian Tominski (Universität Rostock, DE)

License  Creative Commons BY 3.0 Unported license
© Tim Dwyer, Lonni Besancon, Christopher Collins, Petra Isenberg, Tobias Isenberg, Ricardo Langner, Bongshin Lee, Charles Perin, Harald Reiterer, and Christian Tominski

There are several ways in which the term “Mobile Data Vis” may be interpreted. For example, it may describe: visualizations hosted on devices that are mobile; situations where the users of visualizations are mobile relative to the display; and visualizations that are themselves mobile across devices and screens. We focused mainly on defining visualization for mobile devices, and left deeper consideration of the latter two interpretations to future meetings. We explored the characteristics upon which visualizations can be described, focusing on those which, in their extremes, differentiate mobile visualization from other forms of data visualization. These characteristics gave rise to a first set of dimensions of a design space for mobile data visualization, in which instances of mobile data visualization may be positioned. We discussed a number of such example instances to illustrate how the definition makes it possible to describe and compare mobile visualizations. Using the dimensions, we identified gaps and opportunities for future mobile visualizations.

4.2 Evaluating Mobile Data Visualization

Lena Mamykina (Columbia University – New York, US), Frank Bentley (Yahoo Labs – Sunnyvale, US), Eun Kyoung Choe (University of Maryland – College Park, US), Pourang P. Irani (University of Manitoba – Winnipeg, CA), and John T. Stasko (Georgia Institute of Technology – Atlanta, US)

License © Creative Commons BY 3.0 Unported license

© Lena Mamykina, Frank Bentley, Eun Kyoung Choe, Pourang P. Irani, and John T. Stasko

There are many different reasons to evaluate mobile visualizations with end users. Depending on the intention of the system, different methods are needed. In the mobile information visualization domain, there is a broad continuum of research questions that can be answered by a study. Some goals include validating rapid perception of differences in data, while others are interested in examining long-term use of visualizations and whether they achieve their intended impact on users. Very different methods, time-scales of research, and user recruitment strategies are needed. We began to explore the literature and different system goals and evaluation approaches, and plan to continue our work by highlighting best practices and making recommendations for future approaches to evaluating mobile data visualizations.

4.3 Responsive Visualization

Wolfgang Aigner (FH St. Pölten, AT), Dominikus Baur (Volkswagen Data:Lab – München, DE), Matthew Brehmer (Vancouver, CA), Tom Horak (TU Dresden, DE), Alark Joshi (University of San Francisco, US), Harald Reiterer (Universität Konstanz, DE), and Christian Tominski (Universität Rostock, DE)

License © Creative Commons BY 3.0 Unported license

© Wolfgang Aigner, Dominikus Baur, Matthew Brehmer, Tom Horak, Alark Joshi, Harald Reiterer, and Christian Tominski

Due to the proliferation of mobile devices like smartphones and tablets, an increasing number of data visualizations are being used not only on desktop computers but also on mobile devices. But, visualizations designed for desktop computers are often unusable on smaller mobile devices due to differences and restrictions in display size, aspect ratio, and interaction capabilities. Therefore, mobile data visualization applications need to be responsive to the specific constraints of the devices used as well as their users, environment, data, and usage contexts. In our breakout group, we discussed causes of responsiveness, how the contextual information can be sensed on devices, and what needs to be adapted based on this information. Furthermore, we worked towards a conceptual model of responsiveness by extending the simple visualization model of Van Wijk in order to capture all design aspects and data aspects.

4.4 Vis for Good & Ethics

Jo Vermeulen (Aarhus University, DK), Tanja Blascheck (Universität Stuttgart, DE), Sheelagh Carpendale (Simon Fraser University – Burnaby, CA), Raimund Dachzelt (TU Dresden, DE), and Daniel Epstein (University of California – Irvine, US)

License © Creative Commons BY 3.0 Unported license
© Jo Vermeulen, Tanja Blascheck, Sheelagh Carpendale, Raimund Dachzelt, and Daniel Epstein

Our group discussed that visualization is neither good, bad, nor neutral. Visualization is not necessarily objective. We focused on visualization for good and for bad. Due to limited screen space, more interruptions, people are more at the mercy of the visualization designer. We attempted to characterize existing “dark patterns” for mobile visualization.

4.5 Starting Mobile Visualization from Scratch

Dominikus Baur (Volkswagen Data:Lab – München, DE), Sheelagh Carpendale (Simon Fraser University – Burnaby, CA), Daniel Epstein (University of California – Irvine, US), Lena Mamykina (Columbia University – New York, US), and Charles Perin (University of Victoria, CA)

License © Creative Commons BY 3.0 Unported license
© Dominikus Baur, Sheelagh Carpendale, Daniel Epstein, Lena Mamykina, and Charles Perin

Our group discussed what mobile visualization could be like if we shed the restrictions of existing technologies and the influence of existing (desktop) visualizations. We started from several scenarios (e.g., a lecture situation, supporting people’s nutritional choices, team sports) and asked ourselves how support by a visualization system could work there. We categorized the primacy of the visualization task along a spectrum from passive awareness (via supporting the main task) to in-depth analysis and discussed corresponding considerations regarding information displays and timeliness. We also discussed “progressive visualizations” that would increase the information density depending on the amount of available attention and the viewer’s involvement. As a result, we argued the inversion of Ben Shneiderman’s interaction mantra to *details-first, triggering interest, and analysis/overview-on-demand*.

4.6 Beyond Watch/Phone: From Mobile to Ubiquitous Visualization

Christopher Collins (Ontario Tech – Oshawa, CA), Raimund Dachzelt (TU Dresden, DE), Pourang P. Irani (University of Manitoba – Winnipeg, CA), Alark Joshi (University of San Francisco, US), Ricardo Langner (TU Dresden, DE), and Jo Vermeulen (Aarhus University, DK)

License © Creative Commons BY 3.0 Unported license
© Christopher Collins, Raimund Dachzelt, Pourang P. Irani, Alark Joshi, Ricardo Langner, and Jo Vermeulen

Our group talked about visualization beyond the mobile phone and smartwatch. We discussed other approaches including networked small situated displays, cheap disposable (flexible) displays, and textiles. The key characteristic is that these envisioned solutions support people’s ongoing activities. We discussed several possible scenarios including crisis management, large scale communication to the public, and communication between cars,

as well as several existing examples in the literature. Finally, we identified core dimensions such as personal vs. display movement; data that is mobile; public vs. private visualization; information needs; urgency; and situational context.

4.7 (Discoverable) Interaction for Mobile Visualization

Matthew Brehmer (Vancouver, CA), Bongshin Lee (Microsoft Research – Redmond, US), John T. Stasko (Georgia Institute of Technology – Atlanta, US), and Christian Tominski (Universität Rostock, DE)

License © Creative Commons BY 3.0 Unported license
© Matthew Brehmer, Bongshin Lee, John T. Stasko, and Christian Tominski

Our group discussed the challenges and difficulties of interacting with mobile devices, of interacting with visualization on mobile devices, and of interaction in casual contexts. Next, we considered ways of structuring the space of interaction for visualization on mobile devices, such as around existing visualization task typologies, existing visualization interaction typologies, interaction modalities, data types, and chart types. Regardless of how we structure the space of interaction, we will catalog current approaches, gaps, and future opportunities. Finally, we distinguished mobile interaction from desktop/laptop interaction.

4.8 From Perception to Behavior Change: Designing and Evaluating Glanceable Mobile Vis

Tanja Blaschek (Universität Stuttgart, DE), Frank Bentley (Yahoo Labs – Sunnyvale, US), Eun Kyong Choe (University of Maryland – College Park, US), Tom Horak (TU Dresden, DE), and Petra Isenberg (INRIA Saclay – Orsay, FR)

License © Creative Commons BY 3.0 Unported license
© Tanja Blaschek, Frank Bentley, Eun Kyong Choe, Tom Horak, and Petra Isenberg

There is a continuum of uses for mobile visualizations, from solving quick information needs, through systems that provide browsing of data in more detail, to systems that afford deep analysis of larger datasets. This working group focused on systems that solve quick information needs.

Quick information needs are important components of mobile visualizations within applications such as fitness trackers, GPS displays in a car, tracking family members, or weather awareness. Visualizations that require passive interactions and are designed for quick information needs are described under a variety of terms such as glanceable visualizations, glanceable displays, peripheral displays, ambient visualizations, notification systems, or casual visualizations. In this working group, we discussed these individual terms and how they are related across the ubiquitous computing and visualization domains with a focus on how the term “glanceable” differs in the communities.

In addition, we discussed purposes for glanceable displays—from quick awareness, such as indicating to a driver or pilot that there is a serious problem, to systems meant to evoke long-term behavior change, which will be glanced at thousands of times. This working group explored different visualization scenarios, characteristics, and evaluation methodologies for these different purposes.



■ **Figure 7** Group 8 discussing glanceable mobile visualization in the Dagstuhl garden.

4.9 Mobile Vis for 3D Data / AR Vis

Tim Dwyer (Monash University – Caulfield, AU), Wolfgang Aigner (FH St. Pölten, AT), Lonni Besancon (Linköping University, SE), Tobias Isenberg (INRIA Saclay – Orsay, FR), and Harald Reiterer (Universität Konstanz, DE)

License © Creative Commons BY 3.0 Unported license
© Tim Dwyer, Wolfgang Aigner, Lonni Besancon, Tobias Isenberg, and Harald Reiterer

We surveyed the space of 3D mobile visualizations (3D data on mobile 2D displays, abstract and/or 3D data in mobile (HMD) AR/VR displays). As a playful “Case Study” we used a scenario from the film “Aliens”, in which a mobile, small-screen visualisation device is used to track the movements of enemy aliens around a group of space marines. In this scenario, the marines are overrun by aliens in the ceiling, as their device fails to show them the height dimension of the space around them. We used this example to illustrate how different mobile and 3D interaction techniques could have prevented the misunderstanding in the movie, using both hypothetical descriptions of the improved movie action and a scientific discussion of these scenarios and their implications.

5 Outlook and Conclusion

As an outcome of the seminar, we are working towards a joint publication that captures many of the discussed topic related to mobile data visualization. Each chapter will expand on the discussions started at Dagstuhl and will include in-depth explorations of the most of the working group topics mentioned in Section 4. We hope that our book will engage the community to further pursue this exciting topic.

In summary, we had a fruitful and engaged seminar and received positive feedback from the group. The organizers thank Dagstuhl for hosting our seminar and the great research facilities provided.

Participants

- Wolfgang Aigner
FH St. Pölten, AT
- Dominikus Baur
Volkswagen Data:Lab –
München, DE
- Frank Bentley
Yahoo Labs – Sunnyvale, US
- Lonni Besancon
Linköping University, SE
- Tanja Blascheck
Universität Stuttgart, DE
- Matthew Brehmer
Vancouver, CA
- Sheelagh Carpendale
Simon Fraser University –
Burnaby, CA
- Eun Kyong Choe
University of Maryland –
College Park, US
- Christopher Collins
Ontario Tech – Oshawa, CA
- Raimund Dachzelt
TU Dresden, DE
- Tim Dwyer
Monash University –
Caulfield, AU
- Daniel Epstein
University of California –
Irvine, US
- Tom Horak
TU Dresden, DE
- Pourang P. Irani
University of Manitoba –
Winnipeg, CA
- Petra Isenberg
INRIA Saclay – Orsay, FR
- Tobias Isenberg
INRIA Saclay – Orsay, FR
- Alark Joshi
University of San Francisco, US
- Ricardo Langner
TU Dresden, DE
- Bongshin Lee
Microsoft Research –
Redmond, US
- Lena Mamykina
Columbia University –
New York, US
- Charles Perin
University of Victoria, CA
- Harald Reiterer
Universität Konstanz, DE
- John T. Stasko
Georgia Institute of Technology –
Atlanta, US
- Christian Tominski
Universität Rostock, DE
- Jo Vermeulen
Aarhus University, DK



Secure Composition for Hardware Systems

Edited by

Divya Arora¹, Ilia Polian², Francesco Regazzoni³, and
Patrick Schaumont⁴

- 1 Intel – Santa Clara, US, divya.arora@intel.com
- 2 Universität Stuttgart, DE, ilia.polian@informatik.uni-stuttgart.de
- 3 University of Lugano, CH, regazzoni@alari.ch
- 4 Virginia Polytechnic Institute – Blacksburg, US, schaum@vt.edu

Abstract

The goal of the Dagstuhl Seminar 19301 “Secure Composition for Hardware Systems” was to establish a common understanding of principles and techniques that can facilitate composition and integration of hardware systems to achieve specified security guarantees.

Theoretical foundations of secure composition have been laid out in the past, but they are limited to software systems. New and unique security challenges arise when a real system composed of a range of hardware components, including application-specific blocks, programmable microcontrollers, and reconfigurable fabrics, are put together. For example, these components may have different owners, different trust assumptions and may not even have a common language to describe their security properties to each other. Physical and side-channel attacks that take advantage of various physical properties to undermine a system’s security objectives add another level of complexity to the secure composition problem. Moreover, practical hardware systems include software of tremendous size and complexity, and hardware-software interaction can create new security challenges.

The seminar considered secure composition both from a pure hardware perspective, where multiple hardware blocks are composed in, *e.g.*, a system on chip (SoC), and from a hardware-software perspective where hardware is integrated within a system that includes software. The seminar brought together researchers and industry practitioners from fields that have to deal with secure composition: Secure hardware architectures, hardware-oriented security, applied cryptography, test and verification of security properties. By involving industrial participants, we were able to get insights on real-world challenges, heuristics, and methodologies employed to address them and initiate a discussion towards new solutions.

Seminar July 21–26, 2019 – <http://www.dagstuhl.de/19301>

2012 ACM Subject Classification Hardware → Methodologies for EDA, Hardware → Integrated circuits, Security and privacy → Formal security models

Keywords and phrases Hardware, Secure composition, Security, Software

Digital Object Identifier 10.4230/DagRep.9.7.94

Edited in cooperation with Elif Bilge Kavun, The University of Sheffield, UK,
e.kavun@sheffield.ac.uk



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Secure Composition for Hardware Systems, *Dagstuhl Reports*, Vol. 9, Issue 7, pp. 94–116

Editors: Divya Arora, Ilia Polian, Francesco Regazzoni, and Patrick Schaumont



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Ilia Polian (Universität Stuttgart, DE)

Divya Arora (Intel – Santa Clara, US)

Francesco Regazzoni (University of Lugano, CH)

Patrick Schaumont (Virginia Polytechnic Institute – Blacksburg, US)

License © Creative Commons BY 3.0 Unported license

© Ilia Polian, Divya Arora, Francesco Regazzoni, and Patrick Schaumont

Today’s electronic systems consist of mixtures of programmable, reconfigurable and application-specific hardware components, tied together by tremendously complex software. At the same time, systems are increasingly integrated such that a system that was traditionally regarded “harmless” (*e.g.*, an entertainment system in a car) finds itself tightly coupled with safety-critical driving-assistance systems and security-sensitive online payment systems. Moreover, a system’s hardware components are now often directly accessible to the users, making the system vulnerable to physical attacks via its hardware which becomes the system’s “Achille’s heel”. This necessitates a new look on system security from hardware perspective.

The Dagstuhl seminar “Secure Composition for Hardware Systems”, which took place on July 21-26, 2019, focused on secure composition of systems which contain hardware blocks. This is a practically important but a theoretically challenging problem where several foundational questions still lack an adequate answer.

Several formats were used during the seminar. The first phase of the seminar, which focused on prior findings, started with presentations by five pre-selected experts giving their view on secure composition from different perspectives: theory, design automation, trusted execution environments and attacks countermeasures. Then, small-group discussions of relevant state of the art were held, focusing on questions such as “What does it mean to securely compose two elements?” or “What is the role of models in secure composition?” The findings of the small groups were intensively discussed in plenary sessions.

The second phase of the seminar was devoted to discussing research questions. Some of the questions were prepared by the seminar organizers (*e.g.*, “Which protocol-level secure composition methods are applicable in hardware domain?” or “How to counter possible loss of security due to abstraction of hardware components?”) and some additional questions were proposed by the participants (*e.g.*, “How to bootstrap trust in a distributed hardware system?”). The questions were discussed again in small groups, intertwined by individual presentations in plenum (for instance, an in-depth study on the applicability of Universal Composability (UC) in the hardware domain).

Two immediate outcomes grew out of the seminar. First, some participants are organizing a special session on secure compositions in one of the leading scientific conferences; a respective proposal was recently accepted by the “Design, Automation, and Test in Europe Conference” (DATE). Second, there is an ambitious plan to prepare a manuscript on the full variety of aspects in secure composition of electronic systems and submit it as a “Systematization of Knowledge” (SoK) paper to the IEEE Symposium on Security and Privacy (S&P); this effort is ongoing at the time of writing this report.

Overall, we believe that this seminar has provided entirely new insights to most of the participants and has opened new avenues for research on the intersection of security and hardware systems. It brought together researchers from communities who rarely interacted with each other in the past. The seminar helped define new research challenges, and activities are underway to put the topic of secure composition higher on the agenda of the respective communities.

The organizers are thankful to the Dagstuhl team (and in particular to Dr. Andreas Dolzmann who handled the scientific part and Mrs. Heike Clemens who was of invaluable help in organizing the social event and masterly handled all practical issues); to Dr. Elif Bilge Kavun who did a great job in collecting and organizing the documents from participants and in preparing the summarizing texts; and to all the participants for making this seminar a success.

2 Table of Contents

Executive Summary

Ilia Polian, Divya Arora, Francesco Regazzoni, and Patrick Schaumont 95

Overview of Pre-selected Talks

Commercial Trusted Execution Environments (TEEs) – An Overview
Divya Arora 99

Introduction to Universal Composability
Ran Canetti 99

State-of-the-art Implementation Attacks
Elke De Mulder 100

Cryptographic Hardware Design – Challenges and Remarks
Tim Erhan Güneysu 100

State-of-the-art in EDA Security
Francesco Regazzoni 101

Overview of Individual Presentations

Composability of Machine-Learning Resistant PUFs – When Yao Fails -or- Can we build secure composite PUFs?
Fatemeh Ganji 101

Attacks Through Externally-amplified Couplings
Itamar Levi 102

CAD for Physical Attacks – A Fault Attack Perspective
Debdeep Mukhopadhyay 103

Securing Cyber-physical Control Systems – A Formal Perspective
Dey Soumyajit 103

Challenges in Secure Composition from a Practical Perspective
Marc Stöttinger 104

Definition of “Root of Trust (RoT)”
Ingrid Verbauwhede 104

Discussions on State-of-the-art Questions

Working Group A1
Divya Arora, Gaetan Cassiers, Johann Heyszl, Itamar Levi, Debdeep Mukhopadhyay, Kazuo Sakiyama, and Dey Soumyajit 105

Working Group A2
Georg T. Becker, Yaacov Belenky, Shivam Bhasin, and Shahin Tajik 106

Working Group A3
Lucas Davi, Fatemeh Ganji, Tim Erhan Güneysu, Ahmad-Reza Sadeghi, and Fareena Saqib 107

Working Group B1
Ran Canetti, Jean-Luc Danger, Elif Bilge Kavun, Osnat Keren, Johannes Mittmann, and Ilia Polian 107

Working Group B2	
<i>Elke De Mulder, Elena Dubrova, Yunsu Fei, Paolo Palmieri, and Milos Prvulovic</i>	108
Working Group B3	
<i>Annelie Heuser, Michail Maniatakos, Wenjing Rao, Patrick Schaumont, Werner Schindler, and Marc Stöttinger</i>	110
Discussions on Research Questions	
Research Questions:	
“What models and description languages are useful for formalization of security properties?”	
“Which protocol-level secure composition methods are applicable in hardware domain?”	
<i>Patrick Schaumont, Shivam Bhasin, Debdeep Mukhopadhyay, Francesco Regazzoni, Kazuo Sakiyama, Dey Soumyajit, and Ingrid Verbauwhede</i>	112
Research Questions:	
“Can trust start in software, or are hardware roots and anchors of trust indispensable?”	
“How to bootstrap trust in a distributed hardware system?”	
<i>Johann Heyszl, Ran Canetti, Fatemeh Ganji, Michail Maniatakos, Marcel Medwed, Shahin Tajik, and Marten Van Dijk</i>	112
Research Question:	
“Under what circumstances is security additive, and how can this be proven and validated?”	
<i>Jean-Luc Danger, Yaacov Belenky, Elke De Mulder, Elena Dubrova, Osnat Keren, Johannes Mittmann, and Werner Schindler</i>	114
Research Questions:	
“How can existing hardware fulfill expectations and idealistic assumptions of protocols?”	
“How to counter possible loss of security due to abstraction of hardware components?”	
<i>Elif Bilge Kavun, Anupam Chattopadhyay, Annelie Heuser, Johann Knechtel, and Itamar Levi</i>	115
Participants	116

3 Overview of Pre-selected Talks

3.1 Commercial Trusted Execution Environments (TEEs) – An Overview

Divya Arora (Intel – Santa Clara, US)

License © Creative Commons BY 3.0 Unported license
© Divya Arora

A Trusted Execution Environment (TEE) is a hardware/software/firmware framework to allow isolated execution of security-sensitive code and its aim is to reduce Trusted Computing Base (TCB) of sensitive code. Isolated execution, secure storage, remote attestation, secure provisioning, and trusted input output are general properties of a TEE. Today, many commercial TEE solutions are available and different approaches exist for reducing application TCB; however, not all TEEs support all of the listed TEE properties.

This talk provided an overview of commercially available TEEs like ARM® TrustZone® (TZ), Microsoft Virtualization Based Security (VBS), AMD Secure Encrypted Virtualization (SEV), and Intel® Software Guard Extensions (SGX) and asked the question “How do we reason about security of TEEs including hardware/firmware/software?”. Also, a comparison of the supported properties in the commercial TEEs is also provided in the talk: For example, ARM®TZ and Microsoft VBS do not support remote attestation property and Intel®SGX does not support trusted input output. Finally, some examples of challenges that many TEEs face in terms of ecosystem deployment are listed:

- Not all TEEs are available to regular users (*e.g.*, selected usages deployed as part of VTL1 in VBS)
- There may still be a large software attack surface in some cases (*e.g.*, integer overflow in TZ Secure OS)
- Memory integrity & anti-replay are very hard on performance/area
- Many TEEs rely on hardware sharing to amortize the cost of creating a separate environment which may lead to side-channels
- Some TEEs require partitioning of existing applications or “enlightenment” of existing virtual machines which is harder to deploy in the ecosystem

3.2 Introduction to Universal Composability

Ran Canetti (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Ran Canetti

In this talk, a general universal composability framework for describing cryptographic protocols and analyzing their security is presented. The framework allows specifying the security requirements of practically any cryptographic task in a unified and systematic way. Furthermore, in this framework the security of protocols is preserved under a general protocol composition operation, called universal composition.

The proposed framework with its security-preserving composition operation allows for modular design and analysis of complex cryptographic protocols from simpler building blocks. Moreover, within this framework, protocols are guaranteed to maintain their security in any context, even in the presence of an unbounded number of arbitrary protocol instances that

run concurrently in an adversarially controlled manner. This is a useful guarantee, which allows arguing about the security of cryptographic protocols in complex and unpredictable environments such as modern communication networks.

3.3 State-of-the-art Implementation Attacks

Elke De Mulder (Rambus – Sunnyvale, US)


License  Creative Commons BY 3.0 Unported license
© Elke De Mulder

This talk focused on the state-of-the-art in side-channel leakage analysis and mitigation from the point-of-view of compositional security.

Three large research directions are discussed. The first is the non-completeness of our understanding of the physical leakage of devices, whether it is a pure hardware implementation or a software implementation running on an embedded processor or larger system on chip. The second one is the use of formal proofs to guarantee security properties of implementations where composability plays a role for combining smaller provable secure components to create a larger implementations and for combining countermeasures for different types of attack. Are the security properties still valid? The last research direction discussed in this talk is testing. With a growing arsenal of attacks, how can one practically test whether an implementation resist all of part of them in a reasonable amount of time?

3.4 Cryptographic Hardware Design – Challenges and Remarks

Tim Erhan Güneysu (Ruhr-Universität Bochum, DE)

License  Creative Commons BY 3.0 Unported license
© Tim Erhan Güneysu

As a result of digital evolution, today's systems consist of more software than hardware. In contrast, security demand of hardware systems does not decrease due to physical exposure, enhanced security requirements, and advanced networking and connectivity.

This talk focused on challenges in cryptographic hardware design in the presence of attacks and protection measures. Secure hardware elements have to provide security guarantees according to defined attacker model while keeping a trade-off between security, efficiency and cost. Security guarantees often (implicitly) bound to technical/physical device limitations which is important for composability.

It is possible to divide the challenges in cryptographic hardware into two groups: Crypto-level and system-level. On the crypto side, new computing paradigms (*e.g.*, quantum computers makes existing public-key cryptography obsolete) and new/changed requirements (*e.g.*, fault tolerance and verifiable & delegated computation) cause challenges. On the system side, static nature of hardware, non-trivial upgrade/migration, and security validation & testing are the main problems. Some solutions to crypto-level and system-level challenges are also presented in the talk (together with example applications): Hardware implementations of post-quantum cryptography (crypto-level) and integration of cryptographic hardware (system-level).

As a final remark, existing challenges are listed:

- Imperfections of hardware within abstract models and requirements in higher layers
- Composition and multiplicity of (low) confidence from practical security element evaluation
- Lack of a realistic simulator for complex systems
- Long term security, efficiency & cost

3.5 State-of-the-art in EDA Security

Francesco Regazzoni (University of Lugano, CH)

License  Creative Commons BY 3.0 Unported license
© Francesco Regazzoni

Security is one of the most important properties that should be provided by a system. Unfortunately, due to physical attacks, the presence of cryptographic primitives is not sufficient to fulfill this requirement.

In recent years, researchers invested significant efforts implementing optimized security primitives. These blocks are generally produced by expert designers and they are integrated manually into the whole system. This approach however is not optimal, since manual integration is a time consuming and error prone process. Furthermore, this approach is particularly dangerous when used for implementing side-channel resistant designs.


A more effective way to implement secure cryptographic algorithms would enable the automatic application of side-channel countermeasures and would support the verification of their correct application.

This talk revised and summarized the research efforts in this important research direction, starting from the first works implementing hardware design flow for security to the initial steps of automatically driving design tools using security variables and highlights future research direction in design automation for security.

4 Overview of Individual Presentations

4.1 Composability of Machine-Learning Resistant PUFs – When Yao Fails -or- Can we build secure composite PUFs?

Fatemeh Ganji (University of Florida – Gainesville, US)

License  Creative Commons BY 3.0 Unported license
© Fatemeh Ganji

When it comes to composability in the context of cryptography, Yao's lemma [1] plays an important role. This lemma states that if several instances of a somewhat-hard function are XORed together, the resulting function is harder to compute. Moreover, in cryptography and machine learning (ML) theory, it is well-known how to make a connection between the security and provable ML. Here we quote from the seminal work of Rivest, published in 1991 [2]: “In cryptography, the major goal is to ‘prove’ security under the broadest possible definition of security, [...]. [...], in the typical paradigm, it is shown that there is no polynomial-time [learning] algorithm that can ‘break’ the security of the [secure] system.” From these two principles, we can conclude that a physical primitive can become more robust against ML attack if we combine some instances of that by applying the XOR function.


In this talk, we show that this is, unfortunately, not the case. In particular, although the above approach makes physical primitives more resilient to ML attacks, it still requires drastic practical measures to be taken to achieve the ultimate level of security, where the attacker cannot learn the functionality of the respective primitive. We elaborate on this in the context of physically unclonable functions.

References

- 1 A.C. Yao *Theory and Application of Trapdoor Functions*. In 23rd Annual Symposium on Foundations of Computer Science (sfcs 1982) (pp. 80-91), IEEE, 1982, November
- 2 R.L. Rivest *Cryptography and Machine Learning*. In International Conference on the Theory and Application of Cryptology (pp. 427-439), Springer, Berlin, Heidelberg, 1991, November

4.2 Attacks Through Externally-amplified Couplings

Itamar Levi (University of Louvain, BE)

License  Creative Commons BY 3.0 Unported license
© Itamar Levi

Joint work of Itamar Levi, Davide Bellizia, François-Xavier Standaert

Main reference Itamar Levi, Davide Bellizia, François-Xavier Standaert: “Reducing a Masked Implementation’s Effective Security Order with Setup Manipulations And an Explanation Based on Externally-Amplified Couplings”, IACR Trans. Cryptogr. Hardw. Embed. Syst., Vol. 2019(2), pp. 293–317, 2019.

URL <https://doi.org/10.13154/tches.v2019.i2.293-317>

Couplings are a type of physical default that can violate the independence assumption needed for the secure implementation of the masking countermeasure. Recent works put forward qualitatively that couplings can cause information leakages of lower order than theoretically expected. However, the (quantitative) amplitude of these lower-order leakages (*e.g.*, measured as the amplitude of a detection metric such as Welch’s T statistic) was usually lower than the one of the (theoretically expected) d^{th} order leakages. So, the actual security level of these implementations remained unaffected. In addition, the couplings had to be internally amplified in order to make them visible (*e.g.*, by tweaking the placement and routing or iterating linear operations on the shares).

In this talk, firstly, how the amplitude of low-order leakages in masked implementations can be externally amplified by tweaking side-channel measurement setups in a way that they are under control of a power analysis adversary is explained. The experiments put forward that the “effective security order” of both hardware (Field Programmable Gate Array – FPGA) and software (ARM-32) implementations can be reduced, leading to concrete reductions of their security level. For this purpose, instead of the detection-based analyses of previous works, attack-based evaluations are performed in order to allow the confirmation of the exploitability of the amplified lower-order leakages. In the talk, a tentative explanation for the effects based on couplings is provided and a model that can be used to predict them in function of the measurement setup’s external resistor and implementation’s supply voltage is described. In conclusion, the effective security orders observed are mainly due to “externally-amplified couplings” that can be systematically exploited by actual adversaries.

4.3 CAD for Physical Attacks – A Fault Attack Perspective

Debdeep Mukhopadhyay (Indian Institute of Technology – Kharagpur, IN)

License © Creative Commons BY 3.0 Unported license
© Debdeep Mukhopadhyay

Joint work of Sayandeep Saha, S. Nishok Kumar, Sikhar Patranabis, Debdeep Mukhopadhyay, Pallab Dasgupta
Main reference Sayandeep Saha, S. Nishok Kumar, Sikhar Patranabis, Debdeep Mukhopadhyay, Pallab Dasgupta: “ALAFA: Automatic Leakage Assessment for Fault Attack Countermeasures”, in Proc. of the 56th Annual Design Automation Conference 2019, DAC 2019, Las Vegas, NV, USA, June 02-06, 2019, p. 136, ACM, 2019.
URL <https://doi.org/10.1145/3316781.3317763>

The talk presents an overview on automation for fault analysis attacks on cryptosystems. Fault attacks have emerged as a strong attack vector for crypto-implementations and thus need to be properly mitigated using suitable countermeasures. Test and analysis of these countermeasures, particularly in the black-box setting is thus of demand. The talk outlines two approaches: first a prototype tool called ExpFault to analyze differential fault analysis of ciphers at the algorithm level. Secondly, ALAFA, an automated leakage assessment framework was presented which derives its root from classical non-interference theorem and uses t-test based identification of leakage. The tool can be promising for security evaluation for protected crypto-designs and hardware security modules. More details can be found in [1, 2].

References

- 1 Sayandeep Saha, S. Nishok Kumar, Sikhar Patranabis, Debdeep Mukhopadhyay, Pallab Dasgupta: ALAFA: Automatic Leakage Assessment for Fault Attack Countermeasures. DAC 2019: 136
- 2 Sayandeep Saha, Debdeep Mukhopadhyay, Pallab Dasgupta: ExpFault: An Automated Framework for Exploitable Fault Characterization in Block Ciphers. IACR Trans. Cryptogr. Hardw. Embed. Syst. 2018(2): 242-276 (2018)

4.4 Securing Cyber-physical Control Systems – A Formal Perspective

Dey Soumyajit (Indian Institute of Technology – Kharagpur, IN)

License © Creative Commons BY 3.0 Unported license
© Dey Soumyajit

Joint work of Saurav K. Ghosh, Dey Soumyajit

Given the widespread deployment of cyber-physical systems and their safety-critical nature, reliability and security guarantees offered by such systems are of paramount importance. While security of such systems against sensor attacks have garnered significant attention from researchers in recent times, improving the reliability of a control software implementation against transient environmental disturbances need to be investigated further. Scalable formal methods for verification of actual control performance guarantee offered by software implementations of control laws in the face of sensory faults have been explored in recent works. However, the formal verification of the improvement of system reliability by incorporating sensor fault mitigation techniques like Kalman Filtering and Sensor Fusion remains to be explored. Moreover, system designers are bound to face complex trade-off choices for deciding upon the usage of fault and attack mitigation techniques and scheduling them on available system resources as they incur extra computation load.

In this talk, recent contributions for securing cyber-physical control systems are explained. These are threefold:

- Formally analyzing the actual performance guarantee of control software implementations enabled with additional fault mitigation techniques
- Considering task-level models of such implementations enabled with security and fault tolerance primitives and constructing a time-automata based model which checks for schedulability on heterogeneous multi-core platforms
- Leveraging these methodologies in the context of a novel Design-Space-Exploration (DSE) framework that considers target reliability and security guarantees for a control system, and computes schedulable design options while considering well-known platform level security improvement and fault mitigation techniques

The contributions are validated over several case studies from the automotive domain.

4.5 Challenges in Secure Composition from a Practical Perspective


Marc Stöttinger (Continental AG – Frankfurt, DE)

License  Creative Commons BY 3.0 Unported license
© Marc Stöttinger

Nowadays systems require security controls and countermeasures became system of systems with certain level of complexity due to composition of multiple software and hardware components. In this talk, practical challenges of putting security in composed systems were discussed on two examples. The first example demonstrates how an isolation layer for separation (established by two individual electrical components) is overcome by exploitations in the software domain. The second example discusses multiple exploitations on the example of an authentic communication and data exchange between a sensor node and a processing unit. The major issue in this example is the distributed development of both system components with too relaxed requirements.

4.6 Definition of “Root of Trust (RoT)”

Ingrid Verbauwhede (KU Leuven, BE)

License  Creative Commons BY 3.0 Unported license
© Ingrid Verbauwhede

This was an impromptu presentation of four slides to introduce Ingrid Verbauwhede’s definition of a “root of trust.”

First there is a clear distinction between what can be ‘trusted’ and what is ‘trustworthy’. Trusted cannot be verified and if the trust is broken, the system can fail. In the context of hardware design, we want to minimize what needs to be trusted. This is mapped on the design pyramid. So, “a root of trust is a component at a lower abstraction layer, upon which the system relies for its security.”

Some feedback received from the audience on the presentation:

- A component at a lower abstraction layer should be refined to “a component with an associated behavior or usage” for higher abstraction layers.

- A level below PUFs and TRNGs should be added to also include security problems at processing and technology levels (*e.g.*, to protect against Trojan circuits).
- This approach allows for reasoning on ‘defense in depth’. If a security violation is detected at some interface, the consequences can be systematically evaluated.

5 Discussions on State-of-the-art Questions

5.1 Working Group A1

Divya Arora (Intel – Santa Clara, US), Gaetan Cassiers (University of Louvain, BE), Johann Heyszl (Fraunhofer AISEC – München, DE), Itamar Levi (University of Louvain, BE), Debdeep Mukhopadhyay (Indian Institute of Technology – Kharagpur, IN), Kazuo Sakiyama (The University of Electro-Communications – Tokyo, JP), and Dey Soumyajit (Indian Institute of Technology – Kharagpur, IN)

License © Creative Commons BY 3.0 Unported license

© Divya Arora, Gaetan Cassiers, Johann Heyszl, Itamar Levi, Debdeep Mukhopadhyay, Kazuo Sakiyama, and Dey Soumyajit

Group A1 summarized their answers to below questions as follows.

1. How would you handle composability?
 2. What does it mean to securely compose two elements?
 3. Can you give an example of a security failure due to insecure composition?
 4. What is the foundation of composition?
 5. How do you verify remote identity of a connected device?
- **Horizontal and vertical composition:**
The composition of elements can be of two general forms which are different in terms of their interaction possibility. In a horizontal composition, the elements are independent from each other in the sense of running different execution environments (CPU or *e.g.* a state machine). Examples for this are embedded system’s printed circuit boards containing multiple chips. Another example are SoCs including CPUs and peripherals. Interaction is achieved through dedicated interfaces.
In a vertical composition, layers directly depend, respectively run on each other. They would typically share an execution flow. Examples are software layers conceptually running on top of each other (hardware, hypervisor, OS for a running system or hardware, bootloader and OS for the startup process). With vertical compositions, the interactions between elements (software layers) are usually much more complex.
 - **Example for a system failure due to insecure composition:**
In an example for a composed system, the CPU is fetching code from an external ROM and verifying its content before executing it from internal memory (cache). In a sense this can be seen as that the code is included into a trust bubble created by a RoT in the CPU for the purpose of system security. This code is part of the TCB for the system in the sense that the running system can jump into the code’s routines for security purposes. The issue is that the code might be flushed from internal memory (cache) during runtime meaning that it leaves the trust bubble. In the example, the code is then simply fetched from external ROM again but without a repeated verification. The example shows how the composition of the system lead to a critical time of check – time of use issue.

■ Solutions:


State of the art threat modelling helps to assess the security of composed systems.

On top of this, it seems like modelling and formulization are required, if composed systems are designed. This could result in a set of assertion-based checks which ensure security throughout the design flow. The drawback is that this assumes situations where the entire system is designed from scratch using this approach.

Runtime filters, possible configured using the above derived assertions, seem as a valid countermeasure to prevent such situations.

5.2 Working Group A2

Georg T. Becker (ESMT – Berlin, DE), Yaacov Belenky (Intel Israel – Haifa, IL), Shivam Bhasin (Nanyang TU – Singapore, SG), and Shahin Tajik (University of Florida – Gainesville, US)

License  Creative Commons BY 3.0 Unported license

© Georg T. Becker, Yaacov Belenky, Shivam Bhasin, and Shahin Tajik

■ How would you handle composability?

By formalization of threat models and security feature for a composable system and developing frameworks to verify these properties at different levels (*e.g.*, netlist generation, route & placement, *etc.*)

■ What does it mean to securely compose two elements?

1- Preserving the functionalities of secure elements A and B according to the security guarantees of A and B: *e.g.*, shared resources on FPGAs leading to side-channel sources and fault injections

2- Combining secure elements A and B as secure element C to achieve new security feature and functionalities.

■ Can you give an example of a security failure due to insecure composition?

Secure Element A is plugged to some non-security element B:

- Meltdown/Spectre (Memory protection + Speculative execution)

- Error Messages (Error reveals information and leaks sensitive information; *e.g.*, IEEE P1735 » Padding Oracle Attack on CBC mode)

■ What is the foundation of composition?

A possible foundation is a common language/specification in different levels of design and fabrication to assure the coherency between different blocks and assure the security (*e.g.*, constraints for CAD tools, Design Rule Checks (DRC)). However, first, one should overcome the challenge of describing the threat model for the composition.

■ How do you verify remote identity of a connected device?

This verification depends on threat model: There should at least be some secrets and this secret should be bounded to the identity (*e.g.*, identity-based cryptography, public-key infrastructure, device DNA, PUFs, *etc.*)

5.3 Working Group A3

Lucas Davi (Universität Duisburg-Essen, DE), Fatemeh Ganji (University of Florida – Gainesville, US), Tim Erhan Güneysu (Ruhr-Universität Bochum, DE), Ahmad-Reza Sadeghi (TU Darmstadt, DE), and Fareena Saqib (University of North Carolina – Charlotte, US)

License © Creative Commons BY 3.0 Unported license

© Lucas Davi, Fatemeh Ganji, Tim Erhan Güneysu, Ahmad-Reza Sadeghi, and Fareena Saqib

■ How would you handle composability?

In an ideal world, systematic security integration, analysis models, and penetration testing tools for whole system emulation are the ways to handle composability.

However, as there are integration issues in real world, a unified threat model with realistic assumptions together with divide and conquer approach considering security requirements can be the ways to handle composability.

■ What does it mean to securely compose two elements?

In *supply chain*: Security features on chip and the design house/foundry

At *microarchitectural level (Rowhammer attack)*: DRAM and access control & integrity checks on DRAM using software

In *cryptographic designs*: Mathematically strong algorithms and secure interfaces & implementation

At *system level (CANBus)*: Abstract isolation and standard security measures

■ Can you give an example of a security failure due to insecure composition?

Examples lie at different levels of abstraction. For example, there are supply chain threats and piracy. Other examples are microarchitecture level failures (Rowhammer bugs) and cryptographic algorithm failures due to implementation errors (memory corruption).

■ What is the foundation of composition?

- Clear requirements and assumptions that fit reality,
- Security metrics and confidence of metrics,
- Formal construction flow of integrating countermeasures (side effects of the protection countermeasures needs to be modeled),
- New opportunities for attack (for example, self healing logic can be exploited),
- Formal methods to measure and verify the security assumptions,
- Stress testing in order to test all the corner cases for full coverage and assurance.

■ How do you verify remote identity of a connected device?

This can be verified via mutual authentication, remote attestation, hardware fingerprints, and secure hardware protocols.

5.4 Working Group B1

Ran Canetti (Tel Aviv University, IL), Jean-Luc Danger (Telecom ParisTech, FR), Elif Bilge Kavun (University of Sheffield, GB), Osnat Keren (Bar-Ilan University, IL), Johannes Mittmann (BSI – Bonn, DE), and Ilia Polian (Universität Stuttgart, DE)

License © Creative Commons BY 3.0 Unported license

© Ran Canetti, Jean-Luc Danger, Elif Bilge Kavun, Osnat Keren, Johannes Mittmann, and Ilia Polian

■ How can we verify that mathematically proven properties are correctly implemented?

In the current certification practice, highest level (EAL7) includes requirements on formal techniques being used. On top of this, security properties can be checked at run time via

monitoring security properties of the entire system, making sure that secret key never appears on an SoC's bus, and detecting local attacks via sensors. Also, separation of system into small components may help as security then would be considered individually and future attacks would be contained to one component. Finally, techniques like modular redundancy that are used in safety context can be used to verify correct implementation.

■ **What is the starting point of trust?**

Here, the answer depends on the precise definition of “trust”.

We can converge towards root of trust through clarity, so the root of trust can be defined as the “clear” design. However, one has to trust vendor and the whole supply-chain where clarity is not possible (*e.g.*, design details kept secret due to certification).

■ **Does time play a role in composition?**

Time does play a role in composition; for example, timing of the composed system (order of events) is a side-channel. Delays also play a role in security solutions like blockchains.

■ **Can security (strictly) increase, decrease or stay constant as a result of composition?**

All three cases are possible – the point of concepts like Universal Composition is to prevent bad cases. Using composition to increase security sometimes looks obvious but in reality is difficult to prove rigorously. Security is often determined by weakest link of the system (composition may lead to security decrease). Sometimes, a system composed of imperfectly secure components becomes more secure due to composition (*e.g.*, hybrid key exchange via two mechanisms + XOR of the results), or secure components yield an insecure composition.

■ **What is the role of models in secure composition?**

A solution depends on how the problem is modeled. For example, in our context, security analysis will be based on (explicit or implicit) model of the attacker. However, while defining the models, we have to make sure that the unimportant details are omitted from the models and models on different levels of abstraction must be connected with each other.

5.5 Working Group B2

Elke De Mulder (Rambus – Sunnyvale, US), Elena Dubrova (KTH Royal Institute of Technology – Stockholm, SE), Yunsu Fei (Northeastern University – Boston, US), Paolo Palmieri (University College Cork, IE), and Milos Prvulovic (Georgia Institute of Technology – Atlanta, US)

License © Creative Commons BY 3.0 Unported license

© Elke De Mulder, Elena Dubrova, Yunsu Fei, Paolo Palmieri, and Milos Prvulovic

Group B2 answered below questions along with an identification of similarities and differences between them.

■ **Common aspects of questions:**

All of the questions have multiple interpretations in the context of secure composition, which makes them excellent from the perspective of identifying research problems and seeing the big picture rather than focusing on a particular set of solutions.

The group found that all the questions related to models of various aspects of security and models of security-relevant aspects of hardware/software systems. In particular, all the questions related to how assumptions built into the hardware designers' models

of their designs, and into models of potential attacks on those designs, are broken by successful attacks, and how composition tends to make these models' assumptions more problematic.

■ **Differences between questions:**

As the questions are highly inter-related, the group found it difficult to discuss the questions separately unless the problem space is significantly more constrained. However, the group also found that some of the questions, specifically questions 6 and 7, are about problems that are difficult even without composition, *i.e.*, the questions are about problems that exist (and are difficult to address) even in single-component systems.

■ **How can we verify that mathematically proven properties are correctly implemented?**

The group does not believe that it can be done for an arbitrary implementation, they believe that there should be a restriction so that the properties can be proven/verified. In order to verify the correctness of the implementation or properties, one needs to control the *entire* flow including not only the design itself, but also the supply chain.

■ **What is the starting point of trust?**

The group answers this question by asking if there is a need for a root of trust. Ideally, they would design a secure system without a RoT; however, in practice, it is not possible to avoid it. In that case, the group believes that the design, supply chain, and the user all need to be trusted.

■ **Does time play a role in composition?**

If time is understood as how much time an attacker has; yes, it does matter because time is a way of measuring security.

■ **Can security (strictly) increase, decrease or stay constant as a result of composition?**


The question is whether there are systematic ways of composing that maintain or improve security. This probably requires models with proofs and possibly prevents combinatorial explosion by finding which properties are provable for individual components.

■ **What is the role of models in secure composition?**

It is not really possible to have secure composition without models (there would be no systematic security without models). However, composition makes building models much more difficult. In order to deal with this, a hierarchy of models would be necessary. Also, problem-specific models are needed and countermeasures should be designed with respect to problems/models. However, there are still open questions like “Do multiple countermeasures work well together?”

5.6 Working Group B3

Annelie Heuser (IRISA – Rennes, FR), Michail Maniatakos (New York University – Abu Dhabi, AE), Wenjing Rao (University of Illinois – Chicago, US), Patrick Schaumont (Virginia Polytechnic Institute – Blacksburg, US), Werner Schindler (BSI – Bonn, DE), and Marc Stöttinger (Continental AG – Frankfurt, DE)

License  Creative Commons BY 3.0 Unported license

© Annelie Heuser, Michail Maniatakos, Wenjing Rao, Patrick Schaumont, Werner Schindler, and Marc Stöttinger

■ How can we verify that mathematically proven properties are correctly implemented?

More precisely, one usually tries to confirm model assumptions but not the conclusions of the mathematical model. The answer to the modified question depends on the particular feature or functionality. The correctness of implemented functions (*e.g.*, cryptographic algorithms) may be verified by known-answer tests. A known answer test constitutes a special case of a tests for particular properties, which itself is part of the implementation (see FIPS140-2). Another verification path could be the application of formally verified construction methods (see Common Criteria EAL6 or higher). The resistance against implementation attacks (side-channel attacks, fault attacks, *etc.*) may be confirmed by empirical analysis and experimental evaluation although this certainly is not a verification in a strict (mathematical) sense.

Moreover, we identified several open problems, which should be discussed. One question concerns the quantitative impact on the security of the system if some model assumptions are at least to some degree invalid. Another problem is how the completeness of the model assumptions can be/should be verified. Finally, it is not obvious at which stage one should check whether the implementation fulfils the model assumptions.

■ What is the starting point of trust?

There can be potentially several different starting points of trust, depending on the threat model, the amount of resources available, and the solution cost. During the phase of statement collection, many different opinions on the starting point of trust were discussed: No trust, Mathematical theorem and properties, Hardcoded reference value, RTL, hardware, secure tamper proof storage, TPM, smart card, bootloader routine, and BIOS. Some of them (*e.g.*, hardware, TPM) are currently used in practice as starting points of trust. The question, however, is what the optimal starting point of trust is. Approaching the question from the philosophical side, the starting point of trust can be defined as an entity that cannot be divided into smaller entities without affecting the trust assumption of the mathematical model. Or, the starting point of trust is an entity that is believed to guarantee central mathematical security assumptions; depending on the model assumptions it may be advantageous if its complexity is low. Following the complexity discussion, the starting point of trust should be as simple as possible to the extent that it does not need to be verified.

■ Does time play a role in composition?

This question was ranked at the bottom of the list for the group discussion, according to a vote at the initial phase in the interests of prioritizing the questions. The main issue seems to be that there is not a clearly understood foundation for this question, perhaps due to the general nature of “time” – what exactly does “time” mean, in the specific context of this question (about composition)? It would have been more helpful if this question had been narrowed down in the context of security composition, or with some examples.

■ **Can security (strictly) increase, decrease or stay constant as a result of composition?**

Security can either decrease or increase because of composition. First, the security properties of components may not scale up to the composition of the components because the security properties may not be transferable to the composition. An example of that phenomenon can be seen in the composition of individual PUF using an XOR operation into a so-called XOR-PUF. The XOR-PUF construction was proposed to harden the PUF against model-building attempts for the individual PUF. For example, while the arbiter PUF is susceptible to model building, the XOR-composition of arbiter PUF was thought to prevent model-building of the components. However, recent progress in machine learning attacks on XOR-PUF has shown this assumption to be incorrect. The problem is that the XOR operation is linear, and machine learning tools can still classify individual components under such a linear composition. Hence, in this case, one can argue that security decreases as the result of the composition. It decreases because the XOR-PUF offers the illusion of hardening against model-building. The composition is ineffective.

Second, the security properties of components may combine and strengthen the overall composition. An example can be seen in the composition of a better random number generator out of two biases random number generator using an XOR operation. In this case, the XOR result will generally show less bias and better overall distribution. The XOR is able to combine the entropy of each individual random number generator.

These two examples show that an identical operation (XOR) can be detrimental or beneficial to the security properties of the composition. Therefore, the composition of secure elements must be analyzed as well, even when the security properties of individual secure elements is well understood.

■ **What is the role of models in secure composition?**

Models are a key point in secure composition. Models make the analysis of a system with several components feasible and they are the starting point in many scenarios. They are a means for communication between different parties to achieve a common understanding of properties, threats, vulnerabilities, or interfaces. Having precise models make complex problems manageable, but also impose risks to overlook side-effects if an invalid or imprecise model is used. This is a particular risk the field of secure composition as even though individual layers may have been modeled precisely, their composition may include additional unexpected side-effects.

6 Discussions on Research Questions

6.1 Research Questions:

“What models and description languages are useful for formalization of security properties?”

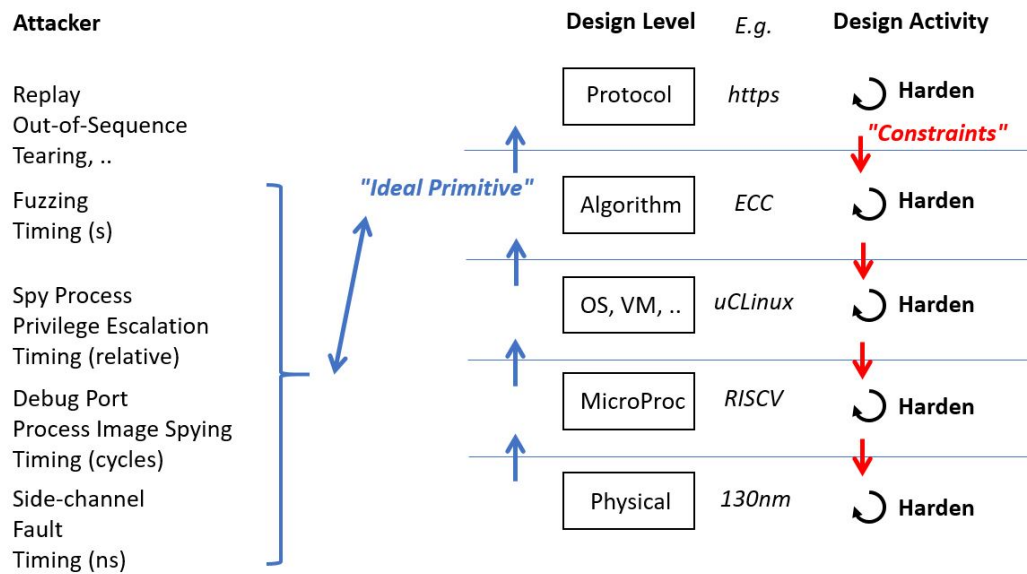
“Which protocol-level secure composition methods are applicable in hardware domain?”

Patrick Schaumont (Virginia Polytechnic Institute – Blacksburg, US), Shivam Bhasin (Nanyang TU – Singapore, SG), Debdeep Mukhopadhyay (Indian Institute of Technology – Kharagpur, IN), Francesco Regazzoni (University of Lugano, CH), Kazuo Sakiyama (The University of Electro-Communications – Tokyo, JP), Dey Soumyajit (Indian Institute of Technology – Kharagpur, IN), and Ingrid Verbauwhede (KU Leuven, BE)

License © Creative Commons BY 3.0 Unported license

© Patrick Schaumont, Shivam Bhasin, Debdeep Mukhopadhyay, Francesco Regazzoni, Kazuo Sakiyama, Dey Soumyajit, and Ingrid Verbauwhede

The group discussed these research questions and built the following “Layered Approach to Secure Composition of Electronic Systems” flow in Fig. 1 as a visual answer.



■ **Figure 1** Layered Approach to Secure Composition of Electronic Systems.

6.2 Research Questions:

“Can trust start in software, or are hardware roots and anchors of trust indispensable?”

“How to bootstrap trust in a distributed hardware system?”

Johann Heyszl (Fraunhofer AISEC – München, DE), Ran Canetti (Tel Aviv University, IL), Fatemeh Ganji (University of Florida – Gainesville, US), Michail Maniatakos (New York University – Abu Dhabi, AE), Marcel Medwed (NXP Semiconductors – Gratkorn, AT), Shahin Tajik (University of Florida – Gainesville, US), and Marten Van Dijk (University of Connecticut – Storrs, US)

License © Creative Commons BY 3.0 Unported license

© Johann Heyszl, Ran Canetti, Fatemeh Ganji, Michail Maniatakos, Marcel Medwed, Shahin Tajik, and Marten Van Dijk

1. There is no HW-free root of trust in embedded computing units

A root of trust is the minimal set of required building blocks (hardware and/or software; keys and/or routines; in all cases immutable) at the lowest possible abstraction layer, upon which the system’s security properties rely upon. The system’s higher level security properties (can we trust the system is behaving as intended) are built upon this RoT, similar like proofs are built upon axioms, by using more software as a TCB (secure boot helps to establish a TCB from a RoT). In the case of embedded security, the attacker is powerful and capable of hardware as well as software attacks. The attacker may, *e.g.*, replace code memory contents. Under these circumstances, to the best of our knowledge, there is no way to create such a RoT without the support of the hardware, hence, its manufacturer. Instead, the executing CPU needs to incorporate a minimal hardware RoT within the same chip (at least an executed routine optionally including either a fixed hash or routines for cryptographic verification of MACs and according symmetric or public key). It seems impossible to bootstrap a RoT on a system purely by supplying a piece of SW to be executed by the system. In such cases, the attacker is always able to execute higher-privileged code (*e.g.*, before and after) which is able to manipulate all system’s behaviour, hence, its security.

2. There is no extension of trust to other executing units without HW based RoTs

We consider multiple computing units within a system which are interconnected through (*e.g.*, low-bandwidth) interfaces. To the best of our knowledge, there is no way of bootstrapping or extending trust from a first one containing a hardware RoT to a second one which does not contain a hardware based RoT. This is under the assumption that one unit does not have highest privileged access to the memory of the other unit (and could hence have complete control over the execution of the other). For example, say one unit S1 is a secure element providing key storage and cryptographic operations. S1 can only trust a connected S2 based on the notion that this S2 would itself contain an equivalently trusted RoT. Hence the composed RoT essentially comprises both individual RoTs. Otherwise, an attacker may, *e.g.*, replace the software of S2 to another one providing all correct answers to S1 during a phase of trust establishment while running manipulated code before and/or afterwards. Similar examples are TPMs connected to CPUs (TPM 2.0 authenticated connection requires a key stored in the CPU – a RoT).

We discussed whether verifiable computing would help to run a small software RoT on S2 but came to the conclusion that it would be impossible since this software can be, *e.g.* run within a hypervisor by a manipulated version of S2. The topic of asking for software-only

RoT has a relation to white-box crypto which is essentially the attempt to have a secure software-based key storage without hardware support. It seems that similar limitations as with white box crypto would apply (white boxes can be lifted and executed in an emulator, hence need obfuscated interconnect to the application as much as possible) leading to the fact that a software-based RoT would only support trust under reduced attacker assumptions. Also the binding to the hardware, distance bounding in a sense, would be important. PUF instantiations help as long as the integrity of all relevant functionality of S2 (for the system's behaviour) is influencing the PUF response. PUFs used as keys storage can be part of a hardware RoT. Ideally, the challenge is, however, to devise a RoT without hardware reliance.

In summary, a hardware RoT is required in all system parts which are not fully controlled by one instance already containing a hardware RoT.

6.3 Research Question:

“Under what circumstances is security additive, and how can this be proven and validated?”

Jean-Luc Danger (Telecom ParisTech, FR), Yaacov Belenky (Intel Israel – Haifa, IL), Elke De Mulder (Rambus – Sunnyvale, US), Elena Dubrova (KTH Royal Institute of Technology – Stockholm, SE), Osnat Keren (Bar-Ilan University, IL), Johannes Mittmann (BSI – Bonn, DE), and Werner Schindler (BSI – Bonn, DE)

License © Creative Commons BY 3.0 Unported license

© Jean-Luc Danger, Yaacov Belenky, Elke De Mulder, Elena Dubrova, Osnat Keren, Johannes Mittmann, and Werner Schindler

A desirable aim is to combine independent security evaluations of component A and of component B (or of countermeasures against attack type A and attack type B, *etc.*) as this would reduce the complexity of the overall evaluation and thereby (hopefully) the probability of evaluation bugs and finally also the costs.

First, such an approach requires the definition of a suitable evaluation metric that is applicable to both the components and to the composed system. An evaluation metric might consider, *e.g.*, the resources required to carry out the most efficient (known) attacks. In evaluations according to the Common Criteria (CC), for example, numeric values for the factors “Elapsed Time”, “Expertise”, “Knowledge of the TOE” (TOE = target of evaluation), “Window of Opportunity”, and “Equipment” are used to derive an attack rating. Moreover, the components or countermeasures need to be ‘independent’ (to be defined) with regard to security properties.

Countermeasures against (A) side-channel attacks and (B) fault attacks, for example, are usually not independent because the latter often use redundancy to detect successfully induced faults. Redundancy, however, might favour side-channel attacks and thus both sets of countermeasures should not be evaluated independently but jointly. (A positive example might be the verification of an RSA-based signature by the exponentiation with the public exponent to prevent the Bellcore attack.)

Discussions suggested that components should allow such an ‘independence splitting’ more often than countermeasures against different attack types. Independence between components usually may not be valid in an information theoretical sense. Instead, it might be the conclusion of a careful ‘best-practice’ evaluation that the components A and B do not interfere in terms of security in an exploitable way. An example might be hardware sensors and the implementation of cryptographic algorithms. In a strict sense, an attacker might gain some local information about the implementation if he is able to identify the

position of hardware sensors. In many scenarios this knowledge yet might not allow to mount a successful attack. The evaluator often yet may be faced with ‘nested’ scenarios where the security evaluation may consider the component A first and then component B under consideration of A.

Finally, we formulate several heuristic criteria, which might justify a ‘practical’ independence assumption between different components A and B.

The components A and B

- are not nested
- have no common functionality related to the processing of secrets
- do not share resources that are used in the processing of secrets
- have no side-channels, which allow to combine information
- ...

These criteria are in general neither necessary nor sufficient for ‘practical independence’ of components but should support the decision making process.

6.4 Research Questions:

“How can existing hardware fulfill expectations and idealistic assumptions of protocols?”

“How to counter possible loss of security due to abstraction of hardware components?”

Elif Bilge Kavun (University of Sheffield, GB), Anupam Chattopadhyay (Nanyang TU – Singapore, SG), Annelie Heuser (IRISA – Rennes, FR), Johann Knechtel (New York University – Abu Dhabi, AE), and Itamar Levi (University of Louvain, BE)

License © Creative Commons BY 3.0 Unported license

© Elif Bilge Kavun, Anupam Chattopadhyay, Annelie Heuser, Johann Knechtel, and Itamar Levi

Protocol-level solutions require certain assumptions; however, these assumptions may not always (or even never) be met by hardware. An example to this is the perfect randomness versus device-level randomness: The expected level of randomness by the algorithm/scheme may not be provided by the randomness source on device. A solution to such problems could be:

1. Transfer of requirements to hardware in a way that they are also accessible to system architects/designers,
2. Finding methodologies so that system architects/designers can verify that the requirements are met.

Notion of abstraction is crucial in modern chip design. Hardware-related vulnerabilities are not well-defined in higher abstraction layers. “Secure abstraction” can be a solution – system design must not introduce vulnerabilities by the fact that some relevant lower-level details are invisible/encapsulated on higher layers.

Participants

- Divya Arora
Intel – Santa Clara, US
- Georg T. Becker
ESMT – Berlin, DE
- Yaacov Belenky
Intel Israel – Haifa, IL
- Shivam Bhasin
Nanyang TU – Singapore, SG
- Ran Canetti
Tel Aviv University, IL
- Gaetan Cassiers
University of Louvain, BE
- Anupam Chattopadhyay
Nanyang TU – Singapore, SG
- Jean-Luc Danger
Telecom ParisTech, FR
- Lucas Davi
Universität Duisburg-Essen, DE
- Elke De Mulder
Rambus – Sunnyvale, US
- Elena Dubrova
KTH Royal Institute of
Technology – Stockholm, SE
- Yungsi Fei
Northeastern University –
Boston, US
- Fatemeh Ganji
University of Florida –
Gainesville, US
- Tim Erhan Güneysu
Ruhr-Universität Bochum, DE
- Annelie Heuser
IRISA – Rennes, FR
- Johann Heyszl
Fraunhofer AISEC –
München, DE
- Elif Bilge Kavun
University of Sheffield, GB
- Osnat Keren
Bar-Ilan University, IL
- Johann Knechtel
New York University –
Abu Dhabi, AE
- Itamar Levi
University of Louvain, BE
- Michail Maniatakis
New York University –
Abu Dhabi, AE
- Marcel Medwed
NXP Semiconductors –
Gratkorn, AT
- Nele Mentens
KU Leuven, BE
- Johannes Mittmann
BSI – Bonn, DE
- Debdeep Mukhopadhyay
Indian Institute of Technology –
Kharagpur, IN
- Paolo Palmieri
University College Cork, IE
- Ilia Polian
Universität Stuttgart, DE
- Milos Prvulovic
Georgia Institute of Technology –
Atlanta, US
- Wenjing Rao
University of Illinois –
Chicago, US
- Francesco Regazzoni
University of Lugano, CH
- Ahmad-Reza Sadeghi
TU Darmstadt, DE
- Kazuo Sakiyama
The University of
Electro-Communications –
Tokyo, JP
- Fareena Saqib
University of North Carolina –
Charlotte, US
- Patrick Schaumont
Virginia Polytechnic Institute –
Blacksburg, US
- Werner Schindler
BSI – Bonn, DE
- Georg Sigl
TU München, DE
- Dey Soumyajit
Indian Institute of Technology –
Kharagpur, IN
- Marc Stöttinger
Continental AG – Frankfurt, DE
- Shahin Tajik
University of Florida –
Gainesville, US
- Marten Van Dijk
University of Connecticut –
Storrs, US
- Ingrid Verbauwhede
KU Leuven, BE



Cybersafety Threats – from Deception to Aggression

Edited by

Zinaida Benenson¹, Marianne Junger², Daniela Oliveira³, and
Gianluca Stringhini⁴

1 Universität Erlangen-Nürnberg, DE, zinaida.benenson@fau.de

2 University of Twente, NL, m.junger@utwente.nl

3 University of Florida – Gainesville, US, daniela@ece.ufl.edu

4 Boston University, US, gian@bu.edu

Abstract

A number of malicious activities, such as cyberbullying, disinformation, and phishing, are becoming increasingly serious, affecting the wellbeing of Internet users both financially and psychologically. These malicious activities are inherently socio-technical, and therefore effective countermeasures against them must draw not only from engineering and computer science, but also from other disciplines. To discuss these topics and find appropriate countermeasures, we assembled a group of researchers from a number of disciplines such as computer science, criminology, crime science, psychology, and education. Through five days of brainstorming and discussion, the participants developed a roadmap for future research on these topics, along four directions: modelling the attackers, measuring human behavior, detection and prevention approaches for online threats to adolescents, and understanding unintended consequences of mitigation techniques.

Seminar July 21–26, 2019 – <http://www.dagstuhl.de/19302>

2012 ACM Subject Classification Social and professional topics → Computer crime, Security and privacy → Human and societal aspects of security and privacy, Security and privacy → Social engineering attacks

Keywords and phrases Cybersafety, Legal and Ethical Issues on the Web, Online Social Networks, Security and Privacy

Digital Object Identifier 10.4230/DagRep.9.7.117

Edited in cooperation with Matthew Edwards


1 Executive Summary

Zinaida Benenson (Universität Erlangen-Nürnberg, DE)

Marianne Junger (University of Twente, NL)

Daniela Oliveira (University of Florida – Gainesville, US)

Gianluca Stringhini (Boston University, US)

License  Creative Commons BY 3.0 Unported license

© Zinaida Benenson, Marianne Junger, Daniela Oliveira, and Gianluca Stringhini

A number of malicious activities are prospering online and are putting users at risk. In particular, cyber deception and cyber aggression practices are increasing their reach and seriousness, leading to a number of harmful practices such as phishing, disinformation, radicalization, and cyberbullying. Attack strategies include controlling and operating fake or compromised social media accounts, artificially manipulating the reputation of online entities, spreading false information, and manipulating users via psychological principles of influence into performing behaviors that are counter to their best interests and benefit the attackers.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Cybersafety Threats – from Deception to Aggression, *Dagstuhl Reports*, Vol. 9, Issue 7, pp. 117–154

Editors: Zinaida Benenson, Marianne Junger, Daniela Oliveira, and Gianluca Stringhini



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

So far, computer science research on cybersafety has looked at the various sub-problems in isolation, mostly relying on algorithms aimed at threat detection, and without considering the implications of the attacks and countermeasures for individual users as well as for society. On the other hand, human factors and social science researchers often consider user interfaces and social interactions without taking full advantage of the algorithmic, data-driven cybersafety research. Moreover, the legal and ethical implications of attacks and countermeasures are often unclear.

The goal of the Dagstuhl Seminar 19302 “Cybersafety Threats – from Deception to Aggression” was to provide a platform for researchers to look at the problem of cybersafety from a holistic and multi-disciplinary perspective. The participants were drawn from a number of disciplines such as computer science, criminology, psychology, and education, with the aim of developing new ideas to understand and mitigate the problems.

At the beginning of the seminar, we asked participants to identify important themes to focus on, and these themes were refined through specific activities and discussions during the first day: Firstly, all participants gave 5-minute talks where they presented their current research related to the seminar, and their expectations and topics they would like to work on during the week. Secondly, we conducted three *introductory panels* on the topics of *Cyber Deception*, *Cyber Aggression* and *Propaganda & Disinformation*. Each panel consisted of five participants. We took special care to represent different disciplines and different career stages in each panel.

By the beginning of the second day, participants had identified four key themes to study in this area, which we describe in detail in the rest of this section. The participants formed working groups (WGs) for each theme.

Theme 1: Attacker modeling

The working group focused on predicting the next steps of an ongoing attack by means of a probabilistic model. The initial model developed by the group consists of 9 variables: attacker goals, characteristics of the attack (e.g., how long the attack takes, tools employed), consequences, authorization, attribution, expected resilience of the victim, expected characteristics of the victim from attacker’s perspective, actual characteristics of the victim, actual responsiveness of the victim. The developed model was verified and refined using two known attacks as case studies: the Internet Worm (1988) and the SpamHaus DDoS attack (2013).

Two most important next steps to refine the model are:

1. Convert the variables into measurable quantities
2. Obtain labeled data on which the model can be trained

The working group started working on a conceptual paper that describes the model, and discussed possible venues for its publication. Several methods of obtaining the data for the model were proposed, such as interviewing CISOs and other defenders, creating financial incentives for organization to share their data, and organizing a stakeholder workshop including not only defenders, but also former attackers who now work as security consultants.

Theme 2: Unintended consequences of countermeasures

This working group focused on an often overlooked aspect of computer security research: the fact that deploying any countermeasure to mitigate malicious online activity can have unexpected consequences and harms to other parties. The members of this working group started by discussing a number of scenarios: intimate partner abuse, CEO fraud, disinformation, online dating fraud, and phishing, and developed a taxonomy of these potential harms.

The taxonomy takes into account not only technical issues that might arise from deploying countermeasures but also socio-technical ones such as the displacement effect of attackers moving to other victims, the additional costs incurred by using the countermeasure, and the issues arising from complacency, for example leaving users desensitized by displaying too many alerts to prevent a certain type of attack.

Theme 3: Measuring human behavior from information security (and societal) perspectives

Measuring online behavior is of fundamental importance to gain an accurate understanding of malicious online activities such as cybercrime. The research community, however, does not have well established techniques to accurately measure this behavior, and this can lead to studies presenting largely contradicting results. This working group focused on identifying techniques relevant to measure and model various types of online behavior, from cyberbullying and disinformation to ransomware and phishing. As a final outcome, the working group drafted two methodological frameworks for researchers aiming to study these problems, one focused on socio-technical threats (cyberbullying and disinformation) and one focused on cybersecurity (phishing and malware).

Theme 4: Prevention, detection, response and recovery.

A key challenge when mitigating socio-technical issues is developing the most effective countermeasures. This group focused on developing detection and prevention approaches focusing on threats encountered by adolescents when surfing the Web (e.g., cybergrooming). A common issue here is that adolescents rarely turn to adults for help, and therefore any mitigation based on direct parental oversight has limited effectiveness. To go beyond these issues, the group developed a mitigation strategy based on a “guardian angel” approach. The idea is to let a minor create a “guardian avatar” that will then advise them on cybersafety practices, with a decreasing level of oversight as the minor grows up. While the children are very young, the guardian avatar will closely supervise them, reporting any suspicious contacts that they have online to a parent or a guardian. Later, as the child enters adolescence, the avatar will gradually take on an advisory role, eventually only providing advice once the adolescent asks for it. The group considered privacy issues and interdisciplinary aspects related to psychology and education, and developed a proposal of how the avatar would work.

Conclusion and Future Work

The seminar produced a number of ideas on how to investigate and mitigate cybersafety threats. It enabled researchers from different disciplines to connect, and set the agenda for potentially impactful research to be carried out in the next years. Joint publications and funding for joint research were discussed in each WG and later in the plenum. For example, WG 3 considered possibilities for a large international grant, such as H2020. The ideas produced as part of theme 4 resulted in the paper “Identifying Unintended Harms of Cybersecurity Countermeasures” to appear at the APWG eCrime Symposium in November 2019.

2 Table of Contents

Executive Summary

Zinaida Benenson, Marianne Junger, Daniela Oliveira, and Gianluca Stringhini . 117

Overview of Talks

Empirically measuring the economic impact of cyber attacks <i>Abhishta Abhishta</i>	122
Teaching People Not to Fall for Cyber Deception Might Be Harmful <i>Zinaida Benenson</i>	122
Inconsistent Deception and Attribution <i>Matt Bishop</i>	123
Research in Social Engineering <i>Jan-Willem Bullée</i>	124
The Federal Trade Commission <i>Joe Calandrino</i>	124
MITRE's Human Behavior and Cybersecurity Research and Capabilities <i>Deanna Caputo</i>	125
Towards Cognitive Security <i>Claude Castelluccia</i>	126
Measuring Online Radicalisation <i>Yi Ting Chua</i>	126
The Neurobiology of Financial Abuse <i>Natalie Ebner</i>	126
Research in Online Fraud <i>Matthew Edwards</i>	127
The Sociology of Phishing <i>Freya Gassmann</i>	127
Caught in the Crossfire / The language of aggression, violence, and cybercrime <i>Alice Hutchings</i>	128
Psychological aspects of Cybercrime <i>Marianne Junger</i>	129
Research in Security Risk Management <i>Katsiaryna Labunets</i>	130
Research in Phishing <i>Elmer Lastdrager</i>	130
Phishing Susceptibility as a Function of Age, Gender, Weapon of Influence, and Life Domain <i>Daniela Oliveira</i>	130
Cyber Deception and Cyber Aggression <i>Simon Parkin</i>	131

Get to know your geek: towards a sociological understanding of incentives to develop privacy-friendly free and open source software	
<i>Stefan Schiffner</i>	131
Characterizing Disturbing and Reactionary Content in Youtube	
<i>Michael Sirivianos</i>	132
Characterization, Detection and Mitigation of Antisocial Behaviour	
<i>Ivan Srba</i>	132
Measuring and Modeling the Online Information Ecosystem	
<i>Gianluca Stringhini</i>	133
DISinformation as a Political Game	
<i>Gareth Tyson</i>	134
Language-based deception detection	
<i>Sophie van Der Zee</i>	135
Applying Routine Activity Theory to Cybervictimization: A Theoretical and Empirical Approach	
<i>Sebastian Wachs</i>	135
Research in Evidence-based Security	
<i>Victoria Wang</i>	136
Deception and deterrence	
<i>Jeff Yan</i>	137

Working groups


Theme 1: Attacker Modeling Group	
<i>Abhishta Abhishta, Zinaida Benenson, Matt Bishop, Joe Calandrino, Natalie Ebner, Manuel Egele, William Robertson, Victoria Wang, and Savvas Zannettou</i>	137
Theme 2: Unexpected Consequences of Countermeasures	
<i>Matthew Edwards, Yi Ting Chua, Alice Hutchings, Daniela Oliveira, Simon Parkin, Stefan Schiffner, and Gareth Tyson</i>	143
Theme 3: Measuring Human Behavior from Information Security and Societal Perspectives	
<i>Ivan Srba, Katsiaryna Labunets, and Sophie van Der Zee</i>	150
Theme 4: Prevention, Detection, Response and Recovery	
<i>Gianluca Stringhini, Freya Gassmann, Marianne Junger, Elmer Lastdrager, Michael Sirivianos, and Sebastian Wachs</i>	152

Participants	154
------------------------	-----

3 Overview of Talks

3.1 Empirically measuring the economic impact of cyber attacks

Abhishta Abhishta (University of Twente, NL)

License  Creative Commons BY 3.0 Unported license
© Abhishta Abhishta

Measuring the economic impact of cyber crime just by the use of surveys does not provide an accurate picture of the real losses. Well, if you ask people what they don't know, they are bound to provide you with the perception of the answer, which might not be the real answer. This is one of the reasons why we see the losses due to cybercrime being reported in millions of dollars.

A solution for this problem is to empirically measure the economic impact of cyber crimes. This can be done by using many of the newly collected datasets such as the OpenINTEL [1]. However, this method has its own short comings. It is not always possible to get the datasets that can be used to measure economic impact (privacy reasons). An example of this is collection of "work study"/"time study" measurements in an IT firm to estimate the true impact of IT downtime due to a cyber attack. "Work study"/"time study" methods have been used in the manufacturing industry to measure the impact of downtime in assembly lines.

The research question I have for this workshop related to the problem described above is: *How can we in a privacy friendly way take "work study" / "time study" measurements at an IT company?*

Another research question that I am interested in and is related to the theme of the workshop is: *As fake news has been around even before the internet, can we learn from how the history has dealt with fake news and use the similar solutions for the current problem.*

References

- 1 Abhishta, A., van Rijswijk-Deij, R., & Nieuwenhuis, L. J. M. *Measuring the impact of a successful DDoS attack on the customer behaviour of managed DNS service providers.* Computer communication review, 48(5), 70-76, 2018

3.2 Teaching People Not to Fall for Cyber Deception Might Be Harmful

Zinaida Benenson (Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 3.0 Unported license
© Zinaida Benenson

In 2014, my colleagues and I conducted a phishing experiment with a (then) novel design: We recruited over 1200 university students for a study on online behavior, but sent to them a simulated phishing message from a non-existing person. The message referred to a party last week, and contained a suspicious link to the party pictures. After several days, we sent to the participants a questionnaire that debriefed them about the true purpose of the study, and asked them for reasons of their clicking behavior. The most frequently reported reason for clicking was curiosity (34 percent), followed by the explanations that the message fit recipient's expectations (27 percent), as they attended a party last week. Moreover,

16 percent thought that they might know the sender. These results show that decisional heuristics for message processing are relatively easy to misuse, if the attack message refers to work or life interests of the people, or spoofs a known sender.

Defense against spear phishing and other targeted attacks seems to be especially challenging because of the ambiguity of the situations that they create, making the context and content of the message look plausible and legitimate. Because of this ambiguity, asking people to be permanently vigilant when they process their messages might have unintended negative consequences. For example, if their job requires processing a lot of invoices sent via email, they might click on a ransomware-infected file called `invoice.doc`, as this fits their job expectations. But if they are taught to be careful with invoices, they might start missing or delaying the real ones, which stands in a direct conflict with the requirements of their job. Under these circumstances, the employees are likely to disregard this kind of user education attempts after some time, because the only way for them to get their job done in time is to process their emails as quickly as possible, without extra security checks. However, in case their organization sends to them simulated phishing messages in order to increase their security awareness, they may become disgruntled and unmotivated, or start blaming themselves for inability to make a correct decision in an ambiguous situation under time pressure.

Although our study led us to hypothesize about negative consequences of the human-centered anti-phishing defenses, we do not have enough evidence to support these hypotheses. Thus, one of the most important directions for future research is development of study designs and measurement procedures for assessing not only effectiveness of anti-phishing measures, but also their impact on the work and life environment of people, and on their psychological well-being.

3.3 Inconsistent Deception and Attribution

Matt Bishop (University of California – Davis, US)

License © Creative Commons BY 3.0 Unported license
© Matt Bishop

Deception is an ages-old tactic for confusing an adversary. In computer science, deception presents a “fiction”, or false reality, to the adversary. The adversary will then act and react based on this false image of the system, and the defenders can have the fiction respond in ways that will cause the attacker to reveal information and methods about the goals and attack techniques. This requires time and resources as well as planning for the attack and developing the fiction.

If the goal is to prevent the attacker from obtaining information, then the defenders must ensure the attackers do not know whether they have succeeded. For this, the consistency of the common fictions is unnecessary. Inconsistent deception confuses the adversary so they do not know what is true; they may know they are being deceived (and probably will), but so long as they cannot determine what is accurate, they cannot know when they have succeeded in finding or altering the information.

Attribution is a key part of defense, because the defenders want to know who or what organization(s) are behind the attack; similarly, the attackers will want to hide that information, possibly using deception to trick the defenders into misattributing the attack. An interesting and relevant question is how and when attribution should be provided, and the effects of different types and levels of assurance of that attribution co-existing on a system or network (such as the Internet).

The research questions I have that are relevant to this workshop are:

- Inconsistent deception is based on the theory that it will confuse an adversary, to the point that the adversary will go away. How would one validate or refute this theory?
- Could one frame inconsistent deception in such a way it seems like the system is flaky rather than the adversary being deliberately deceived?
- Under what conditions do the different types of attribution meet the needs of the involved (and intermediate) entities?
- How would one tie attribution to particular roles, and manage this connection, in a network like the Internet?

3.4 Research in Social Engineering


Jan-Willem Bullée (University of Twente, NL)

License  Creative Commons BY 3.0 Unported license
© Jan-Willem Bullée

I am Jan-Willem Bullee, and I am a Postdoctoral researcher at Linköping University in Sweden and a visiting researcher at Erasmus University Rotterdam in The Netherlands. In this capacity, I work closely with Prof Jeff Yan (LIU) and Dr Sophie van der Zee (EUR). During my doctoral research, I investigated social engineering (a form of cybercrime) in an organisational setting. I was particularly interested in the factors that explain and reduce victimisation of social engineering attacks. I explored three types of social engineering (i.e. face-to-face, telephone and email) in field experiments. Furthermore, I made a meta-analysis on social engineering interventions and a systematic review of the success of phishing emails. I also presented research ideas related to obtaining more insight into email phishing. For example: How can boosters be used to reduce the decay effect of an intervention; and what is the role of culture on the success of a phishing email?

3.5 The Federal Trade Commission


Joe Calandrino (Federal Trade Commission – Washington, US)

License  Creative Commons BY 3.0 Unported license
© Joe Calandrino

The Federal Trade Commission is the US government's primary consumer protection agency. The laws that the agency enforces include ones prohibiting deceptive practices in or affecting commerce. The FTC's Office of Technology Research and Investigation has a number of roles, which include conducting research relevant to the agency's mission. Our research has explored topics from email authentication to targeted advertising. Through research that helps identify, understand, and prevent potential deceptive practices, Dagstuhl attendees can help us protect consumers against such practices.

3.6 MITRE's Human Behavior and Cybersecurity Research and Capabilities


Deanna Caputo (MITRE – Washington D.C., US)

License  Creative Commons BY 3.0 Unported license
© Deanna Caputo

Cybersecurity has been primarily tackled from the technological perspective in academia, government, and industry. Focusing on the human aspects without training in the behavioral sciences reduces effectiveness. Behavioral scientists uniquely bring applied subject-matter-expertise in human behavior to cybersecurity challenges. MITRE, as a not-for-profit who manages federally funded research and development centers, leverages human behavior to reduce cybersecurity risk using the behavioral sciences to understand and strengthen the human firewall through its Human Behavior and Cybersecurity Capability area. We utilize operational research and consultations, as well as direct sponsors' unpublished best practices across projects and portfolios to improve government and national critical infrastructure, particularly insider threat, usable security & technology adoption, cyber risk perceptions & awareness, cyber exercises & teams. Currently, we have been tasked with creating a data-driven insider threat framework that includes psycho-social and cyber-physical characteristics that could be common, observable indicators for insider attacks. Existing frameworks ignore psycho-social characteristics or are based on poor quality data. MITRE will receive, store, structure, hand-code, aggregate, and analyze a large dataset (5-10K) of raw insider threat case investigation files shared directly from multiple organizations. The framework will include: insider attacker's actions before, during, and after an attack; individual-level factors (e.g., role, character, stressors, motivations, intent); organizational factors (organizational procedures, infrastructure elements, security elements, peer information, sector); and key flags and events that led to major decisions in the inquiry/investigation. In addition, to counter the issue of underreporting of insider risks using human sensors, MITRE has conceptualized and developed an Insider Risk Personas Methodology aimed at helping government and critical industry infrastructure to operationalize insider risk in a manner that is relevant, tangible, time-practical and expandable to supervisors/HR. The outcome of the methodology is a set of evidence-based personas that are designed to help supervisors directly challenge the rationalizations that they offer for under-reporting employee risks, increase supervisor confidence and good judgments of employee risk, and increase employee risk reporting in terms of both frequency and quality. We are currently developing and will test and evaluate a set of insider risk personas specifically for the financial critical infrastructure sector. Other problem areas for multi-disciplinary (not interdisciplinary) collaboration between the behavioral and cybersecurity sciences include: imposing costs on cyber threat actors, changing cyber adversary behavior, measuring cybersecurity awareness programs, and the impact of cyberattack response/recovery on public perception/trust.

3.7 Towards Cognitive Security


Claude Castelluccia (INRIA – Grenoble, FR)

License  Creative Commons BY 3.0 Unported license
© Claude Castelluccia

My talk was about Cognitive Security. We tend to think of cyber-attacks, or cybersecurity in general, as network intrusions, malware, Denial Of Service (DOS) attacks or other exploits that compromise physical infrastructures. However recent events, such as the Russian interference attacks on the US election, have shown that humans are increasingly becoming the targets of attacks. Instead of attacking infrastructures, adversaries are using information and existing services, such as social networks, to manipulate people. Adversaries attack humans via weaponized information. Information disorder has evolved from a nuisance into high-stakes information war. It is urgent to secure our “cognitive infrastructure”. My talk discussed the foundations of the field of cognitive security. I presented a systematic analysis framework to help scientists and policy makers to tackle the topic. More specifically, the proposed framework combines the IP (Information Processing) model, used in cognitive psychology, together with the CIA (Confidentiality, Integrity, Availability) triad, used in information security, to conceptualize the field of cognitive security. Although this approach might seem simplistic and should not be taken literally, we believe it provides a useful framework to start building the foundations of cognitive security.

3.8 Measuring Online Radicalisation

Yi Ting Chua (University of Cambridge, GB)

License  Creative Commons BY 3.0 Unported license
© Yi Ting Chua

My presentation focused on the topic of online radicalisation. Using repeated measures analysis of variance (RM-ANOVA) and social network analysis, the study found changes in expressed ideological beliefs both at the forum and individual level. Specifically, differential reinforcement and differential association were the strongest predictors towards changes in expressed far-right ideological beliefs which include beliefs such as xenophobic, anti-semantic and anti-taxation.

3.9 The Neurobiology of Financial Abuse

Natalie Ebner (University of Florida – Gainesville, US)

License  Creative Commons BY 3.0 Unported license
© Natalie Ebner

Financial abuse is one of the most common forms of elder mistreatment, with devastating consequences. A rapidly aging population, combined with changes in decision making, render fraud targeting older adults a public-health concern. Technological advances open novel avenues for fraud. Older adults increasingly navigate the Internet and are at increased risk of becoming victims of cyber social-engineering attacks, such as phishing emails, which lure users into visiting webpages that procure personal information or into clicking on malicious

links. We adopted an ecologically valid approach to uncover age-related vulnerabilities in trust-related decision making. Study 1 recorded browsing activity over 3 weeks, during which young and older participants, unbeknownst to them, received simulated phishing emails. Close to half of the users were susceptible to phishing, with older women most vulnerable. There was a discrepancy, particularly among older users, between self-reported susceptibility awareness and behavior. Examining specific risk profiles, higher susceptibility was associated with lower memory and positive affect among the oldest users. In a complementary study, we contrasted brain structure and function in older adults who were victims of fraud with older adults who had avoided an attempted fraud. The exploited group showed cortical thinning in anterior insula and reduced functional connectivity within default and salience networks, while increased between-network connectivity. Thus, alterations in brain regions implicated in trust-related decision making may signal heightened fraud risk in older adults. Our data advance understanding of brain and behavioral processes underlying age-related vulnerabilities to fraud online and in-person. Determination of cognitive, socio-affective, and neurobiological risk profiles is crucial to develop prevention against victimization in aging, which can have dramatic consequences for the individual and society.

3.10 Research in Online Fraud


Matthew Edwards (University of Bristol, GB)

License  Creative Commons BY 3.0 Unported license
© Matthew Edwards

In this brief introductory presentation, I discussed elements of my research background which were related to the topic of this Dagstuhl seminar: my work on persuasion in 419 email scam exchanges, detecting online dating fraud profiles and ongoing work investigating cybercriminal fora. While I am a computer scientist, my work has been carried out in close collaboration with psychologists, and psychology informs a lot of my research. In the work on 419 scam exchanges, we have been looking at the traces of persuasion principles we can observe by looking at the text of scambaiter and victim interactions with scammers – some of which are extraordinarily long-lived. Our work on dating fraud profiles built upon suggestions that users with more romantic naiveté were more likely to become victims, building automatic classifiers that distinguish between the profiles of scammers and real dating site users. In my ongoing work, I am looking at evidence about the characteristics, historic impact, and careers of cybercriminals in underground forums.

3.11 The Sociology of Phishing

Freya Gassmann (Universität des Saarlandes, DE)

License  Creative Commons BY 3.0 Unported license
© Freya Gassmann

In general, my topics are university research, IT-Security, sport sociology and methods in social science. As a sociologist I am interested in the social science part of IT-security and quantitative data collection and analysis methods. In the last years I worked together with Zina on some projects. The last two papers were about phishing and we tried to figure out, why people click on a link in an email or Facebook message. In a field experiment over 1200

university students received an email or a Facebook message with a link to (non-existing) party pictures from a non-existing person. In a questionnaire there were asked about their clicking behavior. The most frequently reported reason for clicking was curiosity followed by the explanations that the message fit to the circumstances of the participants.

I am interested in the following questions: Why do people act risky (data protection and phishing). Are they careless or do they don't understand the importance of data and data protection? If this would be the case: Do we need better and more education for children, young adults and employees?

References

- 1 Gassmann, F., Benenson, Z., & Landwirth, R. *Kommunikation als Gefahr: Nutzerreaktion auf Nachrichten mit verdächtigen Links per E-Mail und Facebook* Österreichische Zeitschrift für Soziologie, 44(S1), 135–155, 2019 <https://doi.org/https://doi.org/10.1007/s11614-019-00351-6>
- 2 Benenson, Z., Gassmann, F., & Landwirth, R. *Unpacking Spear Phishing Susceptibility*. In M. Brenner, K. Rohloff, J. Bonneau, A. Miller, P. Y. A. Ryan, V. Teague, . . . R. Landwirth (Eds.), *Financial Cryptography and Data Security. FC 2017. Lecture Notes in Computer Science*, vol 10323 (pp. 610–627), 2017
- 3 Gassmann, F., Beck, J., Gourmelon, N. & Benenson, Z. What is more valuable: Confidentiality or availability of data? Work in progress on an online experiment using willingness to pay (WTP) in a ransomware scenario to examine users' valuation of their data. Poster at the conference of the "Akademie für Soziologie". *Digitalsocieties2019*, September 25-27, Konstanz, 2019

3.12 Caught in the Crossfire / The language of aggression, violence, and cybercrime

Alice Hutchings (University of Cambridge, GB)

License  Creative Commons BY 3.0 Unported license
© Alice Hutchings

eWhoring is a term used by offenders to refer to social engineering techniques where they imitate partners in virtual sexual encounters. Victims are asked for money in exchange for pictures, videos, or sexual-related conversations. The harms associated with eWhoring, which involves fraud by misrepresentation, include the exploitation of those being impersonated, usually young women. Some of the images being distributed are indecent images of children, or material leaked as 'revenge porn'. My previous research provides a crime script analysis of eWhoring, identifying the steps involved, the types of actors, and points for intervention. However, one of the concerns about the intervention approaches developed was the impact on those caught in the crossfire. It is important to consider the impact of crime prevention initiatives on the law abiding majority. In some cases, this may cause additional nuisance, such as the time and effort required for account verification. In other cases, it may have particularly adverse impacts on those already marginalised. In the context of eWhoring, this includes those involved in legitimate sex work, particularly if their images are stolen and used fraudulently.

In relation to aggression, my colleagues and I found that the language used on Hackforums was less aggressive than Wikipedia page edit comments. This is perhaps due to its relatively homogenous population. Targets for harassment are likely to be located off, rather than on,

the forum. However, the language used by the computer security is also interesting. Despite cybercrime being relatively non-physical in nature, the language used to describe it is often borrowed from the areas of aggression and violence. For example, we refer to incidents as ‘attacks’, and targets being ‘hit’. ‘Hacking’ has relatively sinister connotations, as does ‘defacing’. There are further examples: ‘brute force’, ‘penetration testing’, ‘smashing the stack’, ‘bomb’ (e.g. logic, fork, zip), ‘Heartbleed’, ‘Rowhammer’, ‘Shellshock’, ‘Bashbug’, and even cyberwarfare. Does this represent something about the way we perceive cybercrime? Does it relate to the way it is represented, is it framed in such a way to be considered newsworthy? Or perhaps it reflects the relative masculinity of the computer security industry?

3.13 Psychological aspects of Cybercrime

Marianne Junger (University of Twente, NL)

License © Creative Commons BY 3.0 Unported license
© Marianne Junger


First, in my presentation I have presented slides based on research on the origins of aggression in humans [1]. I stated that aggression is an innate drive in humans. Accordingly, it is ever-present behavioral option, starting at birth. Therefore, aggression has to be unlearned in childhood and this needs to be done before age 8. This unlearning process is done through a socialization process by parents and teachers. The result is that children are taught self-control. After age 8, behavioral tendencies remain relatively stable over life [2, 3, 4]. The level of self-control that has been reached has many implications. First, humans differ on self-control, not everyone has been socialized equally well. Probably genetic differences may make some children a little harder to socialize. Also, with low-self-control, humans are prone to commit all sorts of deviant behaviors, that is, all sorts of crimes and all types of risky and unhealthy behaviors. Second, I mentioned that humans have ‘truth bias’ [5]. This bias facilitates crime victimization.

References

- 1 Tremblay, R.E., Developmental origins of disruptive behaviour problems: the ‘original sin’ hypothesis, epigenetics and their consequences for prevention. *Journal of Child Psychology and Psychiatry*, 2010. 51(4): p. 341–367.
- 2 Olweus, D., Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin*, 1979. 86(4): p. 852-875.
- 3 Piquero, A.R., et al., Stability in aggression revisited. *Aggression and Violent Behavior*, 2012. 17(4): p. 365-372.
- 4 Heckman, J.J., Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, 2006. 312(5782): p. 1900 – 1902.
- 5 Burgoon, J.K. and T.R. Levine, Advances in deception detection. *New directions in interpersonal communication research*, 2010: p. 201-220.

3.14 Research in Security Risk Management

Katsiaryna Labunets (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Katsiaryna Labunets


My background is in cyber risk management and empirical research. In my PhD thesis, I conducted an empirical comparison of security risk assessment methods and investigated the criteria behind methods' success. However, cyber risk management based just on technical solutions cannot provide 100% security to organisations. Therefore, in the past years, my research focus is on how combined security measures can effectively manage human-related threats. My future research interests include security behaviour definition from organisation management and employees perspective and how actual security behaviour can be measured and explained.

In my talk at Dagstuhl, I proposed a few ideas for the workshop:

- Use a honeypot network to catch, study and suppress cyberbullies;
- Apply a serious gaming approach to train adults about cyberbullying and how to deal with this;
- Develop a catalogue of social/human-specific cyber threats and related countermeasures that can become a part of an information security standard and used by existing cyber risk assessment methods.

3.15 Research in Phishing

Elmer Lastdrager (SIDN Labs – Arnheim, NL)

License  Creative Commons BY 3.0 Unported license
© Elmer Lastdrager

In this introduction talk, I discussed my PhD research on phishing. Specifically, I discussed studies on thinking out loud, teaching children how to recognise phishing emails and websites, and a brief overview of analysing 700.000 phishing emails. After that, I discussed my research interests in Internet of Things (IoT), which cover both technical solutions (e.g., analysing network traffic) and user-oriented solutions (e.g., improving user cyber hygiene). The last part of the introduction talk was a list of ideas for future research.

3.16 Phishing Susceptibility as a Function of Age, Gender, Weapon of Influence, and Life Domain

Daniela Oliveira (University of Florida – Gainesville, US)


License  Creative Commons BY 3.0 Unported license
© Daniela Oliveira

Phishing is key in many cyber attacks. Successful emails employ psychological weapons of influence and relevant life domains. I discussed my research on phishing susceptibility as a function of Internet user age (old vs young), weapon of influence, and life domain. I presented results from a 21-day study conducted with 158 participants (younger and older Internet users). Data collection took place at the participants' homes to increase ecological validity.

Our results show that older women were the most vulnerable group to phishing attacks. While younger adults were most susceptible to scarcity, older adults were most susceptible to reciprocity. Further, there was a discrepancy, particularly among older users, between self-reported susceptibility awareness and their behavior during the intervention. Our results show the need for demographic personalization for warnings, training and educational tools in targeting the specifics of the older adult population

3.17 Cyber Deception and Cyber Aggression

Simon Parkin (University College London, GB)

License  Creative Commons BY 3.0 Unported license
© Simon Parkin

In this talk I discuss two domains of research. Regarding cyber deception, I focus on cyber-enabled fraud and its impact on smaller charities and businesses; organisations such as these may not have sophisticated cybersecurity capabilities to defend from cyber-enabled fraud. I speculate that we may be able to develop capabilities to support these kinds of organisations to assess trustworthiness, and to assess online indicators of trust (and mistrust), which is critical given the importance of trust to how charities and businesses operate online and in electronic communications. Regarding cyber aggression, I highlight challenges in mitigating technology-enabled domestic abuse and violence ('tech-abuse'). Consumer devices may be used to coerce, monitor, or control another person in a shared environment, potentially using standard device features. The capabilities of emerging Internet-of-Things (IoT) devices may have implications for those impacted by interpersonal abuse, as devices such as 'smart' locks and thermostats may be manipulated. This raises questions as to where technology can, and cannot, address related harms of abuse, but also whether there are opportunities for technology to better support those who are able to leave an abusive situation.

3.18 Get to know your geek: towards a sociological understanding of incentives to develop privacy-friendly free and open source software


Stefan Schiffner (University of Luxembourg, LU)

License  Creative Commons BY 3.0 Unported license
© Stefan Schiffner

Overall, we observe a political will that resulted in legislation that mandates developers to provide privacy friendly and secure software. Moreover, when directly asked, software developers do claim that they want to provide secure products. However, privacy incidents are still on the rise and often criminals abuse insecure implementations for their gain. We road map research for a better understanding of software developers motivations and how to create more effective legal incentives for more secure software. For now, we sketched game theoretical model. In a next step we will obtain data through qualitative and quantitative research in FOSS (free and open source software) developer community. This collected data will be used to develop an objective function for a social game. We will use these games to further analyze the current situation in the field of FOSS wrt privacy features. Lastly we will use our findings to propose changes in policy and best practice.

3.19 Characterizing Disturbing and Reactionary Content in Youtube


Michael Sirivianos (Cyprus University of Technology – Lemesos, CY)

License  Creative Commons BY 3.0 Unported license
© Michael Sirivianos

Social networking services have been affected by disinformation, manipulation, and inappropriate content. One of the most popular OSN platforms is Youtube, where a large number of the most-subscribed channels target children of a very young age. While much of this content is age-appropriate, there is also an alarming amount of inappropriate material available. However, Youtube's algorithmic recommendation engine raises many questions related to the "rabbit hole effect", "echo chambers", and other issues. Furthermore, extremists participate extensively in social networks, expressing their aggressive contents and beliefs. They have an outsized impact in communities, campaigns, and political events. For example, Incels have emerged as one of the most influential extremist communities. They define themselves as unable to find a romantic or sexual partner despite desiring one. While being manifestly sexist, their ideology combines various racist and reactionary elements. They express their hate through forums and mainly on videos especially on Youtube. Sovereign citizens are another group of extremists. Any law of the state is rejected by them, they protest taxation, and in the most extreme case, they act violently, usually against government officials. Alarming, pedophiles also form communities around YouTube videos. As these problems persist and grow in size, states are called upon to regulate content moderation in social networks.

3.20 Characterization, Detection and Mitigation of Antisocial Behaviour

Ivan Srba (STU – Bratislava, SK)

License  Creative Commons BY 3.0 Unported license
© Ivan Srba

Growing negative consequences of online antisocial behavior in social media (e.g., fake news, rumours, hating, trolling) have recently elicited many research efforts, aimed at characterization, detection as well as mitigating of this undesired behavior. In our projects REBELION (<https://rebelion.fiit.stuba.sk/>) and MISDEED (<https://misdeed.fiit.stuba.sk/>), we aim to solve a part of open problems related to online antisocial behavior, which persist despite a large body of already existing research. In particular, the main research challenges, that we are addressing, are: 1) a large amount of unlabeled and dynamic data (the existing datasets are static and either too small or labelled by very simplified heuristics), 2) a more extensive utilization of data about content, users and context (the existing methods do not take advantage of the whole spectrum of available data, such as multiple modalities, data from multiple platforms), and 3) a proposal of new mitigation approaches (there is a need for early detection and more extensive involvement of users). In order to obtain suitable data needed to address these research challenges, we proposed and developed a unique platform for monitoring antisocial behavior called Monant [1]. It consists of several modules for web monitoring, integration of various AI methods, platform management as well as a module for providing results to end users (public and experts). In order to evaluate this platform, we conducted a case study in which we monitored 29 unreliable medical news sites and blogs. We obtained about 58 thousand news articles, which we mapped to 131


cancer “treatments” (adapted from the list provided in [2]) which have not been proven to actually cure the patient. A case study revealed us how many articles share the most frequent misinformative treatments and the time evolution of their spreading. In our future work, we plan to work on additional development of Monant platform, gathering a more extensive dataset of medical misinformation, labelling the dataset by a claim presence and stance detection, developing detection methods for various types of antisocial behavior, which will take advantage of feature-rich data provide by the dataset, and finally we will investigate new mitigation strategies, which will be deployed in Monant end-user applications.

References

- 1 Ivan Srba, Robert Moro, Jakub Simko, Jakub Sevcich, Daniela Chuda, Pavol Navrat, Maria Bielikova. Monant: Universal and Extensible Platform for Monitoring, Detection and Mitigation of Antisocial Behaviour. Workshop on Reducing Online Misinformation Exposure – ROME 2019. July 25, 2019, Paris, France.
- 2 Amira Ghenai, Yelena Mejova. Fake Cures: User-centric Modeling of Health Misinformation in Social Media. ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW), 2018.

3.21 Measuring and Modeling the Online Information Ecosystem

Gianluca Stringhini (Boston University, US)

License  Creative Commons BY 3.0 Unported license
© Gianluca Stringhini

The online information ecosystem is complex, with users using multiple online services at the same time, each with its own characteristics. To properly study how malicious activity unfolds on the Web, we need tools that enable us to collect data from these services at scale and enable us to get a comprehensive view of the activity happening on them. To this end, together with my group I developed a number of techniques that enable us to collect data about malicious online activities. Such techniques include developing account honeypots (e.g., on Gmail) and leaking credentials to them so that we can observe how criminals interact with them [1], setting up crawlers for online services, and leveraging social network APIs to collect data in real time [2]. I have then used this data to better understand several types of malicious activity, from cyberbullying [3] to disinformation [4]. Studying these phenomena presents a number of challenges. First, human driven malicious activity (for example cyberbullying) tends to be more nuanced and context dependent than automated one (for example spam), and therefore develop systems to automatically detect it is more challenging. To address this challenge, in my research I apply a mixed method approach in which human annotators label content that is later processed by machine learning techniques [5]. Second, online information is not only conveyed through text, but also through images and videos. To take this into account, in my research I apply image processing techniques to understand how images are used to spread hateful content online [6, 7]. Finally, online services do not operate in a vacuum but information from one service is shared on and can influence other services. To address this challenge, in my research I develop methods to keep track of influence between different online services (e.g., Hawkes Processes) [4, 6].

References

- 1 Onaolapo, J., Mariconti, E., Stringhini, G., What Happens after you are Pwnd: Understanding the Use of Leaked Webmail Credentials in the Wild, ACM SIGCOMM Internet Measurement Conference, 2016.
- 2 Hine, G., Onaolapo, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Samaras, R., Stringhini, G., Blackburn, J., Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan’s Politically Incorrect Forum and Its Effects on the Web, AAAI International Conference on Web and Social Media, 2017.
- 3 Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., Vakali, A., Mean Birds: Detecting Aggression and Bullying on Twitter, ACM Web Science Conference, 2017.
- 4 Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Sirivianos, M., Stringhini, G., Blackburn, J., The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources, ACM SIGCOMM Internet Measurement Conference, 2017.
- 5 Founta, A., Djouvas, C., Chatzakou, D., Leontiadis, I., Blackburn, J., Stringhini, G., Vakali, A., Sirivianos, M., Kourtellis, N., Large Scale Crowdsourcing and Characterization of Twitter Abusive Behavior, AAAI International Conference on Web and Social Media, 2018.
- 6 Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., Blackburn, J., G. Suarez-Tangil, On the Origins of Memes by Means of Fringe Web Communities, ACM SIGCOMM Internet Measurement Conference, 2018.
- 7 Mariconti, E., Suarez-Tangil, G., Blackburn, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Luque Serrano, J., Stringhini, G., “You Know What to Do”: Proactive Detection of YouTube Videos Targeted by Coordinated Hate Attacks , ACM Conference on Computer-Supported Cooperative Work and Social Computing, 2019.

3.22 Disinformation as a Political Game

Gareth Tyson (Queen Mary University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Gareth Tyson

The presentation explored the role of political influence and decision making within the regulation of social media companies. We defined politics as the activities associated with the governance of a country or area, especially the debate between parties having power. This provided an underpinning for exploring how disinformation, and its subsequent regulation, can be best modelled as a political game. In most cases, we found that social media companies shy away from public power, distancing themselves from the responsibilities that it entails. In sum, this leads to a lack of accountability and problems in defining liability for harms derived from disinformation. The presentation concluded with two open-ended questions: 1) who should be given the power decide what misinformation is? and 2) what methods to enforce those decisions should be given?

3.23 Language-based deception detection

Sophie van Der Zee (Erasmus University – Rotterdam, NL)

License © Creative Commons BY 3.0 Unported license
© Sophie van Der Zee

Language use is affected by deception. For example, when lying, people distance themselves more by using more third person pronouns. Usually, this type of research is done on single statements made by many individuals. This time, we analyzed many statements made by one single individual: The US President. Thanks to the fact-checking efforts from the Washington Post, for the first time in history, there are enough fact-checked incorrect statements made by one individual to create a personalised model of deception. We collected 3 months of tweets by @realDonaldTrump, and connected this datafile to the fact-checked database by the Washington Post. We compared language use with LIWC software between factually correct and incorrect tweets. If the US President was aware of the incorrectness of his statements at the moment of sending, one would expect language difference between correct and incorrect tweets in line with the deception literature (deception hypothesis). If the US President was unaware of the incorrectness of his tweets at the moment of sending, little language differences between factually correct and incorrect tweets are expected (misinformation hypothesis). Results showed that almost half of the LIWC word categories differed between his correct and incorrect tweets, suggesting the US President is often aware that his factually correct and incorrect messages are different at the moment of sending, supporting the deception hypothesis. Next, we estimated a logit model to test how well we could predict whether a tweet was factually correct or incorrect based solely on word use. We collected a second dataset, again comprised of three months of tweets by the US President. Both within- and out-of-sample testing results led to a prediction overall accuracy of 73%. In other words, we can correctly predict for 3 out of 4 tweets by the current US President whether it is factually correct or incorrect solely based on word use.

3.24 Applying Routine Activity Theory to Cybervictimization: A Theoretical and Empirical Approach

Sebastian Wachs (Universität Potsdam, DE)

License © Creative Commons BY 3.0 Unported license
© Sebastian Wachs

In my presentation, I proposed the Routine Activity Theory (RAT) as a theoretical framework for cybervictimization among adolescents. RAT has been developed by Cohen and Felson and aims to describe conditions that are favorable for crime [1]. According to the RAT, the following three essential elements must converge for a crime to occur [1]: A likely offender, absence of capable guardians, and a suitable target. I also presented briefly current analyses in which I tested the RAT empirically. In this study, I analyzed whether parental mediation of internet use (absence of capable guardians) is directly as well as indirectly via online disclosure (suitable target) associated with cybergrooming victimization. The sample consisted of self-reports from 5,938 adolescents from six countries ranging in age from 12 to 18 ($M=14.77$, $SD=1.60$). Applying mediation test using the structural equation modeling framework I found that parental mediation, online disclosure and cybergrooming victimization are directly associated. While instructive parental mediation is negatively related with online disclosure

and cybergrooming victimization, restrictive mediation is positively related to both. In addition, online disclosure partially mediates the relationship between parental mediation and cybergrooming victimization. While this analysis confirms the general usefulness of applying the RAT to cybergrooming the findings also highlight the need to educate parents to use certain strategies of mediation and inform adolescents to avoid disclosing online too much private information in the course of prevention programs.

References

- 1 Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, (44), 588–608.

3.25 Research in Evidence-based Security

Victoria Wang (University of Portsmouth, GB)

License  Creative Commons BY 3.0 Unported license
© Victoria Wang

I particularly enjoy applying scientific rigour and academic analysis to real world situations to obtain evidence-based solutions. My current research ranges over cyber/information security, surveillance studies, social theory, technological developments and online research methods. My latest research projects involve: i) data release and its related issues of trust, privacy and security [1, 2]; ii) security threats and management measures in organisations [10]; iii) formal methods for monitoring, data collection and interventions [6]; iv) a general formal theory of digital identity and surveillance [5]; v) developing new techno-social theories such as ‘Phatic Technologies’ as conceptual tools to understand cyberspace and its security issues [6, 7]; vi) cybercrime and threats in various countries, e.g., Nigeria, and various networks, e.g., the Darknet [3]; and vii) cyberbullying [4, 8]. My future research interests include – developing cyber security solutions for critical infrastructure, and developing my Phatic Technology Theory for applications in marginalised urban societies.

References

- 1 Wang, V. & Shepherd, D., Exploring the extent of openness of open government data – A critique of open government datasets in the UK (accepted – Government information quarterly)
- 2 Wang, V., Shepherd, D. & Button, M., The Challenges of Opening Government Data in the UK – A View from the Bottom. *Information Polity*, 24(1), 2019: 59-74.
- 3 Mirea, M., Wang, V. & Jung, J., The not so Dark Side of the Darknet – A Qualitative Study. *Security Journal*, 32(2), 2019: 102-118.
- 4 Edwards, S. & Wang, V., There are Two Sides to Every Story – Young People’s Perspective of Relationship Issue, *Journal of Youth Studies*, 21(6), 2018: 717-732.
- 5 Wang, V. & Tucker, J.V., Surveillance and Identity: Conceptual Framework and Formal Model. *Journal of Cybersecurity*, *Cybersecurity*, 3(3), 2017: 145-158.
- 6 Johnson, K., Tucker, J.V. & Wang, V., Theorising Monitoring: Algebraic Models of Web Monitoring in Organisations. In P. James, M. Roggenbach (eds), *Recent Trends in Algebraic Development Techniques*, 23rd International Workshop, WADT 2016, Revised Selected Papers, *Lecture Notes in Computer Science (LNCS)*, Springer, 10644, 2017: 13-35.
- 7 Wang, V. & Tucker, J.V., Phatic Systems in Digital Society, *Technology in Society*, 46, 2016: 140-148.

- 8 Wang, V. & Edwards, S., Strangers are Friends I haven't Met Yet: A Positive Approach to Young People's Use of Social Media, *Journal of Youth Studies*, 19(9), 2016: 1204-1219.
- 9 Wang, V., Tucker, J.V. & Haines, K., Phatic Technologies in Modern Society, *Technology in Society*, 34 (1), 2012: 84-93.
- 10 Cyber Security Breaches Survey (2016-2019), Commissioned by HM Government (Department for Business, Innovation & Skills), Ipsos MORI.

3.26 Deception and deterrence

Jeff Yan (*Linköping University, SE*)

License © Creative Commons BY 3.0 Unported license
© Jeff Yan

What I have looked into include deception, social engineering, cybercrime and usable security, and we're interested in both technical and sociotechnical aspects. The project on "Deterrence of deception in sociotechnical systems", funded by EPSRC, enabled some exciting research and interaction with brilliant minds including Ross Anderson, Nick Humphrey, Aldert Vrij, Jeff Hancock, Jussi Palomäki and Sophie van der Zee. One of the innovations was a naturalistic behavioural study of Machiavellian individuals on strategic deception. Inspired by Oxford research on the Sicilian mafia, my recent cybercrime study examined the phenomenon of 'scam villages' in China from an economics perspective. My earlier research studied cheating in online games. Research questions which I am curious about and would like to get inspiration for in this week are abundant, for example:

- Deception deterrence: which context, and how?
- Cheat & show-off, or cheat but hide? Is Bernard Madoff the exception, or the norm? Any theory, in psychology, criminology or whatever, explaining either way?
- What research will both CS and social scientists like?
- What is the next big question?

4 Working groups

4.1 Theme 1: Attacker Modeling Group

Abhishta Abhishta (University of Twente, NL), Zinaida Benenson (Universität Erlangen-Nürnberg, DE), Matt Bishop (University of California – Davis, US), Joe Calandrino (Federal Trade Commission – Washington, US), Natalie Ebner (University of Florida – Gainesville, US), Manuel Egele (Boston University, US), William Robertson (Northeastern University – Boston, US), Victoria Wang (University of Portsmouth, GB), and Savvas Zannettou (Cyprus University of Technology – Lemesos, CY)

License © Creative Commons BY 3.0 Unported license
© Abhishta Abhishta, Zinaida Benenson, Matt Bishop, Joe Calandrino, Natalie Ebner, Manuel Egele, William Robertson, Victoria Wang, and Savvas Zannettou

The desired outcome of the group was to determine how to develop one or more probabilistic models that will predict what attackers will do next, or augment the defenses to slow down the attacker, or speed up the defense to handle the attacker better.

The group decided to focus on organizations, because they have some sort of a management plan, giving them coherence and one or more general purposes; they also have different, often complex, technological structures. Although the majority of group members were technical, the group included a criminologist and a psychologist. The group realized that any model developed had to include non-technical factors.

There are a number of ways to develop such probabilistic models. The first is to design a model based on expertise and experience, and then use data to test its accuracy. The alternate approach is to reverse this: gather real world data and develop a model based on that. In this latter case, the model would then be tested against out of sample data. The data will consist of data from attacks, data from defenses is important here, because that data provides both contextual information about the environment, i.e., the organisations involved, and the attack itself, as well as the policies and procedures the defenders use to contain (e.g., minimising its potential damage) or thwart the attack. The procedures here will be those that are used in practice, not simply the ones written in guidelines that the security management (both technical and human) personnel and users are supposed to follow.

This leads to the first step: obtaining real world data required to build such a model. It is unclear at this point what attributes the data must have, and indeed what the data itself must consist of, so an appropriate approach is to see what data is available now, what it consists of and what attributes it has. As the model is developed and refined, aspects of the data and attributes that are missing and yet are necessary for the model to predict effectively will become clear. Also, techniques for obtaining the data are essential, because while much data has been gathered, very little of it is widely available, or indeed available except under the most stringent conditions. In short, even if such data is available, getting access to the data is yet another difficult step. For example, organisations might not want to admit that they have been victimised by cyber attackers. Even if they openly admit to victimisation, they might not be willing to share their log files and other internal documents recording the attacks with researchers. In fact, based on our previous experience, this is rather common, especially within the financial and insurance industries, wherein peer competitions are intense. Thus, an open question is how to relax these constraints while providing the guarantees that the possessors of the data will require in order to share it. This ties into the ethics of gathering data, which vary among legal jurisdictions and types of organizations. For example, in the United States, public institutions must comply with one set of government rules regarding protection of personally identifiable information, whereas private entities comply with a different (but overlapping) set of rules. For another example, the introduction of the GDPR (2018) in Europe might, on the one hand, mean that organisations are under more pressure to share their data; whereas on the other hand, they might become even more cautious in sharing data with researchers.

The data will come from several sources. Technical data will come from places such as logs, network traces, and network- and host-based data; it will include contextual information such as metadata, the organization where it is gathered from and that generated it (which may be different organizations), and the location of the data and its use (for example, if it is stored in a cloud, or stored in encrypted form locally or in a cloud, and whether the computations are done locally or in the cloud, and so forth). Red teaming, also known as penetration testing, will also be a valuable source of data. Less technically complete data will inform motivations, external characteristics of the attack, and other human and organizational aspects of the data. News stories will be a good source of this type of data, as well as law enforcement reports, government analyses, and court records. Relevant questions here relate to the broader picture of attacks. How do attackers advertise their wares? What

are their wares – what tools and methodologies do they use, and do they share or sell these? Further, a series of empirical work might be conducted to gather data from employees of selected organisations, via common social science research methods, such as questionnaires, interviews, and focus groups. For example, we could simply ask employees of an organisation what they think they did right to minimise any possible damage of a recently experienced cyberattack. Here, relevant questions might be: what was their first response? Did the organisation have a Chief Information Security Officer who discover the attack and respond to it very quickly?

The group noted that, in addition to conventional cybersecurity attacks, the above may apply to the dissemination of fake news (defined as news that contains information that is verifiably untrue). The basis for this belief is that Facebook, Twitter, and other social media can be considered large-scale distributed logs, and the organization these logs apply to is the society involved.

This led to a discussion of high-level considerations. Attackers may have many goals, such as getting money, embarrassing someone or some entity, obtaining control of a system (technical or non-technical, such as a political organization) to change things (such as the politics of a society, possibly by the use of fake news), and many other goals. The group agreed to focus on financial institutions to keep the work manageable. Two SWOT (Strengths, Weaknesses, Opportunities, and Threats) analyses examined both the financial institutions (specifically, banks) and the attackers. Tables 1 and 2 summarize the results of these analyses.

From this, the group began work on the model, a preliminary version of which follows. This starting point is definitely not complete. New features will be added, and some of the (existing and new) features will be empty for a given instantiation. Hence, the reader should view what follows as an outline.

The model is based upon goals, which include interrupting services, public shaming, obtaining money or denying others money, obtaining various types of power (social/cultural, political/ideological, economic, and so forth) or denying these to others, gathering information, and other possible goals. These are more detailed than the goals outlined above, and are consonant with them.

The structure of the model consists of 9 basic features:

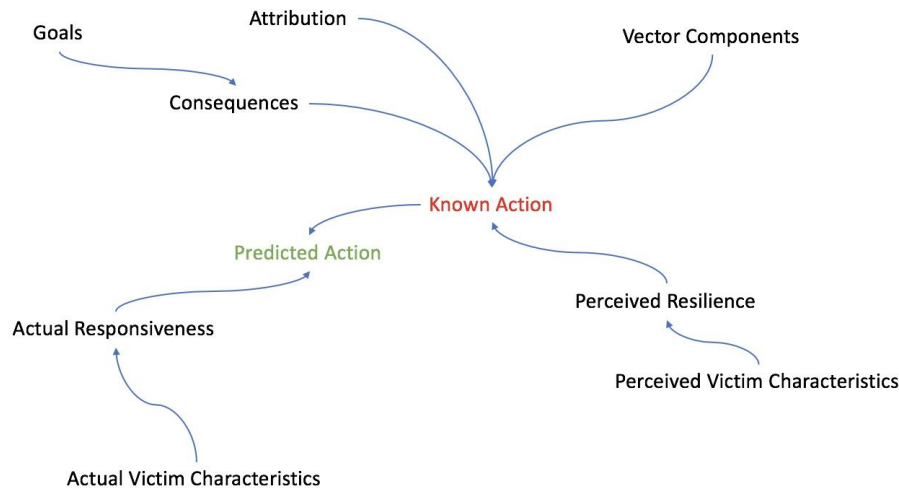
1. Goals
2. Vector components
 - a. How long does the attack take; when does it occur
 - b. Complexity of the attack (technical, non-technical, etc.)
 - c. Technological tools employed
 - d. Access (direct or indirect; social engineering, vulnerability scanning, etc.)
 - e. Communication (density, patterns, etc.)
3. Consequences (intended/unintended; who is harmed, who benefits)
4. Authorisation (authorised/unauthorised; for the latter, open or hidden)
5. Attribution (full, none, false, random)
6. Expected resilience of victim from attacker's point of view
7. Expected characteristics of victim from attacker's point of view
 - a. Location, relationships
 - b. Spread; how large is the attack surface?
 - c. Infrastructure
8. Actual responsiveness of victim
9. Actual characteristics of victim
 - a. Location, relationships
 - b. Spread; how large is the attack surface?
 - c. Infrastructure

■ **Table 1** SWOT table for financial organizations.

<i>Strengths</i> <ul style="list-style-type: none"> ■ High financial resources ■ Historically motivated to invest in security 	<i>Weaknesses</i> <ul style="list-style-type: none"> ■ Focus on financials ■ Reliance on 3rd party software ■ Legacy systems ■ Highly distributed systems
<i>Opportunities</i> <ul style="list-style-type: none"> ■ Sharing of information ■ Availability of finances ■ Substantial political capital 	<i>Threats</i> <ul style="list-style-type: none"> ■ Availability attacks on distributed systems ■ Legacy systems breaking down or compromised ■ Insider attacks (leaking of information on high profile clients) ■ Unauthorized transfers ■ Privacy issues (personal information of clients)

■ **Table 2** SWOT table for attackers of financial organizations.

<i>Strengths</i> <ul style="list-style-type: none"> ■ Force useless investment ■ Availability of cybercrime as a service 	<i>Weaknesses</i> <ul style="list-style-type: none"> ■ High resources and background information required ■ Conversion to hard cash ■ Information asymmetry
<i>Opportunities</i> <ul style="list-style-type: none"> ■ High value data ■ High value money ■ Reliance on implicit trust ■ Attack clients of the bank 	<i>Threats</i> <ul style="list-style-type: none"> ■ Getting caught (for example, when converting the electronic cash to physical cash) ■ Reputational damage to the attacker



■ **Figure 1** Relationship of the components of the structures.

Figure 1 summarizes their relationships. The model uses those structural features that drive the known action (features with paths to the red “Known Action”) to combine with the structural features that enable future actions to be predicted (features with paths to the green “Predicted Action”).

In more formal terms, a known action A_1 (which is a function of features 2-7) leads to a set of probable actions A_{21}, \dots, A_{2n} (which are a result of A_1 and feature 8). To determine the best response strategy, the net payoffs of each need to be computed. The characteristics of the victim (feature 9) drive a penalty, so the calculation must include a DB (for “DisBenefit”) component. Let $P(A_j)$ be the payoff for action A_j . From a purely rational point of view, the next action of the attacker should maximise the net payoff. Then the most likely predicted action is the one maximizing $P(A_1) + P(A_{2k}) - DB$, over k (see Figure 2). How to calculate these payoffs is left for future work. Note the assumption here is that the attacker is following some sort of rational plan; if the attack is a sequence of random actions, the underlying assumption does not hold.

The group then used two case studies to begin validating the model. At least one member of the group worked on each of the incidents using the case studies. Tables 3, 4, and 5 summarize the application of the models to the case studies. Table 3 is the characterization of the Internet Worm of 1988; Tables 4 and 5 are the first and second steps of the SpamHaus attack.

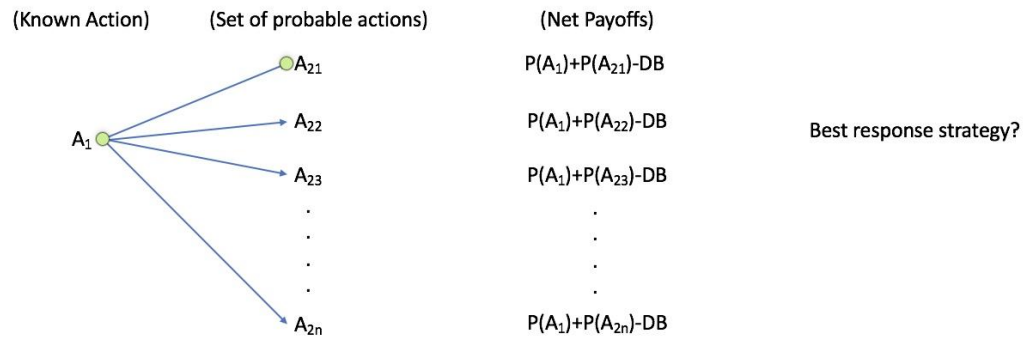
Future work will sharpen the model and make it useful. The key opportunities for improving it are:

1. Convert the variables into measurable quantities

First, data must be found to see if the overall structure of the model works. This data can be used to determine how to measure the attributes. Undoubtedly, some will remain qualitative, and others quantitative; but the values for both types will be refined as data emerges. This will also lead to a refinement of the definitions of the variables.

2. How to obtain labeled data

Obtaining data properly labeled as attack data (as opposed to data that is unlabeled) is critical, and methods to do this must be investigated. Several possibilities were discussed,



■ **Figure 2** Graphical representation of probable actions and net payoffs.

■ **Table 3** Validating the model with the Internet Worm of 1988.

Motive	social/activism
Time	3pm-midnight
Complexity	High (considering historical context)
Technical tools	Reused software from a (suspected) re-research laboratory
Access	Direct access (to MIT public access node)
Communication	N/A (only one attacker)
Consequences	Targeted hosts on network were unusable
Authorization	Unauthorized, intended to be open
Attribution	Fully attributed
Attacker knowledge of responsiveness	None; security community nascent
Perceived victim characterization	Access to ARPANET/ARPANET/Suns and VAXen
Actual victim characterization	Access to ARPANET/ARPANET/All systems with access to ARPANET; only Suns, VAXen taken down

among them providing “data bounties” (much like “bug bounties”) and developing similar incentive structures for encouraging the sharing of data. Threat feeds may be a fertile source, as will interviews with CISOs, incident responders, and other security operations personnel. An alternative is to use the “over the fence” approach. In this approach, others take the model, instantiate it with attack data they can’t share, and give results, including problems with the model, false positives, and false negatives. This will also allow the model to be instantiated with data at different levels of coarseness.

Several open questions remain:

- Are there any higher order attacks or dimensions of attacks we are missing?
- The variables are not orthogonal – is this a problem?
- How do we handle noise in the measurements?
- How do we handle noise the attacker injects?
- How do we handle false positives/negatives? And equally if not more critical: how do we identify them?

■ **Table 4** Validating the model with the SpamHaus attack (step 1).

Motive	revenge
Time	6pm
Complexity	Low
Technical tools	Off-the-shelf tools
Access	None
Communication	IRC (after the attack)
Consequences	None
Authorization	No
Attribution	Random
Attacker knowledge of responsiveness	Low
Perceived victim characterization	SpamHaus/Not distributed/Server
Actual victim characterization	London Exchange + CloudFlare + SpamHaus/Widely distributed/CDN

The group suggested possible next steps. A workshop on defenders and attackers would provide additional insights and understanding. Such a workshop should include CISOs, security operations personnel, and others who defend systems, as well as former attackers who “came over to the light side”. Obtaining funding for this work is critical, and there was considerable discussion about what groups or agencies might fund this international collaboration. A paper on our conceptual model of cyberattacks would be a good starting point for such requests. Possible appropriate venues would be WOOT, IFIP SEC, and the economics workshop WEIS.

4.2 Theme 2: Unexpected Consequences of Countermeasures

Matthew Edwards (University of Bristol, GB), Yi Ting Chua (University of Cambridge, GB), Alice Hutchings (University of Cambridge, GB), Daniela Oliveira (University of Florida – Gainesville, US), Simon Parkin (University College London, GB), Stefan Schiffner (University of Luxembourg, LU), and Gareth Tyson (Queen Mary University of London, GB)

License © Creative Commons BY 3.0 Unported license

© Matthew Edwards, Yi Ting Chua, Alice Hutchings, Daniela Oliveira, Simon Parkin, Stefan Schiffner, and Gareth Tyson

Overview

We tackled the topic of countermeasures enacted in cybersafety and cybercrime often leading to unintended consequences and harm. This problem arises for both technical solutions (classifiers, website takedowns) and administrative solutions (staff training, public advice, policies enacted by staff). We developed a taxonomy of unintended consequences, and transformed this into a set of questions which could be asked of any countermeasure, so that potential consequences might be anticipated and mitigated.

■ **Table 5** Validating the model with the SpamHaus attack (step 2).

Motive	revenge
Time	midnight
Complexity	Low
Technical tools	Off-the-shelf tools
Access	None
Communication	IRC (after the attack)
Consequences	Drop CloudFlare from London Exchange; SpamHaus no longer reachable
Authorization	No
Attribution	Full
Attacker knowledge of responsiveness	Low
Perceived victim char- acterization	London Exchange + SpamHaus/Not dis- tributed/Server
Actual victim charac- terization	London Exchange + CloudFlare + Spam- Haus/Widely distributed/CDN

Section 1: Scenarios

The group approached the problem by first defining a set of cybersafety scenarios as motivating examples, then identifying countermeasures which may be applied to these scenarios. These countermeasures were then used as grounded prompts for consideration of unintended consequences.

1. Intimate partner abuse¹: Bob and Charlie live together. Bob is controlling and monitors Charlie's behaviour using IoT devices. This includes Charlie's smartphone. When suspecting Charlie might be visiting his friends, Bob goes onto Twitter and shares aggressive and fabricated posts.

Countermeasures & Consequences:

- Take away Charlie's tech so Bob cannot use it to harm them. Replace all of Bob's accounts with new ones.
 - Loss of personal information
 - Financial cost
 - Loss of abilities provided by tech (to stay in contact with family and friends)
- Provide training resources for Charlie so they know how to identify and prevent this abuse.
 - Bob might find this advice and become more violent
 - Bob might use this advice to become more stealthy and effective in abuse of Charlie
- Recover and reset devices – as the UK government suggests
 - Loss of personal information
 - Loss of social support structures
- In cases where intimate content is shared – contact social media company, take down material

¹ Lopez-Neira, Isabel, et al. “‘Internet of Things’: how abuse is getting smarter.”, Safe –The Domestic Abuse Quarterly, (63), 22-26. Women's Aid (UK), 2019.

- Takedown mechanism might be misused to implicate innocent users
- Verifying identity might be embarrassing and/or difficult
- Streisand effect – content could become more popular
- Images might instead be shared on platforms where more harm to the victim might originate
- Legal actions – criminal prosecutions
 - Slow pace of justice system, stress
 - Risk of escalation before
- Revenge porn – facebook asks you to upload images in advance
 - Verified connection between image and your identity – future misuse
 - Normalises sharing

2. Disinformation: There is a political campaign, Charlie vs. Bob. A third party, who supports Bob, performs a concerted misinformation campaign to spread false information about Charlie. This is done predominantly via Facebook and Twitter, initiated via a network of social media bots who inject the material.

Countermeasures & Consequences:

- Remove tweets/posts
 - Backlash – spread more often in defiance
 - Takedown used as evidence of conspiracy to suppress ‘truth’
- Remove bots
 - Misclassification, irritation of innocent users
- Removing accounts
 - people move onto Gab and intensify
- Detect collusion in social graph
- Build machine learning model to identify ‘fake news’
 - Leads to complacency, reduction in skepticism
 - Misclassification
- Using fact checkers to highlight fake news
 - Costly to fact-check material
 - Complacency, trusting fact-checker for truth
- Reduce visibility of material considered to be fake news
 - Evidence of ‘suppressing truth’
 - Misclassification, innocent users don’t necessarily know what’s wrong
- Limited number of shares/forwards
 - Limit also applies to legitimate content
- Block entire service
- Promoting correct information

3. CEO Fraud: Bob finds out the name and details of the Footbook’s CEO. Bob emails one of Footbook’s employees, Charlie, asking him to pay a last minute invoice because Bob forgot. Charlie goes ahead and pays the invoice, which transfers money into an off-shore account. Charlie gets sacked.

Countermeasure:

- Change the culture of the company – CEOs can’t send random emails
 - Productivity costs, conflict
- Training

- Additional cost for the low-level employee
- Security best practice, least privilege
- Authentication required for bank transfers/third party checks on all transactions
 - Productivity costs
- Crypto check the sender
- Remove steps from email, and required in-built finance system
 - Implementation costs
- Remove domain squatting
- Automated attacks – looking for anomalous behaviour in transactions
 - Misclassification of important transactions
- Restriction of access to external sites/public email services
- If data leak (e.g. IP theft) could watermark files
 - Company use this to identify whistleblowers

4. Phishing: Bob has recently lost his job, and holds bitter resentment towards his former employer. He believes there has been a conspiracy against him, driven by mistrust of his Northern accent. He therefore formulates a phishing campaign against the HR department of his former employer. Charlie receives an email from Bob, masquerading as a notification of a company award worth £18. Charlie clicks on the link, and is asked to enter his credentials. The website, operated by Bob, is then used by him to retrieve HR data related to his dismissal. Bob was sacked because of his aggressive and inappropriate behaviour in the company toilets.

Countermeasure:

- Training & education (including phishing exercises used as training)
 - Creates a false sense of understanding the problem
 - Allows attackers to adapt to the training
 - Results in victim blaming
 - Might upset people – make them feel stupid
 - Might not help all users (e.g. ones who don't engage with training), but company might then think that the problem is solved
- Email filtering, e.g. using machine learning
 - Misclassified email goes to spam, holds up work
- Website takedown and ISP blocking of websites
 - Website takedown mechanism could be abused to take down legit sites
 - Streisand effect
 - Site might move to more resistant providers
- Website verification
 - Sense of security from verification could be misleading about behaviour
- Safe links

5. Dating Fraud: Bob is innocently swiping on Tinder. He encounters a handsome young woman, Charlie. Bob and Charlie hit it off, and instantly begin to plan their life together. Unfortunately, Charlie lives in Peru and cannot afford to travel to Dagstuhl. After a few weeks of intimate conversation, Charlie requests \$3000 to enable her to book a flight. Once the money has been transferred, Bob never hears from Charlie again.

Countermeasures:

- Get off Tinder
 - No hookups

- Verify accounts
 - Some apps force by providing link to facebook account
 - Might not want to share that information, i.e. privacy invasive
 - People with non-traditional sexual interests have them exposed
 - Might expose people to financial fraud if required to upload credit card details
- Close fraudulent accounts
 - False positives, e.g. person who is very popular
 - People may have had photos stolen from them, and used by fraudsters.
 - Countermeasures often involve collecting more data – data leaks have a greater impact
 - Might be cultural sensitivities that must be catered for, e.g. Tinder vs Grindr
- Advice, tips and prompts (targeted)
 - Annoying for users
- Training
- Waste the time of suspected scammer
 - Wastes timewaster's time/resources
 - Extended contact raises potential for more harm
 - Could provoke e.g. violence

Section 2: Taxonomy & Questions

Working from the list of consequences from countermeasures in each of these scenarios, the group categorised common types of consequence, and then reformulated the taxonomy as a number of questions which should be asked of any proposed countermeasure.

Unintended consequence taxonomy:

Additional Costs: Implementing countermeasures can pose a burden for different stakeholders involved. Training and policy exhaust employee compliance budget², restrictive security controls can hamper business productivity³, staffed reporting systems must be paid for by a social media platform.

Misuse of Countermeasure: The countermeasure itself might be misused by malicious actors to cause harm. Reporting systems can be misused to target competitors for takedown; advice for victims can be used by perpetrators to improve their misbehaviour; abusers can train against classifiers to learn how to go undetected.

False Positives: Incorrect decisions made by/as a result of the countermeasure can cause harm to innocents. Classifiers can misidentify content or users as malicious or deceptive; verification schemes can exclude people legitimately unable to verify their identity.

Displacement: The countermeasure might simply move harm to other targets. Removing extremist accounts pushes them to echo chambers where their views might be reinforced; stricter or more arcane policies may simply cause employees to circumvent policy and suffer all the blame for resulting failures.

Amplification: The countermeasure might actually cause an increase in the behaviour it intended to prevent. A plethora of fact-checkers leads to fragmentation of trust, attempts to take something down can cause it to gather more attention through controversy, harsh crackdowns can lead to reprisals in defiance.

² Beauteament, Adam, M. Angela Sasse, and Mike Wonham. "The compliance budget: managing security behaviour in organisations." Proceedings of the 2008 New Security Paradigms Workshop. ACM, 2009.

³ Kirlappos, Iacovos, Simon Parkin, and M. Angela Sasse. "Learning from "Shadow Security": Why understanding non-compliance provides the basis for effective security." Proceedings of the 2014 Workshop on Usable Security (USEC). Internet Society, 2014.

Insecure Norms: The countermeasure might promote the adoption of insecure norms. Highly-trusted technology or policy can lead to a false sense of security that makes users more susceptible to deception; normalising the sharing of identifying information for verification purposes contributes to phishing success.

Disrupting other Countermeasures: A well-intentioned countermeasure could inadvertently cause problems for another – potentially more effective – countermeasure. Social media sites which remove abusive content are also removing evidence from criminal investigations; requiring users to verify their identity prevents them from using anonymity as a defence; contradictory advice on how to deal with a problem leads to confusion.

Questions

From the above categories of unintended consequence, we extract 6 questions that could be asked of any proposed (or extant) countermeasure to identify potential unintended consequences.

1. In what ways might the countermeasure burden stakeholders?
2. In what ways might the countermeasure be used in attacks?
3. In what ways might the countermeasure displace harm to others?
4. In what ways might the countermeasure amplify harm?
5. In what ways might the countermeasure create insecure norms (e.g. complacency)?
6. In what ways might incorrect classification cause harm?
7. In what ways might the countermeasure disrupt the operation of other countermeasures?

We also identify a cross-cutting concern:

8. Consider for each question which groups are more at risk of experiencing harm.

Section 3: Identifying Further Consequences

We cross-tabulated the above taxonomy with four general categories of countermeasure, to validate the location of specific unintended consequences within the taxonomy, and to make use of the taxonomy to identify new unintended harms in areas our earlier scenarios had not covered (see Table 6).

■ **Table 6** Categories of countermeasures and related unintended harms.

	Categories of Countermeasures			
	Managing content	Verification (controlling users)	Training (changing behaviours)	Takedown (infrastructure)
Displacement	Moves people to echo chambers; Fragmentation	User displacement to less protective platforms	Circumventing work policies	To abuse resistant hosting providers
Insecure Norms	Warnings; Rely on fact-checking; Normalising sharing of explicit images; Preaching to the choir; Groupthink; Non-falsifiability	Normalising sharing personally identifiable information	Makes social engineering problem routine; Desensitization; Habituation; Risk-dumping; Told wrong thing	Assume problems are removed
Additional Costs	Wiping phones; Loss of evidence; Disrupt existing connection	Annoyance / time to verify	Loss of productivity; Adds to compliance budget; Conflict; Chilling effect; Induce mistakes; Victim blaming	Criminal Justice System (slow retaliation); Legitimate sites recovery cost
Misuse	Poisoning fact-checking; Identifying whistleblowers; Sausages identify; Misuse image hashing	Privacy impacts; Misuse by; Data breach; Faking blue ticks / trust seal	Perpetrators learn from advice	Reporting competitors; Censorship
False Positive	Forcing false positives; Cold start problems – new users struggle to gain trust	Users cannot verify identity due to photo stolen	Errors as result of training	Website take-down
Amplification	Fragmentation; Streisand effect	Blue ticks on Twitter	Perpetrator sees advice and escalate	Streisand effect
Disrupting Other Countermeasures	Destroy evidence	Anonymity (e.g. Facebook and phone number)	Contradictory advice	Destroying evidence

Section 4: Directions for Research

Future research on this topic could explore a number of additional directions:

1. Do the devised questions cover enough unintended consequences that they could be used as an instrument in e.g., ethical review of security and cybersafety research proposals concerning countermeasures?
 - What are the limitations of this instrument, and can it be amended to correct for these?

2. How can the likelihood and severity of unintended harms be ascertained?
 - Can anything general be said about the likelihood and severity of the categories of unintended harms, or does this depend too much on the specific countermeasure in question?
 - How can measures of unintended consequences be gathered?
3. Why are unintended harms not already being considered?
 - Is there a facet of decision-making around countermeasures (e.g., lack of incentives) which explains why they are not considered?
 - Are they in fact not considered/seen⁴, or just too difficult to remedy?⁵
4. Are there common mitigations to unintended harms which might complement this taxonomy?
 - Can we produce guidance that allows developing countermeasures to build-in mitigations in a variety of application areas?

4.3 Theme 3: Measuring Human Behavior from Information Security and Societal Perspectives

Ivan Srba (STU – Bratislava, SK), Katsiaryna Labunets (TU Delft, NL), and Sophie van Der Zee (Erasmus University – Rotterdam, NL)

License © Creative Commons BY 3.0 Unported license

© Ivan Srba, Katsiaryna Labunets, and Sophie van Der Zee

Joint work of Ivan Srba, Katsiaryna Labunets, Sophie van der Zee, Jeff Yan, Gabriele Lenzini, Jeremy Epstein, Deanna Caputo, Jean-Willem Bullée, Claude Castelluccia

Introduction. People, organizations, and governments are increasingly using the Internet for a wide range of activities, from socializing to shopping, and working. This increased digitization has brought many benefits, but also comes with downsides. Crime is also increasingly digitized, from hate speech and cyberbullying to hacking and identity theft. Since 2016, hacking has been the most prevalent crime in the Netherlands. Specific numbers are however hard to come by. Victims of cybercrime are not reporting their victimization to the police, which leads to unreliable crime statistics. And estimations of the cost of cybercrime differ substantially between academic researchers and commercial companies offering protection, training, and insurance. In the meantime, organizations are spending much time, effort, and money on training their employees to become more resilient. However, the effectiveness of these interventions are seldom properly measured. In this working group, we aimed to identify and describe techniques to systematically measure digital behaviors relevant to the following two contexts:

1. online misbehavior and false information (e.g., fake news, rumours, hating, cyberbullying).
2. cybersecurity (e.g., phishing, ransomware).

While these online threats are commonly researched from different perspectives, we recognize a lack of well-defined and comprehensive methodological frameworks how to measure interactions between them and human behaviour – how to measure their enablers

⁴ See application of Johari Window to security activity, as in e.g., Beris, Odette, Adam Beautelement, and M. Angela Sasse. “Employee rule breakers, excuse makers and security champions: Mapping the risk perceptions and emotions that drive security behaviors.” Proceedings of the 2015 New Security Paradigms Workshop. ACM, 2015.

⁵ Herley, Cormac. “More is not the answer.” IEEE Security & Privacy 12.1 (2013): 14-19.

(i.e., what makes online threats possible and effective) and influences (i.e., how online threats affect humans and their behaviour). We were particularly interested in measuring:

- reach and effect of online misbehavior and false information threats on influencing human behavior.
- security behavior that may expose individuals and organisations to cybersecurity threats.

The output of the working group are two methodological frameworks for each category of threats that consist of a list of addressed online threats and corresponding security behaviors; and identification of technical measurements that can be practically used to study such online threats and human behaviour

A framework for measuring online misbehavior and false information. As the first part of this framework, we proposed a hierarchical categorization of different types of online threats. We identified 4 main groups on online threats: deception, manipulation, aggression and mischief (we focused in more detail on the first three groups in the framework). Secondly, we identify typical victims and offenders for each online threat (as potential actors we considered individual users, communities, organizations, governments and societies). For each category of threats, we identified how we can determine:

- The reach of threats, e.g. we can measure the speed of spreading for deception and manipulation threats (such as fake news) by number of likes, retweets, shares, replies, comments, etc. per time unit.
- The effect of threats, e.g. we can measure how fake news influenced the political preference by looking at election results or changes in voting behavior.

To sum up, the framework consists of 3 main components: hierarchical categorization of different types of online threats, identification of victims and offenders and metrics to measure reach and effect of threats. In addition, we can summarize our main findings by two take-away messages: 1) We still miss a comprehensive list of definitions and categorizations for individual online threats. 2) While we can measure the reach of threats quite well (reach is well observable), the measurements of their effect cannot be determined precisely (the effect is usually hidden and influenced by a number of additional circumstances).

A framework for measuring human security behavior. A valuable input provided by Sophie van Der Zee became the basis for this work. Her initial framework consists of four components: 1) user groups, 2) possible factors that influence or can be used to influence user's security behaviour, possible 3) metrics and 4) approaches to measure user's behaviours. Based on this input, we decided to focus on possible observable security behaviours. We did a brainstorming session with group members using post-it notes and identified a list of possible security behaviours of individuals or organizational users. In this session, we came up with 26 security behaviours that we grouped in nine categories. These categories are related to browser behaviour, use of a smartphone, software, hardware, emails, passwords, document handling, laptop, and file sharing sites. In the next step, we looked into technical measurements/metrics that can be used to study the corresponding security behaviour in the wild. For example, the data describing security behaviour related to "Changing default passwords (for new accounts, routes, IoT devices)" can be collected by scanning accounts/devices based on the default password list.

As the last component of our framework, we thought about possible research study designs that can be used to investigate each security behaviour using specific technical metric. For the above example of "Changing default passwords" we proposed the following study design: "AB-test: scan for default passwords → provide awareness regarding default passwords → scan same 'population' again after a short time → compare scans."

To sum up, the framework consists of 3 main components: observable security behavior, technical measurement or metric, and suggested research study design to investigate the corresponding behavior using specific metric. This framework aims at providing researchers and practitioners with a practical and structured way of studying human security behaviours.

Conclusion. In summary, our working group drafted two methodological frameworks for researchers and practitioners who are interested in studying and measuring 1) the reach and effect of online threats on influencing human behavior and 2) actual human cybersecurity behavior.

4.4 Theme 4: Prevention, Detection, Response and Recovery

Gianluca Stringhini (Boston University, US), Freya Gassmann (Universität des Saarlandes, DE), Marianne Junger (University of Twente, NL), Elmer Lastdrager (SIDN Labs – Arnheim, NL), Michael Sirivianos (Cyprus University of Technology – Lemesos, CY), and Sebastian Wachs (Universität Potsdam, DE)

License © Creative Commons BY 3.0 Unported license
© Gianluca Stringhini, Freya Gassmann, Marianne Junger, Elmer Lastdrager, Michael Sirivianos, and Sebastian Wachs

The working group focused on Prevention, Detection, Response, and Recovery approaches with respect to cybersafety incidents. To guide the discussion, three very distinct topics were examined: (1) Cyber grooming, (2) phishing; and (3) IoT.

After the brainstorming session, the group decided to focus on a specific topic. It turned out the group was packed with expertise pertaining to research of adolescents. Therefore, the working group decided to focus on cyber grooming as the main topic for further discussions. Cyber grooming occurs when someone (often adult) befriends a child or adolescent and builds an emotional connection with future intentions of sexual abuse and/or exploitation.

The main goal of cyber grooming is to gain the trust of the child, which can for example be exploited to obtain intimate and personal data from the child (often sexual in nature, such as sexual conversations, pictures, or videos). They in turn can be used to threaten and blackmail the child for further inappropriate material or acts.

Unfortunately, adolescents rarely turn to an adult for help when they face problems online. Imposing online restrictions might be perceived as a threat to their freedom and thus induce a psychological reactance process leading to undesired behavior. In order to protect minors, we need to equip them and their guardians with appropriate tools that can tackle challenging situations and empower users to deal with threats in a thoughtful manner.

Considering all these difficulties, the group came up with the “Guardian Angel approach.” This approach entails a proper suite of cybersafety tools that let a minor create a “Guardian Avatar” which can be customized so that it feels familiar. The main goal of the guardian angel is to protect children against groomers and those who plan to abuse their trust and take advantage of them. The avatar pops up when the system detects something suspicious and advises the minor accordingly.

We mentioned a number of requirements, such as that the tool should be age appropriate and culturally appropriate. We also discussed how parents should or could be involved. Depending on the age of the minor, the system provisions for various degrees of privacy. In the first mode the avatar will be invoked only when the user initiates it. This is the least intrusive modality, which is tailored to adolescent users. In the second mode, the avatar is

automatically activated by the system in the response to intelligent detection. This modality is more appropriate for pre-adolescent children. In either case, the avatar will help the victim cope with a dangerous situation. In the case of adolescents, the system engages the minor with a series of questions, answers and advice. Parents will be notified only with the consent of the teenager. In the case of pre-adolescent children, the tool engages the parents as deemed appropriate. Moreover, the avatar-based system can become a learning environment with tutorials. Therefore, the system will also have educational value and can be introduced in classrooms and awareness workshops.

Despite the actions taken from the avatar to prevent cybergrooming, the minor may end up trusting potential attackers more than the avatar, where the minor should only trust the avatar or its parents. Therefore, special care should be taken to gain the trust of the minor by using appropriate UX design and proper settings. As a start, the parent enters the age of the minor and the system should automatically choose the right level of intervention. By analyzing the interactions with its users, the system will progressively learn how to address various types of users and situations.

We stressed that the guardian angel should, ideally, be embedded in more general policies to protect children online, such as media education at school.

Overall, the research will focus on interventions against interpersonal online aggression, ICT- based cybergrooming detection tools, seeking online help, the effectiveness of online assistants, human factors and user experience, effects of alerting parents, experiments with the monitoring of adolescent's mobile and case studies analysis. Furthermore, natural language processing, image analysis, and fact checking modules will be implemented. The evaluation will also entail three user studies which will take place within small and medium scale pilots. In particular, a study of user acceptance for the "Guardian Angel approach" will take place first. Subsequently, a group of adolescents will use the tools and cybersafety-related responses will be compared to a group that does not use the tool. In all studies, the ethnic and socio-economic background of the users will be taken into consideration.

Regarding privacy considerations, ideally the data should be processed only on the user's computer or in Web proxies at the user's residence. At the same time, feedback should be used to update the models of the project and make them more accurate. To this end, privacy-preserving federated learning approaches will be employed. Overall, the application can have various privacy preferences, ranging from 'full monitoring' for younger children, to 'on demand' for adolescents. Every action will follow the GDPR regulations and the users will be fully informed.

The above concepts will be proposed for EU-funding (probably ETN/ITN 2020). In addition, numerous stakeholders will be contacted, including the Cyprus Ministry of Education, the Cyprus Police, Adolescents' Parliament in the Netherlands, foundations (NGOs, GOs) that work with sexuality awareness for teenagers, the Dutch police, teachers, parent associations, schools, to evaluate the idea, collect feedback and to raise awareness.

Participants

- Abhishta Abhishta
University of Twente, NL
- Zinaida Benenson
Universität Erlangen-
Nürnberg, DE
- Matt Bishop
University of California –
Davis, US
- Jan-Willem Bullée
University of Twente, NL
- Joe Calandrino
Federal Trade Commission –
Washington, US
- Deanna Caputo
MITRE – Washington D.C., US
- Claude Castelluccia
INRIA – Grenoble, FR
- Yi Ting Chua
University of Cambridge, GB
- Natalie Ebner
University of Florida –
Gainesville, US
- Matthew Edwards
University of Bristol, GB
- Manuel Egele
Boston University, US
- Jeremy J. Epstein
NSF – Alexandria, US
- Freya Gassmann
Universität des Saarlandes, DE
- Alice Hutchings
University of Cambridge, GB
- Marianne Junger
University of Twente, NL
- Katsiaryna Labunets
TU Delft, NL
- Elmer Lastdrager
SIDN Labs – Arnheim, NL
- Gabriele Lenzini
University of Luxembourg, LU
- Daniela Oliveira
University of Florida –
Gainesville, US
- Simon Parkin
University College London, GB
- William Robertson
Northeastern University –
Boston, US
- Stefan Schiffner
University of Luxembourg, LU
- Michael Sirivianos
Cyprus University of Technology
– Lemesos, CY
- Ivan Srba
STU – Bratislava, SK
- Gianluca Stringhini
Boston University, US
- Gareth Tyson
Queen Mary University of
London, GB
- Sophie van Der Zee
Erasmus University –
Rotterdam, NL
- Sebastian Wachs
Universität Potsdam, DE
- Victoria Wang
University of Portsmouth, GB
- Jeff Yan
Linköping University, SE
- Savvas Zannettou
Cyprus University of Technology
– Lemesos, CY

