



DAGSTUHL REPORTS

Volume 15, Issue 3, March 2025

Guardians of the Galaxy: Protecting Space Systems from Cyber Threats (Dagstuhl Seminar 25101) <i>Ali Abbasi, Gregory J. Falco, Daniel Fischer, and Jill Slay</i>	1
The Future of Games in Society (Dagstuhl Perspectives Workshop 25102) <i>Anders Drachen, Johanna Pirker, and Lannart E. Nacke</i>	39
Computational Complexity of Discrete Problems (Dagstuhl Seminar 25111) <i>Swastik Kopparty, Meena Mahajan, Rahul Santhanam, Till Tantau, and Ian Mertz</i>	56
PETs and AI: Privacy Washing and the Need for a PETs Evaluation Framework (Dagstuhl Seminar 25112) <i>Emiliano De Cristofaro, Kris Shrishak, Thorsten Strufe, Carmela Troncoso, and Felix Morsbach</i>	77
Scheduling (Dagstuhl Seminar 25121) <i>Claire Mathieu, Nicole Megow, Benjamin J. Moseley, and Frits C. R. Spiessma</i> ..	94
Climate Change: What is Computing's Responsibility? (Dagstuhl Perspectives Workshop 25122) <i>Vicki Hanson and Bran Knowles</i>	113
Weihrauch Complexity: Structuring the Realm of Non-Computability (Dagstuhl Seminar 25131) <i>Vasco Brattka, Alberto Marcone, Arno Pauly, Linda Westrick, and Kenneth Gill</i> ..	125
Approximation Algorithms for Stochastic Optimization (Dagstuhl Seminar 25132) <i>Lisa Hellerstein, Viswanath Nagarajan, and Kevin Schewior</i>	159
Categories for Automata and Language Theory (Dagstuhl Seminar 25141) <i>Achim Blumensath, Mikołaj Bojańczyk, Bartek Klin, and Daniela Petrișan</i>	177
Explainability in Focus: Advancing Evaluation through Reusable Experiment Design (Dagstuhl Seminar 25142) <i>Elizabeth M. Daly, Simone Stumpf, and Stefano Teso</i>	201

ISSN 2192-5283

Published online and open access by

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

Publication date

October, 2025

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 <https://dnb.d-nb.de>.

License

This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0).



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

- Elisabeth André
- Franz Baader
- Goetz Graefe
- Reiner Hähnle
- Barbara Hammer
- Lynda Hardman
- Steve Kremer
- Rupak Majumdar
- Heiko Mantel
- Lennart Martens
- Albrecht Schmidt
- Wolfgang Schröder-Preikschat
- Raimund Seidel (*Editor-in-Chief*)
- Heike Wehrheim
- Verena Wolf
- Martina Zitterbart

Editorial Office

Michael Wagner (*Managing Editor*)
Michael Didas (*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
<https://www.dagstuhl.de/dagrep>

Digital Object Identifier: 10.4230/DagRep.15.3.i

Guardians of the Galaxy: Protecting Space Systems from Cyber Threats

Ali Abbasi^{*1}, Gregory J. Falco^{*2}, Daniel Fischer^{*3}, and Jill Slay^{*4}

1 CISA Helmholtz Center for Information Security, DE. abbasi@cispa.de

2 Cornell University – Ithaca, US. gfalco@cornell.edu

3 ESA / ESOC – Darmstadt, DE. daniel.fischer@esa.int

4 University of South Australia – Mawson Lakes, AU. jill.slay@unisa.edu.au

Abstract

This report documents the program and outcomes of Dagstuhl Seminar 25101 “Guardians of the Galaxy: Protecting Space Systems from Cyber Threats,” which brought together 40 participants from 11 countries. It explains why space cybersecurity is distinct from terrestrial contexts and distills the working-group results (attack/prepare, detect, protect, respond) into a focused research-and-action roadmap for agencies, industry, and academia.

Seminar March 02–07, 2025 – <https://www.dagstuhl.de/25101>

2012 ACM Subject Classification Computer systems organization → Embedded and cyber-physical systems; Security and privacy → Security in hardware; Security and privacy → Systems security; Networks → Network security

Keywords and phrases Space Cybersecurity, Satellite Security, Cyber-Physical Systems, Network Security, Embedded Systems Security, System Security, Autonomous Systems Security, Post-Quantum Cryptography

Digital Object Identifier 10.4230/DagRep.15.3.1

1 Executive Summary

Ali Abbasi

Gregory J. Falco

Daniel Fischer

Jill Slay

License © Creative Commons BY 4.0 International license
© Ali Abbasi, Gregory J. Falco, Daniel Fischer, and Jill Slay

This report synthesizes the outcomes of Dagstuhl Seminar 25101, “Guardians of the Galaxy: Protecting Space Systems from Cyber Threats,” which convened 40 experts from academia, industry, and government. The seminar established a clear consensus that space cybersecurity is a qualitatively distinct discipline, not merely an extension of terrestrial challenges. The seminar focused on:

- **Defining the Foundational Challenges:** Articulating why space is different and how this affects the security domain, focusing on the ambiguity created by the harsh physical environment, the necessity of high-stakes autonomy due to extreme latency, and the uniquely asymmetric attack surface.
- **Structuring the Problem Space:** Organizing analysis and solutions around four key operational functions via dedicated working groups: ATTACK/PREPARE, DETECT, PROTECT, and RESPOND.

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Guardians of the Galaxy: Protecting Space Systems from Cyber Threats, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 1–38

Editors: Ali Abbasi, Gregory J. Falco, Daniel Fischer, and Jill Slay



DAGSTUHL
REPORTS

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

- **Formulating a Strategic Roadmap:** Proposing a multi-pillar plan to foster a cumulative and collaborative research ecosystem that bridges the gap between academic innovation and operational needs.

As a major result, the seminar identified the following interconnected problem areas and corresponding future research directions:

1. **The Testbed and Data Gap:** Overcoming the critical shortage of realistic research infrastructure by developing a federated ecosystem of high-fidelity testbeds. A key requirement is that these testbeds must be segment-complete, modeling the entire ground-link-space chain, and support graduated fidelity. This allows researchers to move between pure simulation, hardware-in-the-loop, and testing with unmodified firmware binaries depending on the research question. Furthermore, institutional spacecraft operators should be encouraged to share more representative data sets that can be used in research.
2. **Securing Next-Generation Communications:** Addressing the unique security needs of future space networks. This includes maturing protocols for the Solar System Internet (e.g., Delay Tolerant Networking – DTN) essential for deep space, planning the transition to Post-Quantum Cryptography (PQC), and developing resilient defenses against jamming and spoofing for high-bandwidth optical and RF links.
3. **Building Trustworthy Autonomous Systems:** Ensuring that onboard AI and autonomous systems are secure and safe. This requires developing physics-informed and resource-aware intrusion detection, designing systems to be “forensic-by-design” so that evidence of an attack survives recovery actions, and implementing verifiable and secure software update pipelines.
4. **Strengthening the System Foundation:** Mandating a “secure-by-design” philosophy anchored in hardware. This involves adopting measured boot processes, internal message-level authentication, and robustly managing the cybersecurity of the global supply chain (C-SCRM) for all components.
5. **Establishing a Collaborative Ecosystem:** Creating the necessary non-technical structures for progress. This includes developing clear governance and interoperable standards, establishing “safe-harbor” policies for vulnerability disclosure, and implementing new collaborative models, such as co-funded PhD programs, to grant researchers vital access to realistic systems and data.

2 Table of Contents

Executive Summary

<i>Ali Abbasi, Gregory J. Falco, Daniel Fischer, and Jill Slay</i>	1
--	---

Overview of Talks

Powering Europe's Space Ambition: Cybersecurity Challenges in Space Systems <i>Daniel Fischer</i>	5
Cybersecurity Challenges in Space Systems: Notable Challenges and Research Areas <i>Marcus Wallum</i>	5
Space System Security and the Space Environment <i>Knut Eckstein</i>	6
Down to Earth: Cyber Security Operations <i>Markus Rückert</i>	6
Hack The Planet and Beyond: Security Challenges of the Solar System Internet (SSI) <i>Lars Baumgärtner</i>	7
NASA Mission Resilience & Protection Approach – Including space Cybersecurity <i>Kevin Gilbert</i>	7
Security Units for Satellite Communication Challenges <i>Arne Grenzebach</i>	7
Space Attack Research and Tactic Analysis (SPARTA) <i>Brandon Bailey</i>	8
Migrating Legacy Ground Stations to Cloud-based Zero-trust Stations <i>Mattias Wallén</i>	8
New Space = Secure Space? <i>Steven Arzt</i>	8
A Joint Effort: Stakeholder Cooperation for Better Cybersecurity in Space <i>Florian Göhler</i>	9
Merge/Space: A Security Testbed for Satellite Systems <i>Stephen Schwab</i>	9
HoneySat: A Network-based Satellite Honeytrap Framework <i>Efrén López-Morales</i>	9
Securing the Satellite Software Stack <i>Samuel Jero</i>	10
Developing accessible test beds and data sets <i>Jill Slay</i>	10
On the Security of Non-Terrestrial Networks <i>Gunes Karabulut Kurt</i>	10


Open Problems	11
--------------------------------	----

Working Groups	14
Seminar Organization	14
Working Group on ATTACK/PREPARE	14
Working Group on DETECT	14
Working Group on PROTECT	15
Working Group on RESPOND	15
What Makes Space So Different	16
Summary	19
Future Work and Challenges Ahead	20
Secure Space Communications and Encryption in the Quantum Era	20
AI-Driven and Autonomous Space Cybersecurity	24
Secure-by-Design in Space System Hardware and Software	28
Cyber-Physical Resilience for Multi-Domain Missions	30
Policy, Governance, and Standardization	32
Cyber Security Testbed	33
Recommendations for Future Research and Development	34
Conclusion	36
Participants	38

3 Overview of Talks

3.1 Powering Europe's Space Ambition: Cybersecurity Challenges in Space Systems

Daniel Fischer (ESA / ESOC – Darmstadt, DE, daniel.fischer@esa.int)

License  Creative Commons BY 4.0 International license
© Daniel Fischer

The European Space Agency (ESA) is responsible for the peaceful exploitation of space on behalf of its member states. It is active in all major domains of space, from launchers, human spaceflight, earth observation, and GNSS, to science and communication. Many of the systems developed by ESA, either directly on behalf of its member states, or on behalf of the European Commission (e.g., Galileo, Copernicus, IRIS2), represent critical infrastructure upon which society depends on a daily basis.

Cybersecurity has thus grown to be a major challenge in the development of ESA programs and assets, in particular in today's changing geopolitical landscape. In response to these challenges, ESA has made cybersecurity one of its three main technology priorities in addition to quantum and AI.

The quick development and maturation of space security technologies, together with the European space industry and academia, is fundamental. For this purpose, ESA seeks to connect closer with these entities and exploit synergies. ESA seeks to supply the academic ecosystem with relevant space use cases while benefitting from the resulting research to speed up technology spin-in. Likewise, industry is a valuable partner in picking up the higher technology readiness level (TRL) developments in cybersecurity and creating a diverse ecosystem for operational cybersecure space system assets and components.

3.2 Cybersecurity Challenges in Space Systems: Notable Challenges and Research Areas

Marcus Wallum (ESA / ESOC – Darmstadt, DE, marcus.wallum@esa.int)

License  Creative Commons BY 4.0 International license
© Marcus Wallum

The talk presented an overview of current challenges and potential future research topics. Topics included :

- Digital security engineering, alignment with Model-based System Engineering and formal reference architectures
- Zero trust architectures for space systems
- Post-Quantum Cryptography, its impact on space communications and need for cryptographic agility
- Tailored space system security monitoring and testing solutions including fuzzing of space communication protocols
- Securing legacy systems
- Anomaly detection and responsive resilient self-healing architectures
- Securing the supply chain
- Evolution of avionics security architectures and their secure operation, including on-board IDS/IPS, TEE, remote attestation
- Leveraging AI for security and ensuring secure use of AI
- Applied confidential computing and homomorphic encryption for secure distributed dataset processing
- Proliferation of standards, regulations and certification scheme

3.3 Space System Security and the Space Environment

Knut Eckstein (ESA / ESTEC – Noordwijk, NL, knut.eckstein@esa.int)

License  Creative Commons BY 4.0 International license
© Knut Eckstein

The talk aimed at initiating fruitful discussions between academics and practitioners by positing which aspects of space systems security engineering are the most challenging or the most interesting from an academic Research and Development perspective. It started by noting that spacecraft, compared to other mobile network nodes, have neither the least powerful CPUs, nor the smallest amounts of memory, nor the least predictable communication network topologies, nor the longest periods of communication outages. What is special about spacecraft is that their wireless links are highly asymmetric in nature and are absolutely essential, in absence of any wired links that can be established in drones or aircraft during maintenance phases. Also, spacecraft are fairly unique in their focus of safety and availability over long periods of time without “return to base” i.e. any security mechanism design has to satisfy very stringent safety requirements.

3.4 Down to Earth: Cyber Security Operations

Markus Rückert (ESA / ESOC – Darmstadt, DE, markus.rueckert@esa.int)

License  Creative Commons BY 4.0 International license
© Markus Rückert

Following the NIST Cyber Security Framework (CSF), the talk summarized ESA’s approach to PROTECT, DETECT, and RESPOND at a conceptual level and in order to protect ESA’s Operations, Investments, and Brand Value. The talk created awareness of sector-specific cyber security challenges with the aim of stimulating the ideation for seminar topics.

The talk illustrated the nature and diversity of the assets (infrastructure, services, information) that require protection.


Furthermore, the talk highlighted a series of key challenges in the context of cyber-physical systems as opposed to traditional IT.

The widespread use of shared ground infrastructures, due to cost benefits, exposes a wide attack surface, making it harder to protect from cyber threats. Complex and highly specialized supply chains present unique challenges when it comes to effective identification and management of weaknesses and vulnerabilities, as well as when it comes to the identification of threats and countermeasures. Similarly, the presence of dual-use technologies may limit information sharing among the parties involved. In general, system complexity and interoperability constraints often slow the adoption of new technologies, including improved security controls and protection practices.

In general, the resulting inertia and obstacles affect the evolution of PROTECT, DETECT, and RESPOND. There is a general call for research, development and innovation to take these factors into account.

3.5 Hack The Planet and Beyond: Security Challenges of the Solar System Internet (SSI)

Lars Baumgärtner (ESA / ESOC – Darmstadt, DE, lars.baumgaertner@esa.int)


License  Creative Commons BY 4.0 International license
© Lars Baumgärtner

The SSI is built upon new protocols, technologies, and mechanisms, particularly the concept of ‘store-carry-and-forward’ (SCF) for Delay-Tolerant Networking (DTN). While this approach addresses the fluctuating connectivity and high delays inherent in interplanetary communication, it also creates the need for novel security solutions and prevents the use of existing security measures. Several key areas present major challenges:

- Delay-tolerant key management
- SSI Threat Modelling
- Delay-tolerant networking (DTN) Anom-
- Detection
- Security of inter-planetary multicast
- Scalable network testbeds for SSI

3.6 NASA Mission Resilience & Protection Approach – Including space Cybersecurity


Kevin Gilbert (NASA Goddard Space Flight Center – Greenbelt, US, kevin.w.gilbert@nasa.gov)

License  Creative Commons BY 4.0 International license
© Kevin Gilbert

This talk provides an overview of NASA STD-1006A (NASA’s space protection requirements), then gives an overview of the NASA protection planning process (which includes Candidate Protection Strategies related to space mission cybersecurity), and will conclude with a snapshot of where we think development is needed to find protection solutions for civil space missions.

3.7 Security Units for Satellite Communication | Challenges

Arne Grenzebach (OHB System – Bremen, DE, arne.grenzebach@ohb.de)

License  Creative Commons BY 4.0 International license
© Arne Grenzebach

This talk presents the current challenges of developing security units for satellite communication. This is based on industrial experience within a satellite manufacturing company, namely OHB.

3.8 Space Attack Research and Tactic Analysis (SPARTA)


Brandon Bailey (The Aerospace Corp. – Los Angeles, US, brandon.bailey@aero.org)

License  Creative Commons BY 4.0 International license
© Brandon Bailey

This talk presents an overview of the Tactic Technique, & Procedure (TTP) framework called SPARTA. We describe how it can be used to document attacks on spacecraft along with countermeasures to mitigate or prevent the attacks. The goal was education and awareness of the tool & present future capabilities. SPARTA was the first of its kind repository of knowledge on how to attack or defend spacecraft.

3.9 Migrating Legacy Ground Stations to Cloud-based Zero-trust Stations

Mattias Wallén (Swedish Space Corporation – Solna, SE, mattias.wallén@sscspace.com)

License  Creative Commons BY 4.0 International license
© Mattias Wallén

This talk presents current threats and vulnerabilities to satellite ground stations. The talk was focused on the need to move to cloud-based ground stations and reduce risk by using DevSecOps, loosely coupled systems, Zero trust architectures, policy as code, infrastructure as code, and compliance as code. Compared to the physical industry, computer science, and IT – the space industry is still in a “Steam Power” state and moving towards assembly line and automation.

3.10 New Space = Secure Space?

Steven Arzt (Fraunhofer SIT – Darmstadt, DE, steven.arzt@sit.fraunhofer.de)

License  Creative Commons BY 4.0 International license
© Steven Arzt

The space industry is changing with the “New Space” activities, new technologies and new business models challenge traditional risk models and security measures. As part of the expert group on space security by BSI (Germany’s Federal Office for Information Security), we look into this evolving landscape. Further, governmental missions on new topics such as QKD, space debris, and AI-driven security analysis require us to change existing solutions and insights. What would a world in which anyone can launch a satellite a rent a ground station look like security-wise?

Lastly, we need to bring more bright minds into the intersection of space and cybersecurity. Hacking contests and CTFs can bridge the path into the field and reduce the barrier of entry.

3.11 A Joint Effort: Stakeholder Cooperation for Better Cybersecurity in Space

Florian Göhler (BSI – Bonn, DE, florian.goehler@bsi.bund.de)

License © Creative Commons BY 4.0 International license
© Florian Göhler

Cybersecurity needs to be an integrated part of every space mission, and security aspects should be considered throughout all phases of a project. However, there was a lack of regulation and security standards that address cyber threats in space. To overcome this issue, the German Federal Office for Information Security founded an expert group for cybersecurity in space that invites experts from governmental institutions, industry, and academia to work together on standardization and regulation. In this joint effort, the expert group developed multiple documents that aim to mitigate cyber threats on space and ground segments. Furthermore, the expert group aims to identify emerging new technologies and regulations that may impact cybersecurity in space. These efforts also take international developments into account. This talk will give an overview of the activities of the group and its security documents.

3.12 Merge/Space: A Security Testbed for Satellite Systems

Stephen Schwab (USC/ISI – Arlington, US, schwab@isi.edu)

License © Creative Commons BY 4.0 International license
© Stephen Schwab

Merge/Space (M/S) is a testbed designed to simulate multiple-agent security scenarios in satellite networks. By combining orbital data generated by a simulator such as STK with a synchronized set of images, M/S can accurately simulate bandwidth and connectivity constraints between ground stations and vehicles, enabling analyses of DoS attacks, scanning, malware infiltration, and other analyses. We discuss the development of the testbed, and the sample datasets included for release, and demonstrate the impact of various simulations.

3.13 HoneySat: A Network-based Satellite Honeypot Framework

Efrén López-Morales (Texas A&M University – Corpus Christi, US, elopezmorales@islander.tamucc.edu)

License © Creative Commons BY 4.0 International license
© Efrén López-Morales
Joint work of Efrén López-Morales, Ulysse Planta, Ali Abbasi

Satellites are the backbone of several mission-critical services such as GPS that enable our modern society to function. For many years, satellites were assumed to be secure because of their indecipherable architectures and the reliance on security by obscurity. However, technological advancements have made these assumptions obsolete, paving the way for potential attacks, and sparking a renewed interest in satellite security. Unfortunately, to this day, there is no efficient way to collect data on adversarial techniques for satellites, which severely hurts the generation of security intelligence. In this paper, we present HoneySat, the first high-interaction satellite honeypot framework, which is fully capable of convincingly

simulating a real-world CubeSat, a type of Small Satellite (SmallSat) widely used in practice. To provide evidence of the effectiveness of HoneySat, we surveyed experienced SmallSat operators currently in charge of active in-orbit satellite missions. Results revealed that the majority of satellite operators (71.4%) agreed that HoneySat provides realistic and engaging simulations of CubeSat missions. Further experimental evaluations also showed that HoneySat provides adversaries with extensive interaction opportunities by supporting the majority of adversarial techniques (86.8%) and tactics (100%) that target satellites. Additionally, we also obtained a series of real interactions from actual adversaries by deploying HoneySat on the internet over the span of several months, confirming that HoneySat can operate covertly and efficiently while collecting highly valuable interaction data.

3.14 Securing the Satellite Software Stack


Samuel Jero (MIT Lincoln Laboratory – Lexington, US, samuel.jero@ll.mit.edu)

License  Creative Commons BY 4.0 International license
© Samuel Jero

Satellites and the services enabled by them play an increasingly important in our modern life. To support these services, satellite software is becoming increasingly complex and connected. As a result, concerns about its security are becoming prevalent. While the focus of security has historically been encrypting communication links, we argue that further consideration of the security of satellites is necessary. This talk characterizes the cyber threats to satellites, surveys the unique challenges for satellite software, and presents a vision for future research in this area.

3.15 Developing accessible test beds and data sets

Jill Slay (University of South Australia – Mawson Lakes, AU, jill.slay@unisa.edu.au)

License  Creative Commons BY 4.0 International license
© Jill Slay

To expand and extend the growing area of satellite cybersecurity to larger and more diverse cohorts of cross-disciplinary researchers internationally, we need appropriate datasets and test beds where developed protection solutions can be studied. The emerging challenge is to standardize such research infrastructure to begin to answer wicked space cyber research questions so as to protect humans and their space missions.

3.16 On the Security of Non-Terrestrial Networks

Gunes Karabulut Kurt (Polytechnique Montréal, CA, gunes.kurt@polymtl.ca)

License  Creative Commons BY 4.0 International license
© Gunes Karabulut Kurt

6G networks are expected to be a combination of the terrestrial network and the non-terrestrial network (NTN). Elements of NTN will be base stations with 3D mobility, such as low Earth orbit (LEO) satellites, unmanned autonomous vehicles (UAVs), and high altitude

platform station (HAPS) systems. The presence of such NTN elements introduces new features in terms of coverage, computation, localization, and sensing. However, their presence also makes 6G networks vulnerable to new security threats, especially in the physical layer (PHY). After detailing the NTN evolution, this talk focuses on two different threats. The first threat type emerges from the communication attacks that are expected to increase with the presence of wireless backhaul connectivity. The second threat type is on the localization systems, especially for NTN elements, as the location information of a LEO satellite, a UAV, or a HAPS is an essential network characteristic that will affect the overall network performance. The talk will conclude with the importance of physical layer security for NTNs, an overview of the open issues, and future research directions.

4 Open Problems

Space-security research is hindered by a tooling gap: there is no widely usable way to create mission-realistic attack data. Generic IT labs and pure simulation miss ground–link–space timing, radio effects, and operational modes; export controls and proprietary interfaces further restrict sharing and instrumentation. The result is a shortage of trustworthy datasets for studying adversary TTPs, validating detectors, and training operators. The remedy is a modular testbed strategy comprised of digital twins with selectively inserted hardware-in-the-loop driven by the question under test and instrumented to emit synchronized command/telemetry, process/file, memory-integrity, and bus/link traces. Synthetic data should be generated from these twins with explicit provenance so results are comparable across teams.

Communication and cryptography issues dominate the second cluster of problems. Delay-/disruption-tolerant operation breaks assumptions about freshness and ordering, making key establishment, revocation, and replay defenses fragile on legacy waveforms. Post-quantum cryptography must be planned at the protocol level, not patched in, and optical/QKD concepts need evaluation against pointing loss, weather, and scheduling realities. Internally, many space system platforms remain flat: subsystems share buses without message-level authentication or authorization. Moving toward zero-trust within the vehicle and standardizing minimal, interoperable logging and attestation would close recurring gaps. In parallel, link protection must explicitly address *jamming and spoofing* with sensing-and-mitigation loops and *adaptive* RF/optical protocols that maintain integrity under Doppler, scintillation, and variable contact geometry, and key management and trust must operate across ground–link–space with DTN-aware revocation/rekey and onboard entropy/key health checks.

Detection, autonomy, and response form the third cluster. AI/ML anomaly detection faces sparse labels and physics-induced artifacts (radiation, Doppler, eclipse power transients) that mimic attacks; onboard compute and energy limits constrain model size and update cadence; explainability is required for autonomous action. Beyond security analytics, *predictive maintenance* on flight subsystems and *AI-based threat-intelligence fusion* for space telemetry are needed to anticipate degradations and prioritize hunts. Today's resilience mechanisms often erase the very evidence needed for attribution. Systems, therefore, need provenance-preserving ECC/TMR and scrubbing, append-only anomaly journaling that survives resets, and downlink strategies that trickle forensic records over multiple passes. Response playbooks must assume intermittent contact: contain while preserving observability and commandability, re-key under DTN constraints, and execute ranked recoveries that prioritize mission-critical services.

Finally, space’s cyber-physical character and governance context raise problems that tools alone cannot solve. Security assessment must fuse cyber telemetry with SSA to reason about proximity operations, illumination changes, and constellation effects; redundancy and graceful degradation must preserve control and downlink rather than merely “stay on.” For long missions, *on-orbit servicing and manufacturing (OSAM)* should be planned as controlled security touchpoints, and *secure IoT/edge nodes* treated as first-class participants in command and sensing. Constellation behaviors also introduce *swarming attack/defense* dynamics, requiring coordinated detection and topology-aware degradation. Supply-chain assurance, standardized secure-by-design stacks (measured boot, crypto agility), and explicit end-of-life/serviceability paths are prerequisites for long missions. Policy and standards remain fragmented – liability across commercial/government assets is unclear, and sharing is constrained, so progress depends on harmonized norms for TT&C protection, minimal common data schemas, and adoption of emerging technologies (confidential computing, PQC, quantum/optical links) only when backed by mission-level threat models and viable update paths. Adoption decisions should further consider *dynamic payload adaptability* and *AI-assisted space-traffic management*, each gated by partitioning, attestation, and robust update mechanisms. Environmental extremes, orbital dynamics, and irretrievability make early design choices tricky; getting these foundations right is the only scalable risk reducer. The details of identified issues are listed in Table 1.

■ **Table 1** Consolidated challenges identified by participants.

Category	Identified Problems and Topics
Cyber Range and Simulation	Realistic cyber range simulations; High-fidelity test environments; Digital twins with selective hardware-in-the-loop; Scenario-based threat/attack simulation; Virtualization and emulation; Modular/adaptable testbeds; Operator training and awareness scenarios; Attack modeling and adversary emulation; Synthetic data generation with provenance
Delay/Disruption-Tolerant Networking (DTN)	Store-carry-forward security; Freshness/ordering under long delays; Contact scheduling effects; Robust replay protection and expiry; Routing and identity under fragmentation; DTN-aware revocation/rekey; Protocol interoperability across DSN/cislunar contexts
Secure Communications and Encryption	PQC (algorithms and <i>protocols</i>); QKD/optical feasibility and operations; Robust RF/optical authentication; Jamming/spoofing detection and response; Crypto for ground-space links under high BER/Doppler; Link-layer vs. end-to-end protections
Key Management & Trust Infrastructure	Mission-phase keying (commissioning, cruise, critical ops); Key distribution across ground-link-space; Compromise recovery and re-bootstrap under DTN; HSM/TEE use on ground and onboard; Entropy health and key/credential aging in radiation environments
AI and Autonomous Cybersecurity	AI/ML intrusion detection; Physics-conditioned anomaly detection; Onboard constraints (compute/energy/update cadence); Explainable autonomy and fail-safe action; Predictive maintenance vs. adversarial ML risks; AI-based threat intelligence
Secure-by-Design and Hardware Security	Standardized but diversified stacks; Internal message-level authz/authn (zero-trust within spacecraft); Measured/secure boot; Firmware integrity and updateability (A/B, shadow execute); Embedded crypto modules; Supply-chain security and component provenance
Incident Response and Forensics	DTN-compatible incident playbooks; Autonomous containment that preserves commandability/observability; Provenance-preserving ECC/TMR/scrubbing; Append-only anomaly journals surviving resets; Preplanned recovery options and evaluation
Cyber-Physical Resilience & SSA Coupling	Cyber with SSA (proximity operations, illumination changes); Constellation-level behaviors; Hybrid physical-cyber assessment; Redundancy and graceful degradation preserving downlink/control; Secure IoT/edge nodes in space
Safety-Security Co-engineering & Assurance	Composability of controls across EPS/ADCS/TT&C/payload; V&V for mixed safety-security requirements; Certification and testing under space constraints; Formal interface contracts to avoid harmful emergent behavior
System-of-Systems & Federated Operations	Multi-operator constellations; Cross-domain data sharing; Inter-organisational trust, SLAs, and liability; Mission handover and coalition operations
Policy, Governance, and Standards	International standards and interoperability (CCSDS/DTN/PQC-ready); Threat-intel sharing; Export controls and proprietary interfaces limiting instrumentation and reproducibility; Liability across commercial/government assets; Secure software/hardware supply-chain practices
Emerging Technologies and Trends	Digital twins for vulnerability testing (declared fidelity/limits); Dynamic payloads and modular experimentation; AI-assisted traffic management; Quantum communications; Confidential computing/TEEs
Aerospace Program-matics & Infrastructure	Launcher/ground infrastructure dependencies; Rideshare/hosted-payload risks; Power/propulsion/peripherals constraints; Mission cost/entry barriers
Unique Space Environment Challenges	Radiation, thermal cycling, micrometeoroids; Orbital dynamics and contact geometry; Irretrievability and limited servicing; Communications delay/intermittency; Environment-induced ambiguity that complicates attribution

5 Working Groups

5.1 Seminar Organization

In the afternoon of the first day of the seminar, the participants decided to divide the working groups into four teams: Prepare/Attack, Protect, Detect, and Respond. Each group started by identifying the top three pressing issues within its respective group based on the identified open problems in Table 1.

5.2 Working Group on Attack/Prepare

The ATTACK/PREPARE group opened by enumerating blockers to credible attack research against space systems. Three roots emerged. First, an evidence deficit: there are no trustworthy, shareable attack datasets aligned with mission context. Second, legal and contractual barriers: export controls, proprietary interfaces, and vendor NDAs limit sharing, instrumentation, and reproducibility. Third, a fidelity gap: the coupling of ground-link-space timing, radios, and mission logic means generic IT labs and pure simulation fail to capture the observables that matter for adversary study.

On that basis, the group specified what a study environment must produce: (i) pre-condition metadata (architecture, communications characteristics, software/firmware, and access-control surfaces), (ii) nominal mission traces for the same surfaces, and (iii) aligned attack traces. Existing technique catalogues do not provide worked implementations with synchronized metadata and traces; only an integrated environment can generate all three coherently across the chain.

The resulting outcome was a testbed/digital-twin workflow rather than a stand-alone simulator. Fidelity is chosen by the question under test; hardware-in-the-loop is used where it changes observables (e.g., C&DH, EPS, ADCS, radio/SDR paths); environmental context (eclipses, radiation belts, Doppler) is modeled to shape timing and errors; and minimal instrumentation is standardized so different teams can build threat-driven twins yet still yield comparable datasets. Short-term actions recorded by the group include surveying existing flatsats and ranges, defining the instrumentation and data schemas up front, and packaging adversary scenarios mapped to space-relevant TTP catalogues for reproducible execution.

5.3 Working Group on Detect

The DETECT group treated spacecraft and ground detection as a coupled problem under sparse observability and DTN. It catalogued the data required for practical methods, provenance-rich command/telemetry (counters, origin, timing, mode), process and file events, memory-integrity evidence, and internal bus/link signals, and drafted machine-actionable examples to enable sharing (e.g., command-origin deviations during specific modes; star-tracker reference-hash mismatches; mode-file/process inconsistencies). The group documented why conventional IDS tooling underperforms on mission traffic: freshness and ordering are probabilistic, error bursts and Doppler shifts mimic adversarial behaviour, and semantics are mission-specific.

Outcomes included evaluation expectations and forensic-readiness requirements. Detectors should condition scoring on physics (SAA passages, eclipses, space-weather episodes), separate environmental from adversarial false alarms, and be compared on corpora created in the

ATTACK/PREPARE testbed. Resilience must not erase evidence: corrections and scrubbing events are to be provenance-preserving; anomaly journals must survive safe-mode resets; and downlink strategies must support trickle transmission over multiple passes.

Finally, the group recorded a range design specifically for detection research: start from clear objectives (onboard vs. ground focus), derive fidelity from those objectives, instrument at the points that expose adversary behaviour, and include 0-constellation and deep-space cases so identity, routing, and delay artefacts are exercised in a controlled way.

5.4 Working Group on Protect

The PROTECT group concentrated on architectural measures that hold over long missions and constrained update paths. Recurrent sources of risk were distilled from rideshare/hosted-payload arrangements, evolving network topologies (DSN/DTN, cislunar), standards and legacy components, and “X-as-a-service” ground operations. The baseline recorded by the group comprises strict internal segmentation with message-level authentication/authorization, measured boot with dependable key management and crypto agility (including PQC transition planning and re-key under DTN), and *updateability as a security requirement* (A/B images, shadow execution with telemetry-backed equivalence before commit).

Legacy integration was treated directly: risk cannot be eliminated by isolation alone. The working group specified service wrappers that enforce modern controls around older radios/payloads, dependency-longevity planning and spares, and explicit end-of-life options in contracts. A closing thread addressed composability: safety and security controls must be engineered so interactions across EPS, ADCS, TT&C, and payloads are predictable, with configuration governance to prevent harmful emergent behaviour. The seminar distinction was kept explicit: resilience restores function; protection must also preserve the truth about causes.

5.5 Working Group on Respond

The RESPOND group produced an operational playbook aligned with mission assurance. Preparation and monitoring come first (simulations, validated backups, rehearsed safing procedures). Identification focuses on locating adversary activity across the ground–link–space chain, understanding mechanisms and privileges (including misuse of legitimate tooling), and protecting time-critical services. Isolation is defined as containment that keeps observability and commandability intact.

Immediate recovery proceeds by ranked options, revocation and re-keying, surgical subsystem shutdowns, and only then broader isolation while assessment continues. Longer-horizon recovery restores reachable systems and stands up replacements when assets are unreachable. The cycle closes with learning and evaluation: timelines, costs, and decisions are documented, and specific hardening actions feed back into PROTECT and DETECT. Throughout, actions are chosen to remain safe if symptoms are environmental and to reduce manipulability if they are adversarial, and all steps maintain an audit trail robust to resets and fragmented downlinks.

6 What Makes Space So Different

Addressing space cybersecurity requires a paradigm shift as the challenges are not incremental extensions of terrestrial problems, but represent a qualitative leap in complexity and nature, arising directly from the unique environment, constraints, and operational dynamics of space. At Dagstuhl, we consolidated these challenges into three foundational categories that define why cybersecurity for space assets must be treated differently:

Challenge 1: Space Environment Physical Constraints

Spacecraft operate in an environment defined by physical extremes, unlike any terrestrial system. Radiation is particularly critical: spacecraft are exposed to a complex mix of high-energy charged particles, protons, heavy ions, and electrons from solar, galactic, and extragalactic sources. While missions in Low-Earth Orbit (LEO) benefit from partial shielding, those in higher or interplanetary orbits face far more severe and sustained radiation conditions.

Radiation induces both transient and permanent effects on electronics, including bit flips, logic faults, and cumulative degradation. While these are long-recognized reliability issues, their unpredictable nature also complicates cybersecurity. A single anomaly may be environmental or adversarial, and traditional fault-tolerance techniques such as Triple Modular Redundancy (TMR) or memory checksums restore functionality without questioning causality. As a result, spacecraft may recover from a disruption yet remain blind to whether it originated from natural radiation or deliberate manipulation. This strategic ambiguity gives adversaries plausible cover: disruptions coinciding with solar flares or radiation belt passages may be dismissed as environmental, allowing targeted attacks to masquerade as background noise.

The same uncertainty extends beyond onboard systems to spacecraft communications. Space-to-ground and inter-satellite links face high latency, limited bandwidth, and intermittent availability. Corrupted packets, dropped sessions, or protocol desynchronization may result from Doppler shifts or radiation, but also from replay, delay, or spoofing attacks. In deep space, where space weather forecasting is uncertain and real-time environmental telemetry is limited, distinguishing the two is especially difficult.

Even cryptographic mechanisms are vulnerable to this ambiguity: radiation-induced bit flips in keys or entropy pools may manifest as failed authentication, broken sessions, or malformed messages, while symptoms are indistinguishable from malicious tampering. Thus, both system and communication layers face the same foundational challenge: defending against adversaries in an environment where natural effects can always provide plausible deniability.

While radiation provides the most direct cybersecurity concern, other environmental extremes reinforce the same ambiguity. Thermal cycling can shift timing margins, accelerate component aging, and degrade entropy sources or key storage, producing effects that resemble active tampering. Vacuum-driven outgassing and material fatigue, as well as sporadic micro-meteoroid or debris impacts, primarily threaten reliability but can manifest as resets, sensor drift, or link interruptions that mimic denial-of-service or integrity attacks. In contested settings, such anomalies offer adversaries plausible cover: without physics-informed diagnostics, operators may misclassify malicious interference as natural degradation.

Challenge 2: System Isolation

Space missions operate under a condition of absolute isolation: after launch, hardware can never be retrieved or replaced, and the mission must unfold with the systems committed at liftoff. Spacecraft cannot undergo hardware servicing, trusted forensic inspection, or manual reset. Their only external visibility comes through narrow telemetry channels, and any repair or mitigation must rely on pre-installed onboard logic or constrained, high-risk command uplinks from the ground. The permanence of these constraints is magnified by mission lifespans: probes such as Voyager have remained operational for nearly half a century without physical maintenance, underscoring how design choices made before launch must endure for the full mission lifetime.

A useful comparison is with industrial control systems such as chemical plants. In these settings, the process is the mission, but the cyber-physical control layer remains serviceable: controllers can be replaced, sensors recalibrated, and unit operations re-engineered during maintenance windows while the underlying process continues. By contrast, a spacecraft fuses mission and controller into a single, unreachable asset: its trajectory, sensing geometry, power and thermal envelope, and actuation topology together constitute the mission and cannot be separated from it once deployed.

These properties have concrete security implications. Assurance becomes effectively one-shot: vulnerabilities or misconfigurations that escape pre-launch detection may remain exploitable for years or decades. Monitoring and incident response are limited to the mechanisms designed from the outset. Recovery depends on autonomous mechanisms whose correctness and robustness are themselves part of the attack surface. Certification and trustworthiness, therefore, evolve differently in orbit: security is reinforced not by just periodic patching or audits, but by sustaining resilience in isolation over the mission's full operational life.

Challenge 3: Autonomy Under Extreme Latency

Even when a spacecraft remains functional, communication is constrained by distance and orbital dynamics. Deep space missions experience round-trip latencies of tens of minutes or more, and even low Earth orbit missions can encounter extended blackouts due to orbital dynamics, power constraints, or interference. In such environments, autonomy is not optional but operationally required. Yet autonomy, when combined with extreme communication delay, introduces a distinct class of security challenges.

From a security perspective, autonomy under extreme latency means the spacecraft must serve as its own guardian, at least intermittently. With no possibility for timely human verification, it must assess its state, detect attacks and anomalies, and respond to threats locally and in real-time. Traditional system monitoring mechanisms such as watchdog timers, hardware redundancy, and fail-safe control modes assume that anomalies can eventually be observed or reset through external intervention. Yet in autonomous settings, this assumption does not hold. These mechanisms respond to symptoms, not causes, and are typically agnostic to adversarial intent. A spacecraft may suffer degradation not as a result of accidental faults, but due to subtle manipulation of internal behavior. Partial corruption of command parsing, sensor fusion, or actuator logic may evade fault detection entirely while causing long-term damage. This is particularly dangerous when the degradation affects system-wide processes such as thermal regulation, power control, or attitude adjustment.

For example, a denial-of-service condition exploiting algorithmic complexity, such as a ReDoS (Regular Expression Denial of Service) attack, could induce excessive CPU or bus contention during thermally critical mission phases. If this delays or suppresses heater

activation, the spacecraft may cool below its operational threshold, preventing battery bootstrap and potentially pushing components outside of their specified tolerances. Under autonomous operation, such faults may not be correctly attributed or mitigated in time, leading to cumulative and unrecoverable subsystem degradation or a safe mode configuration that is less resilient than the nominal configuration, opening up additional attack vectors.

Historical incidents show how physical degradation, even when unintentional, can result in irreversible failure. The ROSAT satellite [4], for instance, was equipped with a highly sensitive X-ray telescope that required its detectors to remain covered when not in use. However, due to a software misconfiguration in its attitude control system, the telescope was inadvertently pointed directly at the Sun during an operational maneuver. The onboard logic failed to issue a shutdown, leading to the destruction of the sensor due to solar overexposure [4]. This incident highlights how inadequate safeguards, under autonomous conditions, can lead to catastrophic outcomes from entirely foreseeable edge cases.

In contrast, the Stuxnet malware demonstrated that adversaries can deliberately induce long-term mechanical damage while concealing intent [6]. Stuxnet targeted industrial control systems running on Siemens S7-300 PLCs used in Iran's Natanz uranium enrichment facility. The attack specifically manipulated the rotational frequency of gas centrifuges used to separate uranium isotopes. These centrifuges were designed to operate within a narrow frequency window, typically around 1,064 Hz. Stuxnet intermittently altered this frequency, forcing the centrifuges to accelerate far beyond their nominal speed (reportedly up to 1,410 Hz) and then rapidly decelerate or oscillate unpredictably. These deviations were brief enough to avoid immediate failure but frequent and severe enough to create cumulative mechanical fatigue, misalign rotor assemblies, and eventually cause bearing damage or rupture.

Looking at these two cases, we argue that autonomous security must therefore operate under conditions of incomplete information, degraded sensing, and evolving mission context. The system must reason not only about whether a behavior is faulty, but also whether it is plausible given its location, trajectory, power state, and subsystem interaction, and provide its reasoning to operators when there is a connection window available for further verification.

Additionally, extreme latency and autonomy disrupt core assumptions about identity, authentication, and message integrity. Protocols designed for synchronous or near-real-time networks, such as challenge-response, key renegotiation, or session handshakes, become infeasible. Communication delays, link outages, and high bit-error rates mean that space networks must adopt Delay-tolerant Networking (DTN) principles, where messages are asynchronously relayed, buffered, and reassembled. However, DTN conditions undermine traditional guarantees of freshness, liveness, and ordering primitives on which most terrestrial cryptographic protocols depend.

The result is a security model in which authenticity and trust are probabilistic rather than deterministic. For example, a packet received during an expected contact window, from a plausible antenna orientation, and with correct signal power may be considered more trustworthy than one that deviates from these constraints. Yet this judgment must be made onboard, in real time, with limited context, without external validation, and with all existing power and processing budget limitations.

Challenge 4: Pervasive Exposure and Asymmetric Attack Surface

Space systems possess a fundamentally different attack surface from terrestrial systems, not just in extent but in asymmetry, persistence, and observability. The attack surface spans physical, cyber, and RF domains, each with unique entry points and defense limitations. What sets this domain apart is not the existence of more vectors, but the inability to constrain, observe, or attribute many classes of attacks.

From a security perspective, spacecraft are highly integrated cyber-physical platforms with interconnected subsystems. Each of these may become an attack vector or fault amplifier. For example, access to a thermal management controller or a fault handler may provide indirect control over avionics or memory protection logic. Many existing spacecraft architectures use shared communication buses and unsegmented internal channels, meaning that subsystems can exchange messages over a common interface without isolation or message-level authentication. These designs were historically justified by the assumption that spacecraft are physically inaccessible to adversaries, and therefore internal communications would remain trustworthy and uncontested. This assumption no longer holds in a world where remote code execution, protocol exploitation, or malicious payload injection can be initiated from Earth.

Moreover, modern spacecraft increasingly incorporate commercial off-the-shelf components, third-party software, and open-source libraries [7]. These introduce opaque and often unverified dependencies into mission-critical systems. Vulnerabilities in telemetry handlers, decompression modules, or firmware may go unnoticed until operational deployment, and most spacecraft cannot fully patch or revalidate such components post-launch [8, 3].

Additionally, the most persistent and unavoidable exposure lies in the continuous use of radio frequency or optical interfaces for communication. Spacecraft must maintain always-on RF or optical interfaces for telecommands and data operations. These channels are predictable in time and frequency and inherently exposed.

Additionally, the ground segment introduces a systemic and often underestimated vulnerability. Ground control software, mission scheduling systems, and telemetry storage infrastructure can be compromised to influence spacecraft indirectly. The 2022 Viasat KA-SAT attack is a relevant example [1], where satellite communications were disrupted at scale without modifying satellite firmware, illustrating that terrestrial infrastructure remains a viable entry point.

The asymmetry is further exacerbated by the imbalance between attackers and defenders. Attackers can observe orbital paths, predict visibility windows, and time attacks precisely. Defenders, in contrast, often operate with outdated or intermittent telemetry, lack real-time access, and have little visibility into the presence or behavior of an attacker.

A new facet of this asymmetry is the use of spacecraft to target or probe other spacecraft directly. Nation-state actors have increasingly conducted proximity operations, where one satellite shadows or approaches another to observe its behavior, gather RF emissions or assess response patterns. These interactions, often described as on-orbit reconnaissance or Rendezvous and Proximity Operations (RPO), blur the line between surveillance and preparatory attack, especially when used to map out operational vulnerabilities or test defense thresholds. For example, in 2020, a Russian satellite (Kosmos 2542) maneuvered close to a U.S. military satellite (USA 245), prompting public warnings from U.S. Space Command about potential antisatellite behavior [5]. These actions are rarely transparent, difficult to verify, and often designed to remain below escalation thresholds. As space becomes more contested, inter-spacecraft operations may evolve into a common vector, requiring new security models that consider not only terrestrial threats but also orbital adversaries.

6.1 Summary

Taken together, these challenges show that space cybersecurity is fundamentally unlike any other cyber-physical security problem. Other domains may share fragments of these issues, but only in space do they converge inseparably and persist for the entire mission

lifetime. Harsh physical environments, absolute isolation, extreme communication delays, and pervasive exposure do not just complicate defense; they redefine it. The result is an attack surface that is broader, more persistent, and less observable than in any terrestrial system, forcing defenders to treat incomplete information, degraded sensing, intermittent trust, and adversarial uncertainty as normal operating conditions.

This conclusion reflects the consensus of the Dagstuhl Seminar, where more than forty experts from the space and cybersecurity domains, including specialists in cyber-physical systems, concluded that space constitutes a qualitatively different security environment. Meeting these challenges requires more than adapting terrestrial techniques: it demands fundamentally new approaches that embed physics-informed reasoning, resilience without servicing, and autonomy designed to operate securely under uncertainty for decades at a time.

7 Future Work and Challenges Ahead

The rapid evolution of the space domain, characterized by the proliferation of commercial mega-constellations, increasing autonomy, the steep increase in data throughput capacity, and the extension of terrestrial networking paradigms into orbit, presents a complex and dynamic cybersecurity landscape. The discussions at the Dagstuhl Seminar crystallized a consensus that future research and development must move beyond traditional security research and consider the specificities of the space environment and its constraints. This section outlines the critical frontiers and formidable challenges identified by the seminar participants during the fourth and fifth days of the seminar, providing a roadmap for the academic, industrial, and governmental efforts required to secure the future of space operations. The challenges involve and merge many disciplines in security research, from foundational cryptographic transitions to the complexities of autonomous defense, secure hardware design, cyber-physical resilience, and the establishment of robust international governance.

7.1 Secure Space Communications and Encryption in the Quantum Era

The looming threat of fault-tolerant quantum computers capable of breaking current public-key cryptography (e.g., RSA, ECC) using algorithms like Shor's necessitates a fundamental overhaul of cryptographic systems for space. This is not a distant, theoretical concern but an urgent operational reality. Given the long lifecycles of spacecraft, which can operate for 15–20 years, and the general impossibility of post-launch hardware upgrades, a proactive transition to quantum-resistant security is not merely advisable but mission-critical. Systems launched today with vulnerable cryptography could have their communications intercepted and stored, ready to be decrypted by a future quantum computer, a “harvest now, decrypt later” attack that poses an unacceptable risk to long-term national security and commercial intellectual property. This challenge bifurcates into two primary, and often complementary, research avenues: the near-term deployment of Post-Quantum Cryptography (PQC) and the long-term, ambitious development of Quantum Key Distribution (QKD) for space applications.

7.1.1 The Duality of Quantum Key Distribution (QKD) and Post-Quantum Cryptography (PQC)

The path toward quantum-resistant space systems is defined by the distinct characteristics, trade-offs, and timelines of QKD and PQC. Understanding this duality is fundamental to developing a coherent security strategy.

7.1.1.1 The Promise and Peril of QKD

Quantum Key Distribution (QKD) represents a paradigm shift in secure communications. Its security guarantee is not based on the presumed computational difficulty of a mathematical problem, but on the fundamental laws of quantum physics, such as the no-cloning theorem and Heisenberg's uncertainty principle. This provides information-theoretic security, meaning that an eavesdropper's attempt to intercept and measure the quantum states (e.g., polarized photons) used to generate a key would inevitably disturb the system, revealing their presence to the legitimate parties. In theory, this makes the key exchange impervious to any future advances in computing, including quantum computers.

The feasibility of this technology for space has been convincingly demonstrated. Experiments, most notably China's Micius satellite launched in 2016, have successfully established space-to-ground and inter-satellite QKD links, distributing secure keys over distances exceeding 1,200 km. These missions proved that satellite-based QKD can overcome the distance limitations of terrestrial fiber-optic QKD, which suffers from exponential signal loss, and could form the basis of a future global quantum internet.

However, despite its theoretical promise, QKD faces immense practical challenges that currently limit its widespread deployment. Firstly, QKD is only a partial solution; it secures the distribution of a symmetric key but does not provide authentication. The source of the QKD transmission must be authenticated using classical methods, which today means relying on pre-placed keys or, ironically, PQC, making QKD vulnerable to man-in-the-middle attacks if the authentication layer is weak. Secondly, QKD requires specialized, costly, and inflexible hardware, such as single-photon detectors and precise pointing systems, which are difficult to integrate into satellite buses and impossible to upgrade post-launch. Thirdly, free-space optical links are susceptible to disruption from atmospheric conditions like turbulence and cloud cover, and current key generation rates remain far too low for high-bandwidth applications, with some demonstrations yielding only a few bits of secure key per satellite pass. Finally, the theoretical security of the protocol can be undermined by practical side-channel attacks that exploit imperfections in the physical hardware, and its inherent sensitivity to any disturbance makes it highly susceptible to denial-of-service (DoS) attacks. These significant limitations have led to a cautious stance from security bodies like the U.S. National Security Agency (NSA), which currently does not recommend QKD for securing National Security Systems until these fundamental implementation and security validation challenges are overcome.

7.1.1.2 The Pragmatism of PQC

Post-Quantum Cryptography (PQC) offers a more immediate and pragmatic path to quantum resistance. PQC algorithms are classical, meaning they can run on existing computer hardware, but are based on mathematical problems, such as those found in lattice-based, hash-based, or code-based cryptography, that are believed to be computationally infeasible to solve for both classical and quantum computers.

A major driver for PQC adoption is the progress in standardization. The U.S. National Institute of Standards and Technology (NIST) has completed its multi-year competition and has finalized the first standards for PQC algorithms. These include CRYSTALS-Kyber (standardized as ML-KEM) for key encapsulation and CRYSTALS-Dilithium (standardized as ML-DSA) for digital signatures, providing a vetted foundation for industry to build upon. This has spurred active development, with projects already underway by organizations like the European Space Agency (ESA) to design and implement PQC-based cryptographic systems for securing satellite telecommunication applications, particularly command and control links. Furthermore, space standardisation organisations such as the Consultative Committee for Space Data Systems (CCSDS), are adopting the NIST recommendations already.

PQC is not, however, a simple drop-in replacement for current cryptographic standards. Its security remains computational, not absolute, and the field is still maturing. Several PQC candidate algorithms, including some that reached advanced stages of the NIST process, have been broken by subsequent cryptanalysis using classical computers, highlighting the potential for future vulnerabilities to be discovered. For space systems, the most pressing challenges are practical. PQC algorithms often require significantly larger key sizes and signatures, and are more computationally intensive than their classical counterparts. This poses a substantial problem for the highly constrained Size, Weight, and Power (SWaP) environment of satellites, where processing power and bandwidth are scarce resources. Furthermore, the harsh radiation environment of space introduces the risk of Single Event Upsets (SEUs), bit-flips caused by cosmic rays, which could corrupt complex PQC calculations, potentially leading to authentication failures or security breaches. This makes the research and development of fault-tolerant PQC implementations (which would not necessarily need to depend on expensive radiation-hardened components), likely involving specialized hardware and error-correcting codes, a critical area for future work.

7.1.1.3 A Hybrid and Risk-Stratified Future

The ongoing debate is not a simple choice of “QKD vs. PQC.” Rather, the evidence points toward a future where the two technologies are integrated into a hybrid, risk-stratified architecture. PQC is the only viable path for achieving broad crypto-agility in the near term. Its software-based nature allows it to be deployed on existing and new systems to secure the vast majority of commercial and tactical communications. It will become the workhorse of space cryptography.

However, PQC alone cannot defend against the “harvest now, decrypt later” threat for data that requires confidentiality for decades or longer. This is where QKD finds its crucial niche. A hybrid model is therefore necessary. In this model, PQC provides the robust, authenticated channel required for QKD to operate securely, protecting it from man-in-the-middle attacks. QKD, in turn, provides an information-theoretically secure method for distributing keys for the highest-value, strategic communication links where the cost and complexity are justified. This could include, for example, securing command links for national security satellites, establishing a secure key-exchange backbone for deep space missions, or protecting critical diplomatic communications.

This leads to a tiered cybersecurity model for space assets. High-value, state-owned strategic assets may be equipped with expensive, hardware-based QKD systems for ultimate long-term security. In contrast, large commercial constellations, where cost is a primary driver, will rely on more agile but computationally-based PQC. This inevitable stratification will create profound new challenges for interoperability between different security domains, for the development of international policy, and for the creation of standards, as a single definition of “secure” will no longer apply universally across the space ecosystem.

Table 2 contrasts Quantum Key Distribution (QKD) with Post-Quantum Cryptography (PQC) for satellite links. QKD provides information-theoretic security but demands specialized optics and precise pointing, yields limited/fragile key rates, and supplies key distribution only (no native authentication). PQC relies on computational hardness yet is software-deployable on existing hardware, already standardized (e.g., ML-KEM/ML-DSA), and supports both key establishment and digital signatures. In practice, PQC is the default for securing command and control, while QKD is complementary for bulk key pre-distribution where optical links and SWaP budgets permit.

■ **Table 2** Comparative Analysis of QKD and PQC for Satellite Communications.

Attribute	Quantum Key Distribution (QKD)	Post-Quantum Cryptography (PQC)
Security Basis	Information-theoretic, based on laws of physics; provides forward secrecy against future computational advances.	Computational, based on hardness assumptions believed to resist quantum attacks.
Maturity (TRL)	Low to medium; experimental demonstrations (e.g., Micius) are successful but not yet mature for broad space deployment.	Medium to high; NIST standards exist (e.g., ML-KEM, ML-DSA); space-grade implementations are in development.
Implementation	Hardware-intensive; requires single-photon sources/detectors and precision pointing; not software deployable.	Software-based; runs on existing hardware, deployable via firmware or software updates.
Primary Function	Key distribution only; authentication must be provided separately.	Key encapsulation and digital signatures for authentication.
Suitability for C&C	Challenging; low key rates and DoS susceptibility limit use for critical command links; useful for bulk key pre-distribution.	High; suitable for securing command and control links due to software nature and authentication support.
Key Vulnerabilities	Side-channel attacks on optics/electronics; DoS from atmospheric or malicious interference; lack of built-in authentication.	Potential future cryptanalytic breaks; implementation bugs; software side channels.
SWaP Impact	High; adds dedicated hardware with mass, power, and volume penalties.	Moderate; higher compute and memory than classical crypto, but no new hardware subsystem required.
Fault Tolerance	Highly sensitive to disturbances, atmospheric effects, and pointing errors.	Complex algorithms susceptible to SEUs in radiation, requiring fault-tolerant design.

7.1.2 Securing High-Bandwidth Optical and RF Communications

The transition to high-throughput communication links, particularly optical/laser inter-satellite links (ISLs) and space-to-ground connections, is a key enabler for future space services, from global broadband to massive deep-space science and Earth observation downlinks. While offering unprecedented bandwidth, these links also present high-value targets for sophisticated adversaries seeking to conduct eavesdropping, jamming, or spoofing attacks. Securing these communications on all layers (physical, link, and network) is a critical challenge.

Future research must focus on developing advanced defense mechanisms tailored to the unique physics of these channels. For optical communications, this means moving beyond defenses designed for RF systems. Research is needed into techniques that can distinguish

malicious jamming or spoofing from natural environmental interference, such as atmospheric scintillation, which can cause similar signal degradation. AI-based signal analysis, capable of learning the subtle signatures of both atmospheric conditions and deliberate attacks, presents a promising avenue for robust detection.

Furthermore, as satellites become nodes in large, interconnected mesh networks, resilience cannot depend on a single point-to-point link. Future work must involve the design of secure and resilient routing protocols for these large-scale optical constellations. Such protocols must be able to detect a compromised or failed node and dynamically re-route traffic through trusted paths, ensuring network availability and graceful degradation of service rather than catastrophic failure. This also requires the development of novel signal transformation and Transmission Security (TRANSEC) protocols that enhance resilience against interception and manipulation at the lowest layers of the communication stack.

7.1.3 Authenticated and Resilient Command & Control (C&C)

The command and control (C&C) link is the umbilical cord to a spacecraft; its compromise can lead to the partial or total loss of the asset. Securing this link requires a multi-layered, end-to-end approach that extends from the operator at a control center to the satellite bus in orbit.

A key area for future work is the development of next-generation onboard defenses. This involves moving beyond static defenses to adaptive, space-specific firewalls and on-board Intrusion Detection Systems (IDS). These systems must be capable of operating effectively within the severe resource constraints of a satellite's onboard computer, analyzing command flows and telemetry for signs of unauthorized activity.

Equally important is ensuring the integrity of the entire C&C chain. An attack is just as likely to originate from a compromised ground segment as it is to target the space-to-ground link. Therefore, cryptographic security, likely based on the new PQC standards, must be implemented and rigorously verified across all components of the system, including ground station software, network infrastructure, and operator terminals. This ensures end-to-end trust and prevents an attacker from bypassing space link encryption by compromising a vulnerable terrestrial component. It also needs to include formal security proofs for space-link communication security standards such as the CCSDS Space Data-Link Layer Security (SDLS) standard family.

As the ground segment state-of-the-art moves more into the shared infrastructure approach (e.g. ground station as a service, ground segment as a service) and multi-mission support (one ground segment for many missions) ground segment resilience remains a key aspect with elements such as multi-mission zero-trust architectures taking a major role in research.

7.2 AI-Driven and Autonomous Space Cybersecurity

As space missions venture further from Earth into deep space and constellations grow to encompass thousands of interconnected nodes, the operational paradigm is fundamentally changing. The signal propagation delays, which can range from minutes to hours for deep space missions, combined with the sheer scale and complexity of mega-constellations, make direct human-in-the-loop control for cybersecurity functions untenable. In this new era, autonomy is not an option but a necessity. The central challenge for the research community is to develop Artificial Intelligence (AI)-driven systems that can autonomously detect, reason about, and respond to cyber threats in real-time. The ultimate goal is to create self-defending space assets capable of ensuring their own survival and mission success without constant human intervention.

7.2.1 AI for Advanced Intrusion Detection and Response

Traditional security tools, such as signature-based Intrusion Detection Systems (IDS), are fundamentally reactive. They are effective at identifying known threats but are easily bypassed by novel, zero-day attacks. The future of on-orbit threat detection lies in AI's ability to move beyond pattern matching to behavioral analysis. AI and Machine Learning (ML) models can be trained to establish a high-fidelity baseline of a satellite's normal operations, encompassing everything from bus telemetry and power consumption patterns to network traffic and payload activity, and then identify anomalous deviations that could indicate a compromise.

Key research directions in this area include the implementation and refinement of various AI/ML models tailored for the space environment. Unsupervised learning models like Isolation Forests and Deep Autoencoders are well-suited for anomaly detection in network traffic, while sequence-aware models like Long Short-Term Memory (LSTM) networks can detect deviations in time-series data, such as user activity logs or command sequences. Another critical application is using AI for real-time analysis of the RF spectrum. By learning the characteristics of legitimate signals, an AI system can detect and classify sophisticated jamming and spoofing attacks, providing a layer of cyber-physical defense that bridges the digital and physical domains.

However, deploying these AI models effectively presents significant challenges. A primary hurdle is managing the high rate of false positives, which can overwhelm operators and lead to alert fatigue. This is exacerbated by the inherent data imbalance in cybersecurity, where malicious events are rare compared to normal operations. Furthermore, the computational cost of complex deep learning models can be prohibitive for SWaP-constrained satellites. Finally, the "black box" nature of many AI models poses a challenge for trust and verification; operators need explainable AI (XAI) techniques to understand why an alert was triggered before they can confidently act on it.

7.2.2 Self-Defending and Self-Healing Spacecraft

Detecting a threat is only the first step; a system must also be able to respond effectively to mitigate the threat and restore its core functions. This concept of cyber resilience, the ability to withstand, operate through, and recover from an attack, is the foundation of autonomous cybersecurity.

Future work must focus on developing Tactical Autonomous Systems (TASS), which are AI-driven agents capable of executing pre-defined defensive "playbooks" in response to a detected intrusion. Upon identifying a threat, a TASS could autonomously take action, such as isolating a compromised subsystem from the main bus, rerouting critical data through trusted communication paths, or disabling non-essential functions to preserve the primary mission. In the context of large constellations, this extends to cooperative defense, where satellites can share threat intelligence and defensive postures with trusted peers. This allows the entire constellation to "learn" from an attack on a single node and collectively adapt its defenses, creating a resilient, herd-like immunity.

The ultimate goal is the creation of self-healing architectures. This involves researching systems where a satellite can not only detect and isolate a threat but also autonomously purge malware, restore critical software from a secure, read-only backup, and even apply security patches to remediate the underlying vulnerability, all without human intervention. This capability is especially critical for long-duration missions into deep space, where the extreme communication delays make interactive recovery impossible.

7.2.3 AI for Predictive Maintenance and Fault Tolerance

The line between a system fault and a cyberattack is often blurry. A physical component failure could be a precursor to a cyberattack, a vulnerability an attacker might exploit, or even the direct result of a malicious command. AI can play a crucial role in bridging the gap between traditional Fault Detection, Isolation, and Recovery (FDIR) and cybersecurity.

By applying ML models to analyze historical and real-time telemetry, AI systems can perform predictive maintenance, forecasting component failures before they occur. An unexpected prediction of failure in a healthy component could serve as an early indicator of a subtle, ongoing cyberattack. This fusion of FDIR and cybersecurity enhances overall mission assurance.

A key enabling technology for this is the concept of a digital twin. By creating a high-fidelity, physics-based virtual replica of a satellite and its environment, operators can run simulations that are impossible to conduct on the real asset. Enhanced with Generative AI, these digital twins can be used to simulate a vast range of both physical fault and cyberattack scenarios. This provides an invaluable, safe environment for training and validating AI-based detection and response models before they are deployed on the actual spacecraft, significantly improving their reliability and effectiveness. A critical challenge within this domain is the development of a trusted, automated system for deploying firmware and software patches to an orbiting satellite. Such a system is essential for both fixing bugs and remediating vulnerabilities, but it also represents a prime target for a supply chain attack, where an adversary could inject malicious code into a seemingly legitimate update. Securing this automated patching pipeline is a major research challenge.

7.2.3.1 The Double-Edged Sword of AI

While AI is a powerful enabler for autonomy, it presents a fundamental paradox: the very tools used to manage complexity introduce new, opaque attack surfaces. AI models, particularly deep neural networks, often behave as inscrutable “black boxes” whose performance is brittle under unforeseen conditions and vulnerable to adversarial manipulation, from data poisoning during training to evasion at inference. Critically, traditional software assurance methods like code review and static analysis are insufficient for these learned systems. This creates a dangerous safety-security coupling in autonomous systems like spacecraft, where a security failure (e.g., a spoofed sensor reading) can cascade into an unsafe control action with no immediate human oversight. Countering this requires a security-by-construction approach, integrating multiple layers of protection: formally specified operational envelopes to constrain actions; runtime monitors grounded in physical laws; verifiable provenance for all training data; and signed, versioned models to prevent unauthorized modification.

Table 3 summarizes how common AI/ML methods map to autonomous space-cyber tasks, from anomaly detection and self-healing control to predictive maintenance, constellation-level defense, and RF signal analysis. The techniques promise mission awareness and faster response, but face practical hurdles: scarce high-fidelity data, onboard SWaP limits, safety/verification of autonomous actions, and robustness to interference and data poisoning. Designing for explainability, forensic readiness, and secure digital-twin workflows is essential for reliable deployment.

7.2.3.2 The Rise of Agentic Adversaries

Looking ahead, attackers can field *agentic* AIs that plan, probe, and adapt with minimal human input: autonomously discovering protocol flaws, staging multi-step RF/cyber campaigns, synthesizing spoofed telemetry consistent with orbital dynamics, or timing actions to

exploit DTN delays and attribution ambiguity. This lowers the barrier to persistent, tailored operations against both spacecraft and ground segments. Meeting agentic offense requires agentic defense. Beyond static detectors, we need autonomous, policy-bounded defenders that (i) reason over multi-modal evidence (telemetry, RF, ephemerides, power/thermal), (ii) enact deception, and (iii) recover safely. Practical building blocks include (but are not limited to) *agentic honeypots* (emulated subsystems, decoy services, and DTN honeynodes that absorb and fingerprint probes), moving-target defenses (keying, routing, and software diversity scheduled within power/thermal budgets), physics-aware plausibility filters (rejecting commands or state transitions that violate dynamics), and closed-loop response playbooks with human-on-the-loop authority. Integration should be split: lightweight, fail-safe agents onboard (SWaP-bounded, with hard limits and kill-switches) for fast containment; heavier agents on the ground and within digital twins for red teaming, hypothesis testing, and model retraining. All agents must ship with verifiable policies, audited action histories, and secure update/provenance channels so that an attacker cannot turn the defender into an amplifier.

■ **Table 3** AI/ML Techniques for Autonomous Space Cybersecurity Tasks.

Cybersecurity Task	AI/ML Technique	Primary Function	Key Challenges
Anomaly Detection and Intrusion Response	Deep autoencoders; Isolation Forests; LSTMs	Learn a baseline of normal telemetry/network behavior and flag significant deviations as potential intrusions.	High false positives; need for high-fidelity training data; model explainability; SWaP limits for onboard inference.
Self-Healing and Autonomous Defense	Reinforcement learning; Tactical Autonomous Systems (TASS)	Isolate subsystems, reroute traffic, or disable non-essential functions to neutralize threats and restore functionality.	Reward design; safety during learning; computational cost; formal verification of autonomous actions.
Predictive Maintenance and Fault Tolerance	Supervised learning (SVM, Random Forest); Digital twins	Predict component failures from historical/real-time data and simulate faults/attacks in a virtual replica.	Distinguishing natural faults from malicious actions; twin fidelity; securing the twin itself.
Cooperative Swarm Defense	Federated learning; Swarm optimization	Train shared models across a constellation without sharing raw data and coordinate defensive maneuvers.	Communication overhead; non-IID data; robustness against poisoning from compromised nodes.
RF Signal Analysis	CNNs; Autoencoders	Detect, classify, and localize jamming/spoofing via raw spectrum pattern analysis.	Disentangling interference vs. attacks; real-time processing; large, diverse datasets.

7.3 Secure-by-Design in Space System Hardware and Software

For decades, the prevailing security model for space systems focused on protecting the communication link, treating the satellite itself as a trusted “black box” operating within a secure perimeter. This assumption is now dangerously obsolete. The advent of software-defined satellites, the increasing complexity of global supply chains, and the demonstrated potential for on-orbit malware necessitate a fundamental shift in philosophy. The principles of “secure-by-design” and “secure-by-default,” championed by bodies like the U.S. Cybersecurity and Infrastructure Security Agency (CISA), must be adopted. Security can no longer be an afterthought or a feature to be added on; it must be a foundational requirement, embedded into every layer of the system’s hardware and software from the initial design phase.

7.3.1 Hardware-Rooted Security and Trust

Software-only security measures are inherently vulnerable because they can be bypassed if the underlying hardware or boot process is compromised. To build a truly trustworthy system, trust must be anchored in immutable hardware. This applies to both the platform and payload hardware.

A critical technology for achieving this is the Trusted Platform Module (TPM). A TPM is a dedicated, tamper-resistant microcontroller that provides a hardware root of trust for critical security functions. Adapting and qualifying TPMs for the rigors of the space environment is a key area for future work. A space-grade TPM could provide a secure foundation for a satellite’s entire software stack by enabling a secure boot process, which cryptographically verifies the integrity of each piece of software before it is loaded, from the bootloader to the operating system and flight application. This ensures that only authorized, untampered code can execute. Furthermore, a TPM can provide secure key storage, protecting critical cryptographic keys from being extracted by software-based attacks, and can perform attestation, allowing the satellite to prove its identity and software state to a ground station in a cryptographically verifiable way. Research into using Physically Unclonable Functions (PUFs), such as those based on ring oscillators, can enhance attestation by creating unique, unclonable hardware “fingerprints” for each satellite.

For more computationally intensive cryptographic operations, such as those required by PQC algorithms, dedicated Hardware Security Modules (HSMs) or crypto-accelerator chips are necessary. While common in terrestrial data centers, the challenge lies in developing radiation-hardened, low-power versions of these technologies that can survive and operate reliably in the space environment.

7.3.2 Lightweight and Verifiable On-Orbit Systems

The proliferation of small satellites, particularly CubeSats, introduces a different set of challenges. These platforms have extreme SWaP constraints, which often preclude the use of traditional, resource-heavy security solutions designed for larger satellites. Securing these systems requires innovation in lightweight and efficient security.

A major research focus is the design of lightweight Intrusion Detection and Prevention Systems (IDS/IPS) tailored for resource-constrained embedded systems. This involves developing lightweight ML models, optimizing feature selection to reduce computational load, and creating distributed architectures where complex analysis and model training are offloaded to the ground segment, while a smaller, efficient inference engine runs on the satellite itself.

Another powerful approach for ensuring security in critical systems is the application of formal methods. These are mathematically rigorous techniques used to specify and verify the properties of a system. By creating a formal model of critical flight software, it is possible to mathematically prove the absence of entire classes of vulnerabilities, such as buffer overflows or race conditions, and to verify that the software behaves exactly as specified under all conditions. While historically labor-intensive, advances in tools like model checkers and SMT solvers are making formal methods more accessible. The key challenge is scaling these techniques to handle the ever-increasing complexity of modern flight software.

7.3.3 Cybersecurity for On-Orbit Servicing, Assembly, and Manufacturing (OSAM)

The emergence of OSAM and hosted payload business models fundamentally alters the security paradigm. A satellite is no longer a static, monolithic asset. Instead, the space environment is becoming a dynamic, physically interactive, and potentially multi-tenant ecosystem. This dramatically expands the attack surface and introduces novel threat vectors.

Securing robotic OSAM missions is a primary concern. An attacker who compromises the command link or autonomous logic of a robotic servicer could turn a repair mission into a deliberate kinetic attack, using the servicer to physically damage or de-orbit a target satellite. Future work must focus on securing these robotic operations, including robust authentication, encrypted command links, and verifiable autonomous logic.

These missions also create complex challenges for trust and access control. A typical servicing mission may involve multiple independent entities: the servicer owner, the client satellite owner, and potentially a third-party payload owner. This necessitates the development of new security frameworks for managing trust and access control in these multi-party interactions, including secure protocols for rendezvous, proximity operations, docking, and data exchange.

Finally, the concept of “Space-as-a-Service,” where satellite operators host third-party application code or payloads, requires robust security measures. Future research must focus on creating secure containerization or virtualization environments on satellites. These “sandboxes” must provide strong isolation to prevent a compromised payload from affecting the host satellite bus, other payloads, or accessing data it is not authorized to see.

7.3.3.1 The Supply Chain as the New Perimeter

While on-orbit security technologies like PQC and AI are critical, they can be rendered moot if a system is compromised before it ever reaches orbit. The most immediate and insidious threat vector facing the space industry today is the supply chain. A “secure-by-design” philosophy is meaningless if the components used in that design are already malicious. This elevates Cybersecurity Supply Chain Risk Management (C-SCRM) from a simple procurement issue to a primary national security challenge for space.

The logic is straightforward. Space systems are assembled from a complex, global supply chain of hardware and software components, many of which are Commercial-Off-The-Shelf (COTS) to reduce costs. An adversary can target any point in this long and often opaque chain, from injecting malicious logic into a microchip at a foundry, to inserting a backdoor into open-source software, to tampering with a component during integration. This means a satellite could be launched with a hidden vulnerability or a dormant backdoor already embedded deep within its hardware or software. Such a compromise would completely bypass all link-level encryption and on-orbit defenses, waiting to be activated by an attacker at a time of their choosing. This threat is explicitly recognized as a key concern in foundational policy documents like U.S. Space Policy Directive-5.

This reality demands a “zero-trust” approach to the supply chain itself. Future work must prioritize the development of technologies and policies to ensure the integrity of components from fabrication to launch. This includes developing and mandating cryptographically signed and verifiable hardware and software bills of materials (HBOM/SBOM), establishing programs to source critical components from trusted and vetted suppliers, and using comprehensive frameworks like the NIST Cybersecurity Framework to continuously assess and manage supply chain risk throughout a program’s lifecycle, not just as a one-time check. Ultimately, assuming the supply chain can be compromised, hardware-rooted security technologies like TPMs and secure boot become the final and most critical line of defense, providing a mechanism to detect and prevent the execution of unauthorized, malicious components that may have been inserted during manufacturing. Finally, this would need to be complemented with behavior monitoring to detect changes and the use of a placed backdoor.

7.4 Cyber-Physical Resilience for Multi-Domain Missions

Cyberattacks against space systems are not merely data breaches; they are attacks on physical assets with tangible, real-world consequences. A successful cyberattack could be used to manipulate a satellite’s propulsion system to alter its orbit, disable a critical Earth observation sensor during a natural disaster, or even command a satellite to perform a maneuver that causes a collision, generating a cloud of orbital debris that threatens all space activities. Therefore, future cybersecurity research must focus on the concept of cyber-physical resilience, ensuring mission continuity even when under attack, and deeply integrating cyber defense with our physical understanding of the space environment.

7.4.1 Integrating Cybersecurity with Space Situational Awareness (SSA)

Cybersecurity and Space Situational Awareness (SSA), the practice of tracking and characterizing objects in orbit, are not considered as correlated domains so far. However, a clear relationship exists and needs to be assessed further. A cyberattack can have direct physical manifestations, e.g., an unexpected maneuver or change in RF emissions, and conversely, compromised SSA data can be used as a weapon to enable a cyber or physical attack. Furthermore, cybersecurity can be used as a tool to solve the question of maneuver accountability in case of identified collision risk between two assets that are owned by different stakeholders.

Future work must focus on breaking down these silos. This requires developing a capability for “cyber-informed SSA,” where cyber threat intelligence is used to guide physical monitoring. For instance, a security alert indicating a potential compromise of a satellite’s command and control system should automatically trigger increased tasking of ground-based telescopes and radars to monitor that specific satellite for any anomalous physical behavior. The reverse is also true. “SSA-informed cyber defense” would use physical data as a potential indicator of a cyberattack. For example, the unexpected close approach of an unknown or non-communicative object could trigger a heightened state of cyber monitoring on the high-value asset being approached.

Furthermore, the SSA data ecosystem itself, from the global network of sensors to the data fusion centers and the operators, is a prime target for attack. If an adversary can manipulate the data that operators rely on to understand the space environment, they can cause confusion, hide their own activities, or even induce an operator to perform a disastrous “collision avoidance” maneuver against a phantom object. Research is needed to secure this entire ecosystem against data manipulation and spoofing, ensuring that operators have a trusted, verified picture of the space domain.

7.4.2 Defense Against Autonomous Swarm Threats

The miniaturization of satellites and advances in autonomous coordination are enabling the development of satellite swarms. While these swarms have many beneficial applications, they also represent a novel and potent threat. An adversarial swarm of small, inexpensive satellites could be used to conduct a distributed denial-of-service attack against a target's communication links, perform coordinated, multi-point jamming, or even physically harass or disable a high-value asset through proximity operations.

Defending against such threats requires new approaches. A key area for research is AI-driven swarm defense. This involves developing defensive AI systems that can detect, track, and predict the intent of an adversarial swarm using advanced sensor fusion and behavioral analysis. Beyond detection, future work must focus on designing friendly satellite swarms with inherent resilience. This includes architectures with decentralized command and control, bio-inspired coordination algorithms that allow for emergent, adaptive behavior, and the ability to maintain overall mission capability despite the loss of individual nodes to attack or failure. Such resilient architectures can provide a robust defense, capable of dynamically responding to and mitigating threats from adversarial swarms.

7.4.3 Cybersecurity for Deep Space Operations/ Solar-System Internet

Missions to deep space destinations, meaning Cis-Lunar and beyond, push the boundaries of operational complexity. These missions are defined by extreme communication delays induced by the light speed barrier, minutes to hours as well as frequent and predictable link disruptions due to orbital mechanics. These conditions render traditional, interactive cybersecurity protocols, which rely on real-time communication with an operations centre, inefficient and error-prone. For these missions, security must be highly autonomous and tolerant of prolonged delays and disconnection. In addition, because of the high cost of deep space missions, they are very often comprised of assets owned by different stakeholders that interoperate. This inherently raises questions of trust and routing priority and it requires fully interoperable and standardised secure communication protocols as well as a decentralized key management concept.

A foundational technology for this environment is Delay/Disruption Tolerant Networking (DTN) with the Bundle Protocol (BP) at its core. DTN is a store-and-forward network architecture designed specifically for these conditions, allowing data to be held at intermediate nodes, e.g., a Mars orbiter, until a forward link becomes available. A critical area of future work is to mature, standardize, and deploy the security protocols for DTN, such as the Bundle Protocol Security Protocol (BPsec), to provide robust confidentiality, integrity, and authentication services over these challenging links.

Beyond the network layer, onboard systems for deep space missions must be empowered to make security-critical decisions autonomously. A spacecraft cannot wait for hours for confirmation from Earth to respond to a threat. This requires the development of robust, pre-programmed security policies and advanced AI/ML capabilities that allow the spacecraft to independently assess a situation, such as whether to trust a new communication partner or how to respond to an anomalous sensor reading, and take appropriate action. This also extends to long-term key management, where new protocols are needed to securely manage cryptographic keys and handle credential revocation over mission durations that can span decades, all across high-latency communication links.

7.4.3.1 The Blurring of Lines Between Cyber and Kinetic

The emergence of physically capable autonomous systems in space, such as OSAM servicers and coordinated swarms, effectively erases the clear, traditional distinction between a “cyberattack” and a “kinetic attack.” This convergence of digital and physical threats creates a new and deeply challenging security landscape. A traditional cyberattack targets data or system functions, for example, through jamming or data theft. A kinetic attack involves the application of physical force, such as an anti-satellite missile.

Now, consider a scenario where an OSAM servicer’s command and control system is compromised via a cyberattack. The attacker could then command the servicer to physically grapple and damage another satellite. Is this a cyber or a kinetic attack? It is fundamentally both. Similarly, an autonomous swarm could be commanded to surround a target satellite and use its low-power thrusters to subtly alter its orbit over time, a physical effect achieved through coordinated cyber commands.

This blurring of lines has profound consequences for international law, policy, and military rules of engagement. The foundational legal frameworks for space, including the Outer Space Treaty, were not designed to address such hybrid threats. This leaves a host of critical, unresolved questions that must become priorities for legal and policy research. For example, if a commercial OSAM vehicle from nation A, servicing a satellite from nation B, is hacked by a non-state actor in nation C and subsequently damages a satellite belonging to nation D, who is liable? The current international liability and responsibility frameworks are inadequate to address such a complex, multi-party scenario. Furthermore, at what point does a malicious cyber operation against a physically capable space asset cross the threshold to be considered a “use of force” under international law? The ambiguity of these hybrid threats creates a dangerously high risk of miscalculation and escalation, where a seemingly reversible cyber intrusion could provoke an irreversible physical conflict.

7.5 Policy, Governance, and Standardization

Technological solutions, no matter how advanced, are insufficient to secure the space domain in isolation. They must be developed and deployed within a robust and coherent framework of international policy, clear governance structures, and universally adopted standards. The global, interconnected, and interdependent nature of space operations means that a vulnerability in one nation’s system can pose a direct threat to all others. Establishing this framework is a critical prerequisite for a stable and secure future in space.

A fundamental challenge is that existing international space law, most notably the Outer Space Treaty of 1967, was drafted decades before the digital age and therefore does not explicitly address cybersecurity. This has created a significant legal and policy vacuum regarding critical issues like attribution for cyberattacks, liability for damages, and acceptable norms of behavior for cyber activities in space. Future work must focus on fostering international dialogue, for example through the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS), to establish clear norms of responsible state behavior. This includes defining what constitutes a hostile act in the cyber domain and establishing clear channels and protocols for communication and de-escalation to prevent miscalculation.

Addressing the ambiguity of liability for cyberattacks is another paramount task. This will require a concerted effort to adapt principles from existing treaties, such as the state responsibility and absolute liability concepts from the Outer Space Treaty and the Liability Convention, to the unique context of the cyber domain. While legally and technically complex, establishing clear accountability is essential to deter malicious activity.

Alongside policy development, the promotion of universal technical standards is crucial for interoperability and baseline security. Bodies like the Consultative Committee for Space Data Systems (CCSDS), also known as ISO Technical Committee 20 Subcommittee 13, play a vital role in this process and should be supported in their efforts to develop and promulgate standards for secure command and control, authenticated data formats, and secure inter-satellite communication protocols. This work must be complemented by the harmonization of national-level regulations, such as the EU's proposed space cybersecurity laws and the principles outlined in U.S. Space Policy Directive-5, to create a consistent global security baseline and avoid a fragmented and inefficient patchwork of differing compliance requirements.

Finally, governance must extend to the entire lifecycle of a space system, with a particular focus on the supply chain. Policies must be implemented that mandate robust Cybersecurity Supply Chain Risk Management (C-SCRM) practices for all hardware and software components intended for space systems. Leveraging established frameworks like the NIST Cybersecurity Framework can provide a structured approach to identifying, assessing, and mitigating these risks from a system's inception, ensuring that security is built in, not bolted on.

7.6 Cyber Security Testbed

Recent anomalies from ground-side credential theft to on-orbit bus resets show that security faults in space missions seldom respect organisational or subsystem boundaries. Currently, the space security community lacks a realistic, easily accessible testbed. A scientifically sound *testbed* must therefore function as more than an engineering sandbox: it must be a research instrument that allows hypotheses about security, safety, and resilience to be stated precisely, evaluated systematically, and reproduced independently. The following six design dimensions structure this instrument and crucially explain *why* each is indispensable for scholarly work.

- Segment-complete, abstraction-aware modelling. Attack chains traverse the ground, link, and space segments, and omitting any segment hides entire classes of causality. Therefore, we demand explicit coverage of all three segments even when the RF link is abstracted precisely to keep cross-segment effects observable. Formally recording each segment and its interfaces means that researchers can vary fidelity locally (e.g., replace a hardware ground station with a stochastic delay channel) without invalidating global semantics.
- Graduated fidelity architecture. Simulation is fast and cheap, yet certain timing or radiation effects only surface when real avionics boards are in the loop. We suggest a sliding scale of fidelity that formalises this continuum depending on the research use case, e.g., incident response vs. attack forensics. By binding every experiment to a declarative manifest, a result obtained in a low-fidelity simulation can later be replayed.
- Executable realism with unmodified binaries. Many spacecraft exploits hinge on low-level behaviour (e.g., bus arbitration, watchdog timeouts) that disappears when flight software is re-linked for a laboratory harness. Therefore, it is essential that low-level access (binary) runs within the testbed. Embedding cycle-accurate processor models, flatsats, or physical boards creates an executable ground-truth layer against which analytical models can be calibrated.
- Formal scripting language for faults and threats. The testbed should use one clear language that can describe both random hardware faults and deliberate cyberattacks. By automating techniques from frameworks such as SPARTA, ESA SHIELD, and MITRE

- ATT&CK, and mixing them with classic fault injections, we can measure exactly how much of the threat and fault space we have tested. Each scripted event is tagged and traced to its effect, letting us run solid statistics on how well proposed defences perform.
- Data-first instrumentation and stewardship. Empirical progress depends on transparent, multi-layer telemetry. We insist on capturing three data classes *metadata*, *nominal*, and *attack* traces for every experiment. At the same time, we suggest granular traffic-flow visibility for forensics and recovery research. Embedding synchronised probes at computation, communication, and energy layers satisfies these requirements and produces curated corpora for future machine-learning studies that currently lack representative data.
 - Openness, sustainability, and community extensibility. Proprietary solutions fragment evidence and impede replication. Therefore, for the testbed, we advocate for open standards (CCSDS, ECSS) and a shared registry of benchmark scenarios. A plug-in architecture allows new sensor models, cryptographic stacks, or threat patterns to be added with negligible re-engineering effort, thus ensuring that the testbed evolves alongside mission technology.

8 Recommendations for Future Research and Development

The seminar established that space cybersecurity challenges differ in kind from terrestrial ones, and the working groups on threat preparation, detection, protection, and response surfaced actionable gaps. This section translates those findings into a sequence of interdependent pillars that government agencies and institutions, industry, and academia can use to build a cohesive and cumulative research ecosystem.

Pillar 1: Unified Research Agenda

Progress begins with a shared, public agenda that ties operational pain points to research tasks and evaluation criteria. Space agencies, operators, industry, and academic partners should co-author and annually refresh this agenda, explicitly including long-horizon topics such as secure key distribution for deep space and DTN, physics-informed anomaly detection, provenance-preserving fault tolerance, secure updateability, and verifiable assurance for autonomous systems so foundational science stays aligned with mission needs. This is fundamental for multiple reasons:

- Institutional agencies and industry are dependent on the availability of lower TRL research in order to execute higher TRL prototype development, production, and operationalization.
 - Academia is dependent on use case scenarios and long-term visions of the institutional players and industry in order to be able to select relevant and impactful research topics
- The unified research agenda can capture these dependencies and better link the various actors and their needs.

Pillar 2: Access to High-Fidelity Artifacts and Platforms

Empirical and reproducible research is contingent upon access to realistic *artifacts*, spanning not only telemetry data but also low-level system components. Government and commercial operators should prioritize the creation and dissemination of flight-like datasets. This should include artifacts from missions that are post-mission or have been deorbited, as the operational

risk is eliminated and data can be re-evaluated for release under appropriate agreements. Such datasets should include releasable or anonymized telemetry logs, telecommands, internal satellite bus traffic (e.g., CAN bus, SpaceWire), and firmware images for key subsystems. Where proprietary constraints permit, access to redacted source code offers the highest level of ground truth for formal analysis. When direct release of these artifacts is not feasible, they should be used to curate high-fidelity synthetic corpora anchored to real mission parameters or be made available for analysis within secure data enclaves. Furthermore, a structured framework should be established to grant researchers hands-on access to realistic hardware. This includes creating a repository for retired Engineering Qualification Models (EQMs) from past missions. Additionally, dedicating time to security experiments on operational research platforms, akin to the OPS-SAT [2] model, provides invaluable data on real-world systems. A more ambitious step would be for operators to provide sanctioned research access to in-orbit spacecraft after their primary mission life has concluded. In return, the academic community must commit to the rigorous use of these scarce resources, including documenting dataset limitations and releasing open-source models and data generators to ensure results are comparable and verifiable. At the same time, the research community, supported by space agencies and industry, should work toward developing a space cybersecurity testbed that offers dynamic fidelity to accommodate diverse research experiments.

Pillar 3: Aligned Roles and Incentives.

A sustainable research ecosystem requires a collaborative framework that aligns the distinct roles and incentives of government, industry, and academia. In this tripartite model, government agencies and commercial operators provide the essential context by defining mission constraints, furnishing operational data, and brokering access to hardware. Industry partners serve as the crucial conduit for technology transition, contributing commercial-grade tools, standardized test vectors, and viable pathways to productization. The academic community provides the scientific foundation, delivering novel methods, open-source benchmarks, and the in-depth analysis required to address long-term challenges.

To be effective, collaborative agreements must be structured to recognize the different time horizons inherent to each sector. Projects should be designed to yield both near-term, tangible deliverables (such as software tools and test cases) and to support long-term, foundational research (such as the development of formal methods, principles for autonomy safety, and strategies for PQC migration). This model is designed to be mutually beneficial, creating a virtuous cycle of innovation and capability. Academia benefits from access to relevant problems and data, resulting in peer-reviewed publications and a highly skilled workforce. In return, government and industry gain access to independently validated technologies, robust security evaluations, and ultimately, a reduction in the risk and cost of integrating new security solutions into operational missions.

Pillar 4: Fit-for-Purpose Funding and Collaboration Models

A persistent gap between academic innovation and operational reality is the absence of sustained funding instruments dedicated to low-TRL (TRL 1–3) research. To bridge this, major national and international research programs, such as Horizon Europe and the US National Science Foundation (NSF), alongside other agencies, must establish dedicated, multi-year funding thrusts for space cybersecurity. These programs should be structured as competitive calls for academic-led projects, ensuring publishable results by default and producing deliverables that strengthen the entire research pipeline, including open benchmarks, reference implementations, and curated datasets.

A highly effective structure for these funded projects involves joint research programs, such as industry or agency co-funded PhD positions and fellowships, which embed academic researchers directly within operational environments under pre-negotiated intellectual property agreements. They should be tied to the unified research agenda. This model directly couples funding with commitments for access to data, experimenter time on spacecraft, and controlled use of Engineering Qualification Models (EQMs). By allowing researchers to work with sensitive hardware and data in situ, this approach solves the critical access problem, creating a virtuous cycle: academia gains invaluable access to real-world challenges, while agencies and industry de-risk new technologies and build a direct pipeline to specialized talent.

Pillar 5: Enforceable Standards and Governance.

Technological advances must be codified into enforceable standards and supported by clear governance to ensure a consistent and high security baseline across the space ecosystem. A critical starting point is through procurement and acquisition policy, which should mandate security-by-design principles from inception. Foundational requirements for new systems must include a hardware-rooted secure boot process, cryptographic agility with a clear migration path to Post-Quantum Cryptography (PQC), verifiable software update mechanisms, and policies for post-incident data retention. Furthermore, mandating minimal, interoperable logging and attestation schemas for both command links and internal buses is essential for future incident response and analysis.

Beyond individual systems, community-wide security depends on robust information sharing. This necessitates the creation of a space-focused threat intelligence exchange, building on existing standards such as STIX/TAXII but extending them with space-specific observables and Space Situational Awareness (SSA) context. To encourage proactive defense, clear “safe-harbor” policies for vulnerability disclosure should be established, providing legal protection for good-faith security researchers. Finally, as missions increasingly involve multiple commercial and international partners, unambiguous liability frameworks are required to make collaboration practical by assigning responsibility in the event of a security incident.

Finally, in particular, for solar system Internet scenarios, communication security solutions should be standardized through the Consultative Committee for Space Data Systems (CCSDS) to ensure maximum impact and interoperability.

9 Conclusion

This Dagstuhl Seminar convened 40 leading experts from academia, industry, and space agencies, many of whom possess significant experience securing terrestrial cyber-physical systems such as industrial control systems and autonomous vehicles. This diverse group reached a clear consensus: space cybersecurity is not an incremental extension of terrestrial challenges but a qualitatively distinct discipline. The unique interplay of a deceptive physical environment that creates attribution ambiguity, the operational necessity of high-stakes autonomy under extreme latency, and a uniquely asymmetric and expanding attack surface forge a security paradigm that demands a fundamental shift in our approach. To address these foundational challenges, the seminar’s four working groups translated this diagnosis into concrete priorities.

Moving forward, progress cannot be achieved in isolated silos. The path to secure and resilient space systems is not paved by technological solutions alone but is built upon a strategic foundation of collaboration and shared resources. The recommendations outlined in

this report, from establishing a unified research agenda and providing access to high-fidelity artifacts to aligning stakeholder incentives and creating fit-for-purpose funding or designing a testbed dedicated to space cybersecurity research, are not independent objectives but an interdependent roadmap. The central message of this seminar is a call to action: for space agencies, industry, and academia to collaboratively build the open, reproducible, and cumulative research ecosystem required to safeguard our critical infrastructure in orbit and beyond.

References

- 1 Nicolò Boschetti, Nathaniel G. Gordon, and Gregory Falco. Space cybersecurity lessons learned from the viasat cyberattack. In *ASCEND 2022*. American Institute of Aeronautics and Astronautics (AIAA), 2022.
- 2 David Evans and Mario Merri. Ops-sat: A esa nanosatellite for accelerating innovation in satellite control. In *SpaceOps 2014 Conference*, page 1702, 2014.
- 3 Courtney Fleming, Mark Reith, and Wayne Henry. Securing commercial satellites for military operations: A cybersecurity supply chain framework. In *Proceedings of ICCWS 2023: The 18th International Conference on Cyber Warfare and Security*, pages 85–92. Academic Conferences and Publishing Limited, 2023.
- 4 Max Planck Institute for Extraterrestrial Physics. ROSAT – the end of an exceptional satellite – mpe.mpg.de. https://www.mpe.mpg.de/229897/News_20111114. [Accessed 27-05-2025].
- 5 Loren Grush. A Russian satellite seems to be tailing a US spy satellite in Earth orbit – theverge.com. <https://www.theverge.com/2020/1/31/21117224/russian-satellite-us-spy-kosmos-2542-45-inspection-orbit-tracking>. [Accessed 28-05-2025].
- 6 Ralph Langner. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE security & privacy*, 9(3):49–51, 2011.
- 7 Banks Lin, Wayne Henry, and Richard Dill. Defending small satellites from malicious cybersecurity threats. In *International Conference on Cyber Warfare and Security*, volume 17, pages 479–488, 2022.
- 8 Syed Shahzad, Keith Joiner, Li Qiao, Felicity Deane, and Jo Plested. Cyber resilience limitations in space systems design process: Insights from space designers. *Systems*, 12(10):434, 2024.

Participants

- Ali Abbasi
CISPA – Saarbrücken, DE
- Steven Arzt
Fraunhofer SIT – Darmstadt, DE
- Brandon Bailey
The Aerospace Corp. – Los Angeles, US
- Lars Baumgärtner
ESA / ESOC – Darmstadt, DE
- Nesrine Benchoubane
Polytechnique Montréal, CA
- Simon Birnbach
University of Oxford, GB
- Antonio Carlo
Tallinn University of Technology, EE
- José Manuel Diez López
TU Berlin, DE
- Knut Eckstein
ESA / ESTEC – Noordwijk, NL
- Gregory J. Falco
Cornell University – Ithaca, US
- Daniel Fischer
ESA / ESOC – Darmstadt, DE
- Kevin Gilbert
NASA – Greenbelt, US
- Florian Göhler
BSI – Bonn, DE
- Arne Grenzebach
OHB System – Bremen, DE
- Gürkan Gür
ZHAW – Winterthur, CH
- Jessie Hamill-Stewart
University of Bristol, GB
- Wayne “Chris” Henry
Air Force Inst. of Technology – Wright-Patterson, US
- Eric Jedermann
RPTU – Kaiserslautern, DE
- Samuel Jero
MIT Lincoln Laboratory – Lexington, US
- Gunes Karabulut Kurt
Polytechnique Montréal, CA
- Syed Ibrahim Khandker
New York University – Abu Dhabi, AE
- Vincent Lenders
armasuisse – Thun, CH
- Efrén López Morales
Texas A&M University – Corpus Christi, US
- Mark Manulis
Universität der Bundeswehr – München, DE
- Carsten Maple
University of Warwick, GB & Alan Turing Institute – London, GB
- Ulysse Planta
CISPA – Saarbrücken, DE
- Aanjhan Ranganathan
Northeastern University – Boston, US
- Markus Rückert
ESA / ESOC – Darmstadt, DE
- Peter Y. A. Ryan
University of Luxembourg – Esch-sur-Alzette, LU
- Harshad Sathaye
ETH Zürich, CH
- Stephen Schwab
USC/ISI – Arlington, US
- Mridula Singh
CISPA – Saarbrücken, DE
- Jill Slay
University of South Australia – Mawson Lakes, AU
- Joshua Smailes
University of Oxford, GB
- Fiona Stone
UK Space Agency – London, GB
- Martin Strohmeier
armasuisse – Thun, CH
- Rosa Szurgot
Embry-Riddle Aeronautical University – Prescott, US
- Mattias Wallén
Swedish Space Corporation – Solna, SE
- Marcus Wallum
ESA / ESOC – Darmstadt, DE
- Johannes Willbold
Ruhr-Universität Bochum, DE



The Future of Games in Society

Anders Drachen^{*1}, Johanna Pirker^{*2}, and Lannart E. Nacke^{*3}

1 University of Southern Denmark – Odense, DK. adrac@mmmi.sdu.dk

2 TU München, DE. johanna.pirker@tugraz.at

3 University of Waterloo – Stratford, CA. lennart.nacke@uwaterloo.ca

Abstract

The *Dagstuhl Perspectives Workshop 25102: The Future of Games in Society*, addressed the growing disconnect between gaming's massive global influence – reaching over four billion players in a \$230+ billion industry – and its unrealized potential for societal benefit. While digital games drive technological innovation in, for example, AI, data science, and HCI, and serve as social infrastructures and educational tools, the field faces significant challenges including exploitative monetization, health concerns, and a widening academia-industry gap, that limits research impact. This extends to public policy, where games research is not serving the public interest. This interdisciplinary workshop convened stakeholders to develop strategic directions that realign gaming's influence with societal imperatives, and to establish a bold vision for the future role of games in society. The initiative established a number of key priorities, notably: 1) Ensuring that games promote and facilitate human flourishing; designing games that promote mental health and wellbeing and that promote inclusive online communities. 2) Realizing the potential of educational games to transform education; such as embedding evidence-based learning in education systems. 3) Building sustainable, large-scale research infrastructure, thus enabling industry-academia-policy maker collaboration. Furthermore, utilizing the scale of games for large-scale behavioural research. 4) Developing standardized evaluation frameworks. Enhancing the rigour, assessment, evidence, and knowledge generated from games research and mobilizing this to ensure the positive impact of games on society. This Dagstuhl Perspectives Workshop aimed to unify fragmented efforts into a coherent agenda so that digital games realize their potential as instruments of meaningful societal benefit in our increasingly digital world.

Seminar March 2–5, 2025 – <https://www.dagstuhl.de/25102>

2012 ACM Subject Classification Applied computing → Computer games; Software and its engineering → Interactive games; Information systems → Massively multiplayer online games; Human-centered computing → Human computer interaction (HCI)

Keywords and phrases Game development, games research, artificial intelligence, HCI, player research

Digital Object Identifier 10.4230/DagRep.15.3.39

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

The Future of Games in Society, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 39–55

Editors: Anders Drachen, Johanna Pirker, and Lannart E. Nacke



Dagstuhl Reports

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

1 Executive Summary

Anders Drachen (University of Southern Denmark – Odense, DK, adrac@mmmi.sdu.dk)

Johanna Pirker (TU München, DE, johanna.pirker@tugraz.at)

Lennart E. Nacke (University of Waterloo – Stratford, CA, lennart.nacke@uwaterloo.ca)

License  Creative Commons BY 4.0 International license
© Anders Drachen, Johanna Pirker and Lennart E. Nacke

The full list of participants is included at the end of this document. This document is a collection of the thoughts and writings of all participants.

Digital games engage over four billion individuals globally, accounting for more than a trillion hours of play annually. This activity underpins a \$230+ billion industry and sustains a dynamic, cross-disciplinary field of academic inquiry. Games now operate as social infrastructures, educational tools, and platforms for citizen science. Yet this broadening role brings significant challenges, including concerns around physical and mental health, exploitative monetization models, and contested public narratives about gaming's value.

Gaming continues to drive technological innovation in areas such as artificial intelligence (AI), data science, and human-computer interaction (HCI). Early academic work identified the transformative potential of games to support education, foster social bonds, and enable personalized digital experiences.

However, many of these ambitions remain unrealized. Issues such as predatory monetization and problematic play have come to dominate discourse, overshadowing the field's more constructive capacities. At the same time, research leadership has increasingly shifted from academia to industry, widening the gap between the two despite overlapping expertise. The resource asymmetry limits academia's ability to scale initiatives aimed at producing broader societal benefits. While recent policy frameworks from the EU, UK, and UN have articulated clear expectations for the societal contribution of games, the research community has yet to deliver a coordinated response.

The *Dagstuhl Perspectives Workshop 25102: The Future of Games in Society*, sought to address this gap by formulating strategic directions for both academic policy and industry actors. Our goal was to realign the scale and influence of gaming with emerging societal imperatives. The workshop employed a cross-disciplinary structure, organizing participants into focused working groups tasked with producing actionable strategies to inform future developments across research, practice, and policy.

The digital gaming ecosystem is approaching a pivotal moment. By convening a broad coalition of stakeholders, one can leverage the medium's vast cultural and technological reach while confronting its pressing challenges. The related manifesto will seek to unify currently fragmented efforts into a coherent and strategic agenda that benefits players, researchers, developers, and society as a whole.

Through this initiative, we reaffirm our commitment to ensuring that digital games realize their potential as instruments of meaningful and lasting societal benefit in an increasingly digitized world. We also establish the following key priorities for the future of games in society (these cut across the nine themes of the seminar):

1. **Design for human flourishing:** Ensuring games promote and facilitate human flourishing and designing games that promote mental health and wellbeing and create inclusive online communities and digital environments. This includes creating games that explicitly promote psychological well-being and prosocial behavior, grounded in robust research

and guided by ethical design principles. It also involves implementing proactive design strategies to combat the harms of games, e.g., to reduce toxic behavior.

2. **Realizing the potential of educational games to transform education:** Integrating games into formal curricula through evidence-based design, teacher training, and rigorous evaluation to enhance student engagement and educational outcomes.
3. **Establish sustainable research infrastructure at scale:** Building sustainable, large-scale research infrastructure, which enables industry-academia-policymaker collaboration but also provides the evidence needed to drive good policymaking in games and beyond. Furthermore, utilize the scale of games for large-scale behavioural research. This also involves creating shared platforms, accessible data repositories, and standardized tools to support research between academic and industry actors.
4. **Develop standardized evaluation frameworks:** Creating robust, context-sensitive metrics and assessment tools to evaluate the impact of games on learning, behavior, and social dynamics across diverse populations. There is also a need to enhance the rigour, assessment, evidence, and knowledge generated from games research and mobilize it to ensure the positive impact of games on society.

This Dagstuhl Report summarizes the outcomes of our seminar through a series of abstracts that introduce the nine themes.

2 Table of Contents

Executive Summary

Anders Drachen, Johanna Pirker and Lennart E. Nacke 40

Overview of the Themes

Theme 1: Games for human flourishing
Catherine Flick, Julian Frommel, Linda Hirsch, Simone Kriglstein, Sebastian Deterding, and Rachel Kowert 43

Theme 2: Realizing the potential of educational games to transform traditional education
Aleshia Hayes, Simone Kriglstein, Fabio Zünd, Magy Seif El-Nasr, and Casper Hartevelt 45

Theme 3: Harms of games: shifting the paradigm from mitigation to prevention
Regan Mandryk, Julian Frommel, Guo Freeman, and Kathrin Gerling 46

Theme 4: Games as large-scale behavioural research platforms
Alessandro Canossa, Fabio Zünd, David Melhart, Günter Wallner, Vero Vanden Abeele, Magy Seif El-Nasr, and Regan L. Mandryk 47

Theme 5: Building sustainable and scalable research infrastructure
Vero Vanden Abeele, Günter Vallner, Linda Hirsch, Katja Rogers, Kathrin Gerling, Regan Mandryk, and Michael Young 49

Theme 6: Building a Global Funding Infrastructure for Games Research
Michael Young, Yvette Wohn, Casper Hartevelt, Aleshia Hayes, and Guo Freeman 50

Theme 7: Bridging Industry and Academia: Knowledge Translation and Policy Mediation for Digital Well-being
David Melhart, Enrica Loria, Alena Denisova, Magy Seif El-Nasr, and Pejman Mirza-Babaei 51

Theme 8: Societal Awareness of Games' Impact through Rigour in Assessment, Evidence, and Knowledge Mobilization
Katja Rogers, Alena Denisova, David Melhart, Lannart E. Nacke, Anders Drachen, Catherine Flick, Vero Vanden Abeele, Kathrin Gerling, Regan L. Mandryk, Magy Seif El-Nasr, Günter Wallner, R. Michael Young, and Linda Hirsch 53

Theme 9: Games Research: Responsibility and Impact
Anders Drachen, Johanna Pirker, and Lannart E. Nacke 54

Participants 55

Remote Participants 55

3 Overview of the Themes

Games research is an evolving field, shaped by rapid technological development and changing societal conditions. To effectively navigate this complexity and work toward maximizing the societal value of games, it is essential to identify and engage with the foundational themes that will shape the future of the field.

This *Dagstuhl Report* offers a high-level synthesis of eight key themes identified during the *Dagstuhl Perspectives Workshop 25102: The Future of Games in Society*. These themes, drawn from extensive dialogue and collective analysis within the games research community, represent at the same time the most promising opportunities for games expanding their positive societal impact, and the most urgent challenges confronting the field of games research. Each is explored in greater depth in the accompanying *Dagstuhl Manifesto* for the seminar. The themes are as follows:

1. Games for human flourishing
2. Realizing the potential of educational games to transform traditional education
3. Harms of games: shifting the paradigm from mitigation to prevention
4. Games as large-scale behavioural research platforms
5. Building sustainable and scalable research infrastructure
6. Building a global funding infrastructure for games research
7. Bridging industry and academia: knowledge translation and policy mediation for digital well-being
8. Societal awareness of games' impact through rigour in assessment, evidence, and knowledge mobilization
9. Games Research: Responsibility and Impact

The need to act on these themes has never been more pressing. As games increasingly influence digital culture, human behavior, and technological innovation, a coherent framework is required to guide their development and societal integration. These themes are not only central to the future of games research but also reflect broader systemic issues across academia, industry, and public policy.

Although each theme targets a specific domain, they are inherently interdependent. Their intersections reveal common barriers and shared possibilities for change — many of which extend beyond gaming itself.

Understanding and addressing these themes is essential for realizing the transformative promise of games. By articulating their underlying visions and interconnections, this report provides a roadmap for shaping the next generation of games research and aligning it with societal needs.

3.1 Theme 1: Games for human flourishing

Catherine Flick (University of Staffordshire, GB) (team leader), Julian Frommel (Utrecht University, NL), Linda Hirsch (University of California Santa Cruz, US), Simone Kriglstein (Masaryk University – Brno, CZ), Sebastian Deterding (Imperial College London, GB), Rachel Kowert (University of Cambridge, GB)

License © Creative Commons BY 4.0 International license

© Catherine Flick, Julian Frommel, Linda Hirsch, Simone Kriglstein, Sebastian Deterding, and Rachel Kowert

Our vision focuses on how games can contribute meaningfully to human flourishing, both at the individual and societal levels. Flourishing is defined here not merely as well-being, but

as the capacity to thrive before, during, and after play – manifesting in psychological growth, social connection, critical reflection, and collective values. Games hold unique potential in this regard due to their interactive nature, capacity for role-play, and ability to simulate complex systems. However, realizing this potential requires a holistic and systemic approach that accounts for the broader cultural, industrial, and political contexts in which games are produced and played.

We critique the current overemphasis on “well-being” in the literature and propose a broader, more nuanced understanding rooted in eudaimonic traditions. While much is known about how games can support individual mental, emotional, and physical well-being through inclusivity, accessibility, skill development, and positive experiences, less is understood about how to embed these practices within the mainstream games industry, particularly AAA development, and how to measure long-term societal impact.

To address these challenges, we propose a five-step roadmap:

1. Defining prerequisites such as freedom of expression, inclusive stakeholder engagement, and transparent governance;
2. Building shared concepts through co-creation with diverse actors from industry, academia, policy, and civil society;
3. Developing participatory metrics and heuristics for measuring flourishing, especially from marginalized perspectives;
4. Fostering individual flourishing through best-practice design, critical reflection, and support for self-determination;
5. Advancing societal flourishing by embedding games in shared spaces like education, community programs, and policy frameworks.

Significant challenges remain. Industry pressures prioritize profit over player well-being; governments are retreating from regulation; community moderation is inconsistent; and academic research remains fragmented. Nonetheless, key enablers – such as engaged community leaders, ethical developers, and interdisciplinary researchers – are already present and active.

We call for an ethics of co-creation rather than prescription, acknowledging the risks of bias, exclusion, and unintended consequences. We advocate for ongoing monitoring, broad representation, and institutional support to ensure that interventions are effective and equitable.

Advancing human flourishing through games is not simply a matter of better design or regulation, but requires systemic change across multiple sectors and levels. When supported by rigorous research, inclusive practices, and sustained collaboration, games can contribute to a more just, connected, and thriving society.

3.2 Theme 2: Realizing the potential of educational games to transform traditional education

Aleshia Hayes (University of North Texas – Denton, US) (Team Leader), Simone Kriglstein (Masaryk University – Brno, CZ), Fabio Zünd (ETH Zürich, CH), Magy Seif El-Nasr (University of California at Santa Cruz, US), Casper Hartevelde (Northeastern University – Boston, US)

License © Creative Commons BY 4.0 International license

© Aleshia Hayes, Simone Kriglstein, Fabio Zünd, Magy Seif El-Nasr, and Casper Hartevelde

Educational gaming, while demonstrating significant potential across diverse learning contexts, continues to fall short of its transformative promise due to systemic barriers and implementation challenges. The chocolate on broccoli phenomenon persists, where games fail to effectively integrate educational content with engaging gameplay. Policy constraints, scalability limitations, funding bottlenecks, and technical barriers prevent widespread adoption despite decades of development from pioneering titles like *The Sumerian Game* (1964) to modern platforms like *Minecraft: Education Edition*.

Current educational games often lack understanding of learners, content, and context, resulting in products that are neither sufficiently educational nor engaging. Policy frameworks built around traditional education models create regulatory barriers, while localization challenges limit global applicability. Funding disparities create bottlenecks between development and implementation, and technical issues including restrictive firewalls and inadequate teacher training impede classroom integration.

To realize educational gaming's societal potential, we outline four key enablers:

1. **Comprehensive stakeholder buy-in through evidence-based advocacy**, demonstrating clear return on investment and measurable improvements in student performance, teacher satisfaction, and cost savings.
2. **Systematic capacity building and teacher preparation**, integrating game-based pedagogical approaches into university curricula and ongoing professional development programs.
3. **International collaboration networks and sustainable funding models**, establishing cross-border research partnerships and diversified funding mechanisms including public-private partnerships and outcome-based models.
4. **Strategic integration frameworks and quality assurance systems**, determining optimal opportunities for game-based approaches while preserving effective traditional practices through rigorous evaluation standards.

The roadmap toward impact spans multiple phases:

1. Immediate actions include synthesizing existing successful models, documenting global implementations, and establishing baseline effectiveness metrics.
2. Short-term goals (1-3 years) prioritize building foundational stakeholder networks, engaging policymakers, and creating communication platforms across disciplines.
3. Medium-term objectives (3-5 years) focus on systematic partnerships, validation frameworks, teacher certification programs, and quality assurance protocols.
4. Long-term outcomes (5-10+ years) include scaled pilot implementations, universal access solutions, and institutionalized game-based learning through policy integration and self-sustaining ecosystems.

These interventions hold transformative societal potential. Educationally, they promise personalized learning experiences that enhance digital literacy and reduce inequities. Socially, they create communities of practice connecting educators, students, and families around shared learning objectives. Economically, they prepare digitally literate workforces while supporting sustainable industry growth through more sophisticated consumers and creators.

By shifting from fragmented individual products to coordinated systematic implementation, this agenda envisions educational transformation where games enhance rather than replace traditional methods, creating engaging, equitable, and effective learning environments that serve diverse global communities while maintaining the human connections essential to meaningful education.

3.3 Theme 3: Harms of games: shifting the paradigm from mitigation to prevention

Regan Mandryk (University of Victoria, CA) (team leader), Julian Frommel (Utrecht University, NL), Guo Freeman (Clemson University, US), Kathrin Gerling (Karlsruhe Institute of Technology, DE)

License © Creative Commons BY 4.0 International license
© Regan Mandryk, Julian Frommel, Guo Freeman, and Kathrin Gerling

Digital gaming, while globally pervasive and socially significant, continues to produce harms that are inadequately addressed by current practices. Toxic behaviour, deceptive design, problematic play, and inequitable access persist across platforms, with reactive moderation and fragmented policies proving insufficient. This manifesto proposes a paradigm shift from harm mitigation to proactive harm prevention, grounded in multidisciplinary research and actionable socio-technical strategies.

Toxicity – including hate speech, harassment, and extremist content – is widespread and difficult to regulate, in part due to its subjectivity and normalization within gaming culture. Deceptive design, such as loot boxes and exploitative reward structures, prioritizes monetization over player well-being. Problematic gaming behaviour, while controversial as a clinical diagnosis, causes demonstrable harm for some players, necessitating nuanced frameworks that avoid pathologizing healthy play. Meanwhile, barriers to equitable access, such as inaccessible interfaces and non-inclusive content, continue to marginalize diverse player groups. To address these interlinked harms, we outline four key enablers:

1. **A robust, accessible evidence base on the antecedents, mechanisms, and consequences of harm**, supported by interdisciplinary methods and large-scale in-situ studies.
2. **A strategic shift toward predictive modelling and real-time detection systems** that enable pre-emptive intervention.
3. **Cross-platform, context-sensitive intervention tools** – including algorithms, player-facing resources, and frameworks for ethical design – integrated with industry practices and community norms.
4. **Empowered players and resilient communities** equipped with improved literacy, transparent content communication, and mechanisms for rejecting harmful designs.

The roadmap toward impact spans multiple horizons:

1. Immediate actions include gathering evidence, identifying expertise, and setting pathways for responsible industry collaboration.

2. Short-term goals prioritize refining research agendas, prototyping tools, and developing educational resources.
3. Medium-term goals focus on predictive models, efficacy trials, and ethical data-sharing practices.
4. Long-term outcomes include public-facing repositories, policy implementation, and automated harm prevention systems.

These interventions hold promise for wide societal impact. Culturally, they aim to reframe gaming spaces as inclusive and safe. Educationally, they provide stakeholders – from players to policymakers – with the tools to understand and navigate digital harms. Economically, fostering healthier relationships with games will support the sustainable growth of the industry.

By shifting from fragmented responses to proactive, evidence-informed systems, this agenda envisions a future where games contribute not only to entertainment but to well-being, equity, and collective resilience in digital play.

3.4 Theme 4: Games as large-scale behavioural research platforms

Alessandro Canossa (Royal Danish Academy – Copenhagen, DK) (Team Lead), Fabio Zünd (ETH Zürich, CH), David Melhart (University of Malta – Msida, MT), Günter Wallner (Johannes Kepler Universität Linz, AT), Vero Vanden Abeele (Katholieke Universiteit Leuven, BE), Magy Seif El-Nasr (University of California at Santa Cruz, US), Regan L. Mandryk (University of Victoria, CA)

License © Creative Commons BY 4.0 International license

© Alessandro Canossa, Fabio Zünd, David Melhart, Günter Wallner, Vero Vanden Abeele, Magy Seif El-Nasr, and Regan L. Mandryk

Digital games represent untapped laboratories for understanding human behavior at unprecedented scale and granularity, offering researchers the ability to capture the full spectrum of cognition, social interaction, and decision-making through naturally engaging digital environments. Unlike traditional research methodologies constrained by artificial laboratory settings and self-report biases, games provide controlled yet ecologically valid spaces where millions of participants exhibit authentic behaviors over extended periods, generating rich longitudinal datasets that reveal the fundamental algorithms underlying human psychology and social dynamics.

Current research demonstrates games' potential as behavioral research platforms through three key areas: game-based digital biomarkers for mental health assessment, games as microcosms enabling controlled social experimentation, and personality modeling through gameplay patterns. Studies show that behavioral traces from commercial games can serve as proxies for psychological traits, while virtual environments like Minecraft's anarchy servers provide natural experiments in self-organizing social structures. Advanced techniques now enable researchers to create surprisingly accurate personality profiles from gameplay data, while AI-powered synthetic humans offer unprecedented control over social experimentation variables.

However, significant barriers limit this potential. Ethical challenges around consent, privacy, and data ownership create complex legal landscapes where players may unknowingly contribute psychological profiles while simply seeking entertainment. Misaligned incentives between game companies focused on profit and researchers seeking scientific insight limit data

access and research independence. The potential for algorithmic harm through discrimination in hiring, insurance, or healthcare based on gaming-derived behavioral models raises profound concerns about dual-use applications of this technology.

To realize games' transformative potential as behavioral research platforms, we outline four critical enablers:

1. **Research-oriented game design and intelligent analytics infrastructure**, featuring modular architectures for systematic variable manipulation, comprehensive behavioral data capture, and AI-powered analysis systems capable of identifying complex patterns across massive heterogeneous datasets while maintaining player engagement.
2. **Robust ethical frameworks and interdisciplinary collaboration models**, establishing meaningful informed consent processes, advanced anonymization techniques, and partnership structures that align game development expertise with behavioral research rigor while navigating divergent objectives and success metrics.
3. **Universal data instrumentation and experimental manipulation systems**, developing cross-platform frameworks for behavioral data collection and unified modding systems enabling controlled experimental modifications across any game environment regardless of built-in research support.
4. **Advanced evaluation tools and synthetic content generation**, implementing AI-assisted analytics dashboards, procedural scenario creation, and synthetic human agents that enable scalable, controlled experiments while providing explainable insights into complex behavioral phenomena.

The roadmap toward impact spans multiple development phases:

1. Immediate priorities include automatic data instrumentation systems, memory-level capture frameworks, and standardized ethical collection protocols that work universally across gaming platforms.
2. Short-term development focuses on universal modding systems, dynamic game modification tools, and standardized experimental manipulation frameworks for controlled research within existing games.
3. Medium-term objectives emphasize procedural content generation, AI-driven synthetic humans for controlled social experimentation, and adaptive game environments that adjust parameters based on research requirements.
4. Long-term outcomes include comprehensive evaluation systems, AI-assisted analytics platforms, benchmarking frameworks for reproducible research, and meta-theories of human behavior in digital spaces that inform broader scientific understanding.

These interventions promise transformative societal impact across multiple domains. Scientifically, they enable data-driven policy innovation through virtual testing environments, revolutionize psychological research through digital behavioral twins, and advance healthcare through early mental health detection and personalized therapeutic interventions.

Educationally, they support adaptive learning systems that personalize instruction based on individual cognitive patterns while providing immersive professional training simulations. Socially, they enhance understanding of human dynamics across cultures while informing urban planning through virtual city simulations that predict actual resident behavior patterns.

By transforming games from entertainment platforms into sophisticated behavioral laboratories, this agenda envisions a future where digital environments serve as “petri dishes” for human psychology, capturing the performative rather than declarative aspects of behavior while providing ethical frameworks for studying sensitive social phenomena that would be impossible to replicate safely in real-world settings. Success requires sustained collaboration

between game developers, behavioral scientists, policymakers, and players themselves to ensure that these powerful research capabilities serve human understanding rather than exploitation, ultimately contributing to more nuanced, data-driven approaches to addressing complex social challenges through unprecedented insights into the fundamental nature of human behavior.

3.5 Theme 5: Building sustainable and scalable research infrastructure

Vero Vanden Abeele (KU Leuven, BE) (team leader), Günter Vallner (Johannes Kepler University Linz, AT), Linda Hirsch (University of California Santa Cruz, US), Katja Rogers (University of Amsterdam, NL), Kathrin Gerling (Karlsruhe Institute of Technology, DE), Regan Mandryk (University of Victoria, CA), Michael Young (University of Utah – Salt Lake City, US)

License © Creative Commons BY 4.0 International license

© Vero Vanden Abeele, Günter Vallner, Linda Hirsch, Katja Rogers, Kathrin Gerling, Regan Mandryk, and Michael Young

Games research, as an emerging field, lacks the foundational infrastructure necessary for sustainable growth and scholarly impact. Despite rapid expansion, the discipline suffers from fragmented knowledge, limited resource sharing, and inadequate professional development structures that hinder both individual researchers and collective progress. The organic growth of games research has created a situation where foundational theories remain unconsolidated, research artifacts are rarely preserved or shared, and early-career researchers struggle to navigate complex academic-industry ecosystems without adequate mentorship support.

Current challenges stem from the field's youth and interdisciplinary nature. Researchers frequently “reinvent the wheel” due to insufficient awareness of prior work, while empirical studies often create unique games and tools that remain inaccessible to other researchers, limiting reproducibility and progress. Professional development relies on scattered, often region-specific programs that fail to address the global nature of games research or provide sustained career guidance across academic and industry transitions.

To address these structural deficiencies and establish games research as a mature, impactful discipline, we propose three integrated infrastructure components:

1. **Establishing a comprehensive canon of seminal work**, featuring curated narrative reviews by leading experts, editorial oversight for quality and scope, and an online repository with visualization systems documenting key contributions, theories, and research artifacts to ensure coherent knowledge progression.
2. **Creating a platform for archiving and sharing of artifacts related to games**, providing technological sustainability, methodological rigor and long-term accessibility for research games, tools, and data sets while addressing legal, technical and infrastructure challenges through secure storage, analytics integration, and standardized documentation protocols.
3. **Developing a global mentoring center for professional development**, connecting researchers across academia and industry through structured programs, career guidance resources, and networking opportunities that support talent development from early career advancement to senior researcher transitions between sectors.

The implementation roadmap spans coordinated development phases.

1. Immediate actions include assembling editorial teams for canon development, defining platform functionality requirements for artifact sharing, and identifying global networks of mentoring representatives from academia and industry partnerships.
2. Short-term development focuses on creating narrative review standards, establishing modular platform architectures with data storage and analytics capabilities, and launching pilot mentoring programs with defined formats and evaluation mechanisms.
3. Medium-term objectives emphasize publishing canonical works through indexed venues, deploying comprehensive artifact platforms with versioning and accessibility features, and scaling mentoring networks through organizational partnerships and structured program expansion.
4. Long-term outcomes include maintaining dynamic canon updates reflecting field evolution, ensuring platform sustainability through continued technical and legal support, and establishing institutionalized mentoring frameworks that support sustained professional development across career stages.

These infrastructure investments promise significant social impact through improved research quality and accessibility. Canonical knowledge will improve the acceptance of research findings in games across disciplines, including education and psychology, allowing for more valuable scientific contributions. Artifact platforms will accelerate research progress while increasing transparency and reducing funding waste through improved reproducibility. Global mentoring networks will strengthen community effectiveness and competitiveness, empowering individual researchers while building collective capacity to address complex societal challenges through games research.

By establishing a robust infrastructure for knowledge preservation, resource sharing, and professional development, this agenda transforms games research from a fragmented emerging field into a mature discipline capable of sustained scholarly impact and meaningful social contribution across multiple domains.

3.6 Theme 6: Building a Global Funding Infrastructure for Games Research

Michael Young (University of Utah – Salt Lake City, US, Team Lead), Yvette Wohn (New Jersey Institute of Technology – Newark, US), Casper Hartevelde (Northeastern University – Boston, US), Aleshia Hayes (University of North Texas – Denton, US), Guo Freeman (Clemson University, US)

License  Creative Commons BY 4.0 International license

© Michael Young, Yvette Wohn, Casper Hartevelde, Aleshia Hayes, and Guo Freeman

Games research continues to grow in complexity and scope, yet it lacks the dedicated funding infrastructure necessary to support sustained global collaboration. This manifesto outlines a vision for a global research consortium and advocacy group that connects researchers, aligns national and international efforts, and secures stable, long-term investment in the field. By combining funding infrastructure with strategic advocacy, we aim to elevate games research as a legitimate and impactful domain across borders and disciplines.

The field faces numerous challenges: fragmented funding mechanisms, dispersed researchers, and limited institutional recognition. Its interdisciplinary nature, while a strength, complicates collaboration and grant eligibility. In addition, current advocacy efforts are

uncoordinated and often regionally constrained, limiting visibility and public support. Addressing these systemic issues requires the creation of a global framework that supports both research and advocacy efforts through dedicated leadership, stable funding, and broad coalition building.

We identify four key enablers for realizing this vision:

1. **A sustainable, global research consortium** that supports international coordination, secures dedicated funding, and provides infrastructure for long-term collaboration and support for researchers.
2. **Internal leadership and operational capacity**, including the establishment of financial, strategic, and advocacy roles with expertise in diverse funding ecosystems and cross-sector partnerships.
3. **An international advocacy group** that unites existing organizations, counters public stigma, and promotes the societal impact of games research through targeted communication and public engagement.
4. **Cross-sector alliances and stakeholder incentives** that foster participation, enable interdisciplinary cooperation, and link research output to global educational, policy, and cultural goals.

The roadmap spans multiple horizons:

1. Immediate actions include definition of the mission, community building, founding committees, and outreach to stakeholders and partners.
2. Short-term goals focus on staffing, securing seed funding, legal structure, and launching initial collaborative activities and resources.
3. Medium-term goals include expanding international nodes, refining funding strategies, supporting junior scholars, and evaluating impact.
4. Long-term outcomes include a self-sustaining infrastructure with endowments, annual reports, content studios, and matchmaking platforms for researchers, media, and policy-makers.

This dual infrastructure of research and advocacy has the potential to reshape the social understanding of games. Culturally, it affirms games as legitimate and meaningful parts of life. Educationally, it opens new pathways for learning and behavioral change. Economically and politically, it enables targeted, evidence-driven investment and public policy. By institutionalizing global collaboration, this agenda sets the foundation for a new era of game research with lasting impact.

3.7 Theme 7: Bridging Industry and Academia: Knowledge Translation and Policy Mediation for Digital Well-being

David Melhart (University of Malta – Msida, MT, Team Lead), Enrica Loria (Keen Software House – Prague, CZ), Alena Denisova (University of York, GB), Magy Seif El-Nasr (University of California at Santa Cruz, US), Pejman Mirza-Babaei (Ontario Tech University, CA)

License © Creative Commons BY 4.0 International license

© David Melhart, Enrica Loria, Alena Denisova, Magy Seif El-Nasr, and Pejman Mirza-Babaei

The game industry's rapid expansion is marked by fragmentation, ethical tensions, and disparities between large studios and smaller developers. Despite the capacity of academia for critical insight and innovation, structural misalignment and limited collaboration prevent

meaningful integration of research into development practices. This manifesto proposes a global consortium to institutionalize partnerships between academia, industry and policymakers, supporting ethical innovation, policy mediation, and knowledge translation for digital well-being.

Barriers include conflicting timelines and incentives, power asymmetries, bureaucratic overhead, and a lack of shared language and infrastructure. Collaboration is often limited to informal networks, excluding underrepresented groups and smaller studios. Although pressing, ethical concerns are difficult to address without trust, transparency, and mutual accountability. To build a responsible ecosystem, all stakeholders must engage in a structured and sustained collaboration.

We identify four key enablers to support this transition:

1. **Scalable academic infrastructures and actionable research outputs** that translate theory into tools, frameworks, and recommendations aligned with industry timelines and production needs.
2. **Formalized bridge roles and translational ecosystems** that connect academic and industrial actors, supported through fellowships, joint appointments, and co-development platforms.
3. **Industry engagement models** that encourage responsible design and participation in ethical certification, matchmaking systems, and collaborative research.
4. **Policy mediation and advocacy mechanisms** that translate academic evidence into practical regulation and support the co-creation of enforceable standards promoting digital well-being.

The roadmap to implementation includes:

1. Immediate actions: define consortium structure and goals; initiate trust-building dialogue; engage policymakers and stakeholders.
2. Short-term goals: develop ethical toolkits and training programs; formalize partnerships; launch shared matchmaking and collaboration platforms.
3. Medium-term goals: introduce certification systems; scale bridge-building roles; support interdisciplinary training and research initiatives.
4. Long-term outcomes: embed ethical development as an industry norm through governance models, co-created policy, and cross-sector accountability.

This agenda envisions a mature game development ecosystem rooted in ethical innovation and shared responsibility. It supports safer, more inclusive digital spaces. Culturally, it strengthens trust and public legitimacy. Economically, it reduces risk and boosts long-term sustainability. By institutionalizing collaboration and mutual respect, games can evolve into a sector that not only entertains, but also champions societal well-being and equity.

3.8 Theme 8: Societal Awareness of Games' Impact through Rigour in Assessment, Evidence, and Knowledge Mobilization

Katja Rogers (University of Amsterdam, NL, Team Lead), Alena Denisova (University of York, GB), David Melhart (University of Malta – Msida, MT), Lannart E. Nacke (University of Waterloo, CA), Anders Drachen (University of Southern Denmark – Odense, DK), Catherine Flick (University of Staffordshire, GB), Vero Vanden Abeele (KU Leuven, BE), Kathrin Gerling (Karlsruhe Institute of Technology., DE), Regan L. Mandryk (University of Victoria, CA), Magy Seif El-Nasr (University of California at Santa Cruz, US), Günter Wallner (Johannes Kepler Universität Linz, AT), R. Michael Young (University of Utah – Salt Lake City, US), Linda Hirsch (University of California at Santa Cruz, US)

License © Creative Commons BY 4.0 International license

© Katja Rogers, Alena Denisova, David Melhart, Lannart E. Nacke, Anders Drachen, Catherine Flick, Vero Vanden Abeele, Kathrin Gerling, Regan L. Mandryk, Magy Seif El-Nasr, Günter Wallner, R. Michael Young, and Linda Hirsch

Games research holds the potential to create substantial societal impact, but this impact is limited by fragmented assessment practices, disconnected evidence bases, and underdeveloped knowledge mobilization. This manifesto envisions a future where rigorous, large-scale assessment is supported by shared infrastructures, where interdisciplinary research is recognized through meta-assessment criteria, and where evidence is mobilized beyond academia to inform policy, education, industry, and public discourse.

Currently, games research is often siloed, with limited opportunities for scaling, replication, or cumulative knowledge-building. The field's interdisciplinary nature, while a strength, creates inconsistencies in how research is assessed and understood, complicating collaboration and slowing progress. Meanwhile, the societal value of games remains poorly communicated to key stakeholders due to a lack of accessible, tailored evidence. Addressing these challenges requires systemic change across infrastructure, methods, and outreach.

We identify four key enablers to realize this vision:

1. **Sustainable, large-scale research infrastructure and data ecosystems** that support coordinated assessment, data donation, and resource sharing between projects, institutions, and countries.
2. **Shared meta-assessment criteria and vocabulary** that respect disciplinary diversity while supporting mutual understanding, interdisciplinary collaboration, and coherent quality standards.
3. **Community-driven evidence maps and customized communication strategies** to translate research for policy makers, funding bodies, educators, and the public.
4. **Systemic incentives and support structures** to encourage interdisciplinary practices, knowledge translation, and long-term collaboration between academia, industry, and stakeholders.

The roadmap toward implementation includes:

1. Immediate actions: form advisory groups for large-scale infrastructure and knowledge mobilization; expand evaluation criteria for games research contributions.
2. Short-term goals: identify key resources, platforms and stakeholder needs; develop example-based meta-assessment materials; engage policy makers and educators.
3. Medium-term goals: launch collaborative projects; embed assessment tools into games; prototype audience-specific evidence maps and summaries; advocate for targeted funding.
4. Long-term outcomes: maintain and expand large-scale platforms, implement interdisciplinary training, deploy evidence maps for policy advocacy, and normalize societal impact framing in games research.

This agenda supports a more cohesive, visible, and impactful field. Culturally, it empowers diverse narratives and strengthens the legitimacy of games. Academically, it fosters collaboration and quality. Politically and economically, it informs policy and unlocks funding. Through this transformation, games research can serve not only players and developers, but society as a whole.

3.9 Theme 9: Games Research: Responsibility and Impact

Anders Drachen (SDU Metaverse Lab, DK), Johanna Pirker (TU München, DE & Graz University of Technology, AT), Lennart E. Nacke (University of Waterloo, CA)

License  Creative Commons BY 4.0 International license
© Anders Drachen, Johanna Pirker, and Lennart E. Nacke

There is a gap between the potential and the realized societal impact of games research. While the field has matured considerably over the past two decades - spanning education, health, public policy, and technological innovation - transformative outcomes remain limited in scale and frequency. Despite numerous funded projects and scholarly outputs, few have translated into lasting, wide-reaching societal benefit. The report outlines both the historical contributions of academic research and the systemic barriers that constrain its broader impact.

Notable academic contributions include advancements in domains such as game AI, analytics, and user research, as well as successful but isolated interventions such as *Foldit* (citizen science), *Re-Mission* (health), and *SnowWorld* (VR pain management) and *WEAVR* (academia-industry collaboration). These cases demonstrate the potential for games to support learning, therapy, and civic engagement. However, most initiatives remain stuck at the prototype stage due to scalability issues, funding discontinuities, and limited integration into institutional systems.

Key barriers include structural misalignments between academic and industry incentives, a lack of dedicated support for scaling beyond research prototypes, and insufficient recognition of societal impact in academic evaluation metrics. Furthermore, large-scale industry research often overshadows academic efforts, reducing the visibility and uptake of scholarly innovations. Ethical concerns - particularly around data privacy, exploitative mechanics, and negative public perception - add complexity to adoption, especially in regulated sectors like healthcare and education.

To address these challenges, we propose a structured roadmap. Immediate actions (1–2 years) include forming cross-sector working groups and developing standard assessment tools. Medium- and long-term priorities (2–8+ years) involve aligning incentive structures, institutionalizing ethical frameworks, building shared infrastructure, and embedding game-based solutions into public systems such as schools and hospitals. Continuous improvement cycles and robust ethical oversight are essential for sustainable progress.

A cultural shift is required: games research must prioritize real-world outcomes over academic prestige. Enabling this transformation demands new evaluation criteria, sustained funding models, and cross-sector collaboration that centers on shared societal goals. The United Kingdom-based *Smart Data Donation Service* is one potential model of this future: an initiative that empowers citizens, supports research, and informs policy by bridging data asymmetries between industry and academia.

Ultimately, we argue that while the path to impact is difficult - requiring institutional change across multiple sectors - it is not unattainable. Strategic coordination, ethical rigor, and a focus on societal value can allow games research to fulfill its transformative potential.

Participants

- Alessandro Canossa
Royal Danish Academy –
Copenhagen, DK
- Alena Denisova
University of York, GB
- Anders Drachen
University of Southern Denmark –
Odense, DK
- Catherine Flick
University of Staffordshire –
Stoke-on-Trent, GB
- Guo Freeman
Clemson University, US
- Julian Frommel
Utrecht University, NL
- Kathrin Gerling
KIT – Karlsruher Institut für
Technologie, DE
- Casper Hartevelde
Northeastern University –
Boston, US
- Aleshia Hayes
University of North Texas, US
- Linda Hirsch
University of California –
Santa Cruz, US
- Simone Kriglstein
Masaryk University –
Brno, CZ
- Enrica Loria
Keen Software House –
Prague, CZ
- Regan L. Mandryk
University of Victoria, CA
- David Melhart
University of Malta – Msida, MT
- Pejman Mirza-Babaei
UOIT – Oshawa, CA
- Lannart E. Nacke
University of Waterloo –
Stratford, CA
- Johanna Pirker
TU München, DE
- Katja Rogers
University of Amsterdam, NL
- Magy Seif El-Nasr
University of California at
Santa Cruz, US
- Vero Vanden Abeele
KU Leuven, BE
- Günter Wallner
Johannes Kepler Universität
Linz, AT
- Donghee Wohn
NJIT – Newark, US
- R. Michael Young
University of Utah –
Salt Lake City, US
- Fabio Zünd
ETH Zürich, CH

Remote Participants

- Sebastian Deterding
Imperial College London, GB
- Rachel Kowert
Discord – San Francisco, US



Computational Complexity of Discrete Problems

Swastik Kopparty^{*1}, Meena Mahajan^{*2}, Rahul Santhanam^{*3},
Till Tantau^{*4}, and Ian Mertz^{†5}

1 University of Toronto, CA. swastik.kopparty@gmail.com

2 The Institute of Mathematical Sciences & HBNI – Chennai, IN.
meena@imsc.res.in

3 University of Oxford, GB. rahul.santhanam@cs.ox.ac.uk

4 Universität zu Lübeck, DE. tantau@tcs.uni-luebeck.de

5 Charles University – Prague, CZ. iwmertz@gmail.com

Abstract

This report documents the program and activities of Dagstuhl Seminar 25111 “Computational Complexity of Discrete Problems,” which was held during March 09–14, 2025. The seminar brought together researchers working in many diverse sub-areas of computational complexity, promoting a vibrant exchange of ideas. Following a description of the seminar’s objectives and its overall organization, this report lists the different major talks given during the seminar in alphabetical order of speakers, followed by the abstracts of the talks, including the main references and relevant sources where applicable.

Seminar March 9–14, 2025 – <http://www.dagstuhl.de/25111>

2012 ACM Subject Classification Theory of computation → Complexity theory and logic; Theory of computation → Complexity classes; Theory of computation → Problems, reductions and completeness; Theory of computation → Circuit complexity; Theory of computation → Proof complexity

Keywords and phrases circuit complexity, communication complexity, computational complexity, lower bounds, randomness

Digital Object Identifier 10.4230/DagRep.15.3.56

1 Executive Summary

Swastik Kopparty

Meena Mahajan

Rahul Santhanam

Till Tantau

License  Creative Commons BY 4.0 International license

© Swastik Kopparty, Meena Mahajan, Rahul Santhanam, and Till Tantau

Overview

Computational complexity studies the amount of resources (such as time, space, randomness, communication, or parallelism) necessary to solve discrete problems – a crucial task both in theoretical and practical applications. Despite a long line of research, for many practical problems it is not known if they can be solved efficiently. Here, “efficiently” can refer to polynomial-time algorithms, whose existence is not known for problems like Satisfiability or Factoring. For the large data sets arising for instance in machine learning, already cubic or

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Computational Complexity of Discrete Problems, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 56–76

Editors: Swastik Kopparty, Meena Mahajan, Rahul Santhanam, Till Tantau, and Ian Mertz



Dagstuhl Reports

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

even quadratic time may be too large, but may be unavoidable as research on fine-grained complexity indicates. The ongoing research on such fundamental problems has a recurring theme: the difficulty of proving lower bounds. Indeed, many of the great open problems of theoretical computer science are, in essence, open lower bound problems.

This Dagstuhl Seminar, the 14th in a long-standing series of seminars on this theme, addressed several of these questions in the context of circuit and formula sizes, meta-complexity, proof complexity, fine-grained complexity, communication complexity, and classical computational complexity. In each area, powerful tools for proving lower and upper bounds are known, but particularly interesting and powerful results often arise from establishing connections between the fields. The seminar aimed to bring together a diverse group of leading experts and promising young researchers in these areas, to discuss and to discover new, further connections.

Technical Talks

The Dagstuhl Seminar saw thirty technical lengths, of durations ranging from 10 to 50 minutes, covering and going beyond the themes discussed above. While detailed talk abstracts appear later in this report, here is a brief topic-wise overview.

Hardness of Approximation and Local Testing

When faced with a computational problem for which an efficient solution is hard to find (e.g. if the problem is NP hard), we can hope to efficiently find *approximate* solutions. *Hardness of approximation* is the study of computational limitations to what kinds of approximation can be achieved efficiently, and it has been a flourishing subfield of theoretical computer science. A key ingredient for hardness of approximation theorems are probabilistically checkable proofs and *local testing*: where one wants to check properties of some large object while only querying a small randomly chosen part of it. Yuval Filmus presented a new unified approach to local testing of polymorphisms, generalizing linearity testing and monomial testing, previously proved using quite different techniques. Prahladh Harsha presented an optimal analysis of the classical “lines vs points” low degree test, which can detect when a given function has even just 1%-fraction agreement with a low-degree multivariate polynomial. Such local tests are central ingredients in state-of-the-art probabilistically checkable proofs and hardness of approximation results. Amey Bhangale described a long series of works that are part of a program to classify the hardness of approximating constraint satisfiable problems that are promised to be satisfiable. Shuichi Hirahara presented new results on *average-case* hardness of approximation for matrix multiplication, a topic that has seen much interest in recent years. The key ingredient here is a new proof of the classical Yao XOR lemma, a hardness amplification result with origins in cryptography. Sasha Golovnev gave a talk on barriers to proving exponential time complexity hardness (known as “SETH-hardness”) for many classical problems with unknown complexity like Hamiltonian cycles. Radu Curticapean gave a survey of a recent line of work on hardness of finding subgraphs. This line of work can now determine for every graph H the fine-grained complexity of finding copies of H in a given input graph; remarkably, the hardness results match classical algorithms based on dynamic programming and treewidth.

Meta-Complexity

Meta-complexity studies relationships between lower bounds, learning, pseudorandomness, cryptography and proofs, based on analysing the complexity of compression problems. Valentine Kabanets discussed how a central question in cryptography, namely whether witness encryption exists for NP, is equivalent to a central question in learning theory, namely whether computational learning is hard for NP. Zhenjian Lu defined the Heavy Avoid problem, which asks whether “heavy” elements for a samplable distribution, can be identified efficiently, and showed that this problem is closely related to uniform probabilistic lower bounds. Oliver Korten described connections between the Range Avoidance problem for NC^0 circuits and previously well-studied problems about cell-probe lower bounds and NC^0 pseudo-random generators. In each of these cases, the meta-complexity perspective leads to the identification of new connections and approaches.

Space-bounded Computation

Space-bounded computation was another important theme of the seminar; recent advances in *catalytic* computation have generated much excitement. Ian Mertz discussed the recent breakthrough result of Ryan Williams showing that time can be simulated in nearly square-root space. Michal Koucký described a range of collapses of catalytic classes, including the results that catalytic non-deterministic space and catalytic randomized space are equivalent to catalytic deterministic space. Roei Tell presented work on the long-standing open question of whether randomized logarithmic space can be derandomized, showing that for two standard algorithmic tasks, namely solving connectivity and computing random walk probabilities for graphs, at least one is solvable more efficiently than was hitherto known. Amit Chakrabarti and Sumegha Garg discussed various models of streaming algorithms, which are special kinds of space-bounded algorithms analyzed using various information complexity techniques.

Query and Communication

Kaave Hosseini gave a sweeping overview of various kinds of measures for Boolean matrices – algebraic, analytic, and combinatorial – and their relative strengths in pinpointing the communication complexity of specific Boolean functions. Yogesh Dahiya described recent work focusing on the size of decision trees (a measure of the space required for storing Boolean functions), including a surprising application using size bounds in simple decision trees to derandomize depth (i.e. query complexity) in a generalized decision tree model. Avishay Tal described the connection between query and communication complexities via lifting theorems, and sketched a simpler proof of the lifting theorem of Göös, Pitassi, and Watson for randomized query and communication.

Proof Complexity and Circuits

Several connections between proof complexity and circuit complexity were highlighted in a series of talks. For different complexity measures of the same type of object (proofs, circuits, algorithms), tradeoff results describe the extent to which we can optimize one measure while simultaneously controlling the others. Supercritical tradeoffs describe the phenomenon where a procedure optimizing one complexity measure may make other measures shoot up even beyond the generic worst case bound. Jakob Nordström described recent tightening of supercritical tradeoffs in multiple settings, including cutting-plane proof size vs depth, monotone circuit size vs depth, and more; all hinge upon tradeoff results in propositional

proof complexity. Susanna de Rezende described a generalized query-to-communication lifting theorem and its applications to obtaining lower bounds for monotone circuits and propositional proof sizes. Olaf Beyersdorff sketched a broad framework for translating computational hardness in varied circuit models into QBFs with no short proofs in QBF proof systems naturally corresponding to many real-world solvers.

Pseudorandomness and Combinatorial Constructions

We had several talks on explicit constructions of pseudorandom combinatorial objects – of the kind that are useful for pseudorandom generators and other derandomization tasks. Rachel Zhang presented her new explicit constant-degree expander graphs, breaking a barrier on what is achievable by spectral methods. Gil Cohen presented a new result computing optimal spectral bounds for the zig-zag product, a method for construct expander graphs. Remarkably, their method uses tools from very distant areas of mathematics: free probability and complex analysis. Siqi Liu showed how high dimensional expanders, a hypergraph analogue of expander graphs, could be used to give new locally testable codes with the pointwise multiplication property. Eshan Chattopadhyay presented explicit constructions of extractors from multiple independent sources, that can extract pure randomness when even just three of them are assumed to be weakly random. This result involves several ideas, and in particular develops extractors that fool multiparty communication protocols. Pavel Pudlák showed that nonmalleable affine extractors, due to their strong pseudorandomness, are hard to compute for certain branching programs. Thomas Thierauf gave a survey of several computational problems surrounding graph rigidity. Finally, Makrand Sinha talked about how to generate pseudorandom matrices using random sequences of elementary operations: the study of such problems is motivated by issues in quantum computation.

Open Problems

The seminar also included an open problems session. Interesting research directions and open problems were posed by Sumegha Garg, Mika Göös, Ian Mertz, Jakob Nordström, Hanlin Ren, and Robert Robere.

The seminar included ample time for informal discussions, and interactions in smaller groups. The discussion spaces in the Schloss were put to good and frequent use!

Social Events

The social interactions during the seminar were significantly enhanced by the traditional and well-attended hike on Wednesday afternoon, and the music night on Thursday night (thanks to Antonina Kolokolova for organizing, and to her and Rahul Ilango, Ian Mertz, Noga Ron-Zewi, Avishay Tal, Roei Tell, Rachel Zhang, for actively contributing to this).

Acknowledgments

The organizers, Swastik Kopparty, Meena Mahajan, Rahul Santhanam, and Till Tantau, thank all participants for the many contributions they made. We also especially thank the Dagstuhl staff, who were – as usual – extremely friendly, helpful, and professional regarding all organizational matters surrounding the seminar. Finally, we are deeply grateful to Ian Mertz for his invaluable help assembling and preparing this report.

2 Table of Contents

Executive Summary

<i>Swastik Kopparty, Meena Mahajan, Rahul Santhanam, and Till Tantau</i>	56
--	----

Overview of Talks

Computationally Hard Problems Are Hard for QBF Proof Systems Too <i>Olaf Beyersdorff</i>	62
A New Approximation Algorithm for Satisfiable Constraint Satisfaction Problems <i>Amey Bhangale</i>	62
Leakage-Resilient Extractors Against Number-on-Forehead Protocols <i>Eshan Chattopadhyay</i>	63
Can You Link Up With Treewidth? <i>Radu Curticapean</i>	63
Lifting with Colourful Sunflowers <i>Susanna de Rezende</i>	64
BLR for arbitrary Boolean predicates <i>Yuval Filmus</i>	64
A New Information Complexity Measure for Multi-pass Streaming with Applications <i>Sumegha Garg</i>	65
Polynomial Formulations as a Barrier for Reduction-Based Hardness Proofs <i>Alexander Golovnev</i>	65
An Improved Line-Point Low-Degree Test <i>Prahladh Harsha</i>	66
Error-Correction of Matrix Multiplication Algorithms <i>Shuichi Hirahara</i>	66
Algebraic, Analytic, and Combinatorial complexity measures of boolean matrices <i>Kaave Hosseini</i>	67
Witness Encryption and NP-hardness of Learning <i>Valentine Kabanets</i>	67
Stronger Cell Probe Lower Bounds via Local PRGs <i>Oliver Korten</i>	68
Collapsing Catalytic Classes <i>Michal Koucký</i>	69
High Dimensional Expanders for Error-correcting Codes <i>Siqi Liu</i>	69
On the Complexity of Avoiding Heavy Elements <i>Zhenjian Lu</i>	70
Simulating Time with Square Root Space <i>Ian Mertz</i>	70
Truly Supercritical Trade-offs for Resolution, Cutting Planes, Monotone Circuits, and Weisfeiler–Leman <i>Jakob Nordström</i>	71

Non-malleable affine extractors	
<i>Pavel Pudlák</i>	71
Recent development in the construction of efficient t-wise independent permutations and unitary designs	
<i>Makrand Sinha</i>	72
Lifting Barriers: towards query-to-communication lifting with smaller gadgets	
<i>Avishay Tal</i>	72
When Connectivity is Hard, Random Walks are Easy	
<i>Roei Tell</i>	73
Graph Rigidity	
<i>Thomas Thierauf</i>	73
Explicit Vertex Expanders Beyond the Spectral Barrier	
<i>Rachel Zhang</i>	73
Open problems	
Can Sherali–Adams prove the totality of rwPHP(PLS) in low degree?	
<i>Hanlin Ren</i>	74
Improving SPACE versus NSPACE via Tree Evaluation	
<i>Ian Mertz</i>	74
Participants	76

3 Overview of Talks

3.1 Computationally Hard Problems Are Hard for QBF Proof Systems Too

Olaf Beyersdorff (Friedrich-Schiller-Universität Jena, DE)

License © Creative Commons BY 4.0 International license
© Olaf Beyersdorff

Joint work of Agnes Schleitzer, Olaf Beyersdorff

Main reference Agnes Schleitzer, Olaf Beyersdorff: “Computationally Hard Problems Are Hard for QBF Proof Systems Too”, in Proc. of the AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 – March 4, 2025, Philadelphia, PA, USA, pp. 11336–11344, AAAI Press, 2025.

URL <https://doi.org/10.1609/AAAI.V39I11.33233>

There has been tremendous progress in the past decade in the field of quantified Boolean formulas (QBF), both in practical solving as well as in creating a theory of corresponding proof systems and their proof complexity analysis. Both for solving and for proof complexity, it is important to have interesting formula families on which we can test solvers and gauge the strength of the proof systems. There are currently few such formula families in the literature.

We initiate a general programme on how to transform computationally hard problems (located in the polynomial hierarchy) into QBFs hard for the main QBF resolution systems that relate to core QBF solvers. We illustrate this general approach on three problems from graph theory and logic. This yields QBF families that are provably hard for QBF resolution (without any complexity assumptions).

3.2 A New Approximation Algorithm for Satisfiable Constraint Satisfaction Problems

Amey Bhangale (University of California – Riverside, US)

License © Creative Commons BY 4.0 International license
© Amey Bhangale

Joint work of Amey Bhangale, Subhash Khot, Dor Minzer

Main reference Amey Bhangale, Subhash Khot, Dor Minzer: “On Approximability of Satisfiable k -CSPs: V”, CoRR, Vol. abs/2408.15377, 2024.

URL <https://doi.org/10.48550/ARXIV.2408.15377>

Two algorithms are well-known in the CSP world: Gaussian Elimination and rounding semi-definite program relaxation. In this talk, I will discuss a new ‘hybrid’ approximation algorithm that non-trivially combines these two algorithmic techniques. I will also discuss why we hope that this hybrid algorithm is an optimal approximation algorithm for satisfiable instances of certain CSPs.

References

- 1 Amey Bhangale, Subhash Khot and Dor Minzer. *On Approximability of Satisfiable k -CSPs: V*. In 57th Annual ACM Symposium on Theory of Computing 2025 (to appear), Prague, Czech Republic, 2025

3.3 Leakage-Resilient Extractors Against Number-on-Forehead Protocols

Eshan Chattopadhyay (Cornell University – Ithaca, US)

License © Creative Commons BY 4.0 International license
 © Eshan Chattopadhyay
Joint work of Eshan Chattopadhyay, Jesse Goodman
Main reference Eshan Chattopadhyay, Jesse Goodman: “Leakage-Resilient Extractors against Number-on-Forehead Protocols”, in Proc. of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23–27, 2025, pp. 604–614, ACM, 2025.
URL <https://doi.org/10.1145/3717823.3718272>

Given a sequence of N independent sources X_1, X_2, \dots, X_N , each on n bits, how many of them must be good (i.e., contain some min-entropy) in order to extract a uniformly random string? This question was first raised by Chattopadhyay, Goodman, Goyal and Li (STOC ’20), motivated by applications in cryptography, distributed computing, and the unreliable nature of real-world sources of randomness. In their paper, they showed how to construct explicit low-error extractors for just $K \geq N/2$ good sources of polylogarithmic min-entropy. In a follow-up, Chattopadhyay and Goodman improved the number of good sources required to just $K \geq N/0.01$ (FOCS ’21). In this paper, we finally achieve $K = 3$. Our key ingredient is a near-optimal explicit construction of a new pseudorandom primitive, called a leakage-resilient extractor (LRE) against number-on-forehead (NOF) protocols. Our LRE can be viewed as a significantly more robust version of Li’s low-error three-source extractor (FOCS ’15), and resolves an open question put forth by Kumar, Meka, and Sahai (FOCS ’19) and Chattopadhyay, Goodman, Goyal, Kumar, Li, Meka, and Zuckerman (FOCS ’20). Our LRE construction is based on a simple new connection we discover between multiparty communication complexity and non-malleable extractors, which shows that such extractors exhibit strong average-case lower bounds against NOF protocols.

3.4 Can You Link Up With Treewidth?

Radu Curticapean (Universität Regensburg, DE)

License © Creative Commons BY 4.0 International license
 © Radu Curticapean
Joint work of Radu Curticapean, Simon Döring, Daniel Neuen, Jiaheng Wang
Main reference Radu Curticapean, Simon Döring, Daniel Neuen, Jiaheng Wang: “Can You Link Up With Treewidth?”, CoRR, Vol. abs/2410.02606, 2024.
URL <https://doi.org/10.48550/ARXIV.2410.02606>

Marx showed that $n^{o(k/\log k)}$ time algorithms for detecting colorful H -subgraphs would refute the exponential-time hypothesis ETH, even when H is a k -vertex expander of constant degree. This shows that colorful H -subgraphs are hard even for sparse H , and this result is widely used to obtain almost-tight conditional lower bounds.

We show a self-contained proof of this result that further simplifies very recent works. For this, we introduce a novel graph parameter, the linkage capacity $\gamma(H)$, and we show that detecting colorful H -subgraphs in time $n^{o(\gamma(H))}$ refutes ETH.

A very simple construction of communication networks credited to Beneš gives k -vertex graphs of maximum degree 3 and linkage capacity $\Omega(k/\log k)$. Additionally, we obtain new tight lower bounds for certain patterns by analyzing their linkage capacity. For example, we prove that almost all k -vertex graphs of polynomial average degree $\Omega(k^\beta)$ for some $\beta > 0$ have linkage capacity $\Theta(k)$, which implies tight lower bounds for such patterns H .

3.5 Lifting with Colourful Sunflowers

Susanna de Rezende (Lund University, SE)

License © Creative Commons BY 4.0 International license
© Susanna de Rezende

Joint work of Susanna de Rezende, Marc Vinyals

Main reference Susanna F. de Rezende, Marc Vinyals: “Lifting with Colorful Sunflowers”. Computational Complexity Conference (CCC), 2025, to appear.

In this talk we will show that a generalization of the DAG-like query-to-communication lifting theorem, when proven using sunflowers over non-binary alphabets, yields lower bounds on the monotone circuit complexity and proof complexity of natural functions and formulas that are better than previously known results obtained using the approximation method. These include an $n^{\Omega(k)}$ lower bound for the clique function up to $k \leq n^{1/2-\epsilon}$, and an $\exp(\Omega(n^{1/3-\epsilon}))$ lower bound for a function in P.

3.6 BLR for arbitrary Boolean predicates

Yuval Filmus (Technion – Haifa, IL)

License © Creative Commons BY 4.0 International license
© Yuval Filmus

Joint work of Yaroslav Alekseev, Yuval Filmus

The celebrated BLR linearity test states that if a Boolean function f satisfies $f(x) \oplus f(y) = f(x \oplus y)$ with probability close to 1, then f is close to a linear function, that is, a function that satisfies this equation for all x, y . Another way to view the BLR test is through the lens of *polymorphisms*, a notion from universal algebra. Linear functions are polymorphisms of the predicate $P_{\oplus} = \{(a, b, c) \in \{0, 1\}^3 \mid a \oplus b = c\}$. The BLR test states that an approximate polymorphism of P_{\oplus} (with respect to the uniform distribution) is close to an exact polymorphism. Other results of a similar sort include Mossel’s approximate Arrow theorem, a result of Friedgut and Regev about Kneser graphs, and a result about AND testing which is a prequel to the present work.

In this work, we show that a BLR-like result holds for all predicates on bits, with respect to any distribution which is fully supported on the predicate (this includes BLR for arbitrary distributions). As in the case of AND testing, the statement needs to be changed to allow “multi-sorted” polymorphisms. The proof resembles the classical proof of the triangle removal lemma using the regularity lemma, with Jones’ regularity lemma replacing Szemerédi’s, and It Ain’t Over Till It’s Over an essential ingredient for the counting lemma.

3.7 A New Information Complexity Measure for Multi-pass Streaming with Applications

Sumegha Garg (Rutgers University – New Brunswick, US)

License © Creative Commons BY 4.0 International license

© Sumegha Garg

Joint work of Mark Braverman, Sumegha Garg, Qian Li, Shuo Wang, David P. Woodruff, Jiapeng Zhang
Main reference Mark Braverman, Sumegha Garg, Qian Li, Shuo Wang, David P. Woodruff, Jiapeng Zhang: “A New Information Complexity Measure for Multi-pass Streaming with Applications”, in Proc. of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24–28, 2024, pp. 1781–1792, ACM, 2024.

URL <https://doi.org/10.1145/3618260.3649672>

In this talk, we will introduce a new notion of information complexity for one-pass and multi-pass streaming problems, and use it to prove memory lower bounds for the coin problem. In the coin problem, one sees a stream of n i.i.d. uniform bits and one would like to compute the majority (or sum) with constant advantage. We show that any constant pass algorithm must use $\Omega(\log n)$ bits of memory. This information complexity notion is also useful to prove tight space complexity for the needle problem, which in turn implies tight bounds for the problem of approximating higher frequency moments in a data stream.

3.8 Polynomial Formulations as a Barrier for Reduction-Based Hardness Proofs

Alexander Golovnev (Georgetown University – Washington, DC, US)

License © Creative Commons BY 4.0 International license

© Alexander Golovnev

Joint work of Tatiana Belova, Alexander Golovnev, Alexander S. Kulikov, Ivan Mihajlin, Denil Sharipov,
Main reference Tatiana Belova, Alexander Golovnev, Alexander S. Kulikov, Ivan Mihajlin, Denil Sharipov: “Polynomial Formulations as a Barrier for Reduction-based Hardness Proofs”, ACM Trans. Algorithms, Association for Computing Machinery, 2025.

URL <https://doi.org/10.1145/3721134>

The Strong Exponential Time Hypothesis (SETH) asserts that for every $\varepsilon > 0$ there exists k such that k -SAT requires time $(2 - \varepsilon)^n$. The field of fine-grained complexity has leveraged SETH to prove quite tight conditional lower bounds for dozens of problems in various domains and complexity classes, including Edit Distance, Graph Diameter, Hitting Set, Independent Set, and Orthogonal Vectors. Yet, it has been repeatedly asked in the literature whether SETH-hardness results can be proven for other fundamental problems such as Hamiltonian Path, Independent Set, Chromatic Number, MAX- k -SAT, and Set Cover.

In this paper, we show that fine-grained reductions implying even λ^n -hardness of these problems from SETH for *any* $\lambda > 1$, would imply new circuit lower bounds: super-linear lower bounds for Boolean series-parallel circuits or polynomial lower bounds for arithmetic circuits (each of which is a four-decade open question).

We also extend this barrier result to the class of parameterized problems. Namely, for every $\lambda > 1$ we conditionally rule out fine-grained reductions implying SETH-based lower bounds of λ^k for a number of problems parameterized by the solution size k .

Our main technical tool is a new concept called polynomial formulations. In particular, we show that many problems can be represented by relatively succinct low-degree polynomials, and that any problem with such a representation cannot be proven SETH-hard (without proving new circuit lower bounds).

3.9 An Improved Line-Point Low-Degree Test

Prahladh Harsha (TIFR – Mumbai, IN)

License © Creative Commons BY 4.0 International license
© Prahladh Harsha

Joint work of Prahladh Harsha, Mrinal Kumar, Ramprasad Saptharishi, Madhu Sudan

Main reference Prahladh Harsha, Mrinal Kumar, Ramprasad Saptharishi, Madhu Sudan: “An Improved Line-Point Low-Degree Test”, in Proc. of the 65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024, pp. 1883–1892, IEEE, 2024.

URL <https://doi.org/10.1109/FOCS61266.2024.00113>

In this talk, I’ll show that the most natural low-degree test for polynomials over finite fields is “robust” in the high-error regime for linear-sized fields. This settles a long-standing open question in the area of low-degree testing, yielding an $O(d)$ -query robust test in the “high-error” regime. The previous results in this space either worked only in the “low-error” regime (Polishchuk & Spielman, STOC 1994), or required $q = \Omega(d^4)$ (Arora & Sudan, Combinatorica 2003), or needed to measure local distance on 2-dimensional “planes” rather than one-dimensional lines leading to $\Omega(d^2)$ -query complexity (Raz & Safra, STOC 1997).

Our main technical novelty is a new analysis in the bivariate setting that exploits a previously known connection (namely Hensel lifting) between multivariate factorization and finding (or testing) low-degree polynomials, in a non “black-box” manner in the context of root-finding.

3.10 Error-Correction of Matrix Multiplication Algorithms

Shuichi Hirahara (National Institute of Informatics – Tokyo, JP)

License © Creative Commons BY 4.0 International license
© Shuichi Hirahara

Joint work of Shuichi Hirahara, Nobutaka Shimizu

We present an optimal “worst-case exact to average-case approximate” (non-uniform) reduction for Matrix Multiplication: Given an oracle that correctly computes, in expectation, a $(1/p + \epsilon)$ -fraction of pairs (A, B) of uniformly random matrices over a finite field of order p , we design an efficient oracle non-uniform algorithm that computes Matrix Multiplication exactly for all the pairs of matrices. The proof is based on a simple proof for Yao’s XOR lemma, whose complexity overhead is independent of the output length.

3.11 Algebraic, Analytic, and Combinatorial complexity measures of boolean matrices

Kaave Hosseini (University of Rochester, US)

License © Creative Commons BY 4.0 International license

© Kaave Hosseini

Joint work of Hamed Hatami, Ben Cheung, Kaave Hosseini, Morgan Shirley, Toni Pitassi, Alexander Nikolov

Main reference Tsun-Ming Cheung, Hamed Hatami, Kaave Hosseini, Aleksandar Nikolov, Toniann Pitassi, Morgan Shirley: “A Lower Bound on the Trace Norm of Boolean Matrices and Its Applications”, in Proc. of the 16th Innovations in Theoretical Computer Science Conference, ITCS 2025, January 7-10, 2025, Columbia University, New York, NY, USA, LIPIcs, Vol. 325, pp. 37:1–37:15, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025.

URL <https://doi.org/10.4230/LIPICS.ITCS.2025.37>

I will discuss several fundamental complexity measures for Boolean matrices, such as rank, approximate rank, sign-rank, factorization norm, approximate factorization norm, etc. Then, I will discuss the relationship between these measures by addressing the following question: for two measures, X and Y , is it true that for all Boolean matrices M , if $X(M)$ is small, then $Y(M)$ is small? The quantitative aspects of this question have been an important line of work for several decades, with applications in many areas such as communication complexity, learning theory, dimensionality reduction, etc. However, the question is still poorly understood for several pairs of measures X and Y . I will discuss a few collaborative works to address this question.

3.12 Witness Encryption and NP-hardness of Learning

Valentine Kabanets (Simon Fraser University – Burnaby, CA)

License © Creative Commons BY 4.0 International license

© Valentine Kabanets

Joint work of Halley Goldberg, Valentine Kabanets

We study connections between two fundamental questions from computer science theory. (1) Is *witness encryption* possible for NP [1]? That is, given an instance x of an NP-complete language L , can one encrypt a secret message with security contingent on the ability to provide a witness for $x \in L$? (2) Is *computational learning* (in the sense of [2]) hard for NP? That is, is there a polynomial-time reduction from instances of L to instances of learning?

Our main result is that a certain formulation of NP-hardness of learning (very close to one described in [3]) characterizes the existence of witness encryption for NP. More specifically, we show:

- witness encryption for NP secure against non-uniform polynomial-size adversaries is equivalent to a “half-Levin” reduction from NP to the Computational Gap Learning problem [3];
- witness encryption for NP secure against *uniform* polynomial-time adversaries is equivalent to a BPP-black-box half-Levin reduction from NP to a search version of the same problem;
- witness encryption for NP with ciphertexts having logarithmic length, along with a circuit lower bound for E, are together equivalent to a half-Levin reduction from NP to a “distributional” version of the Minimum Circuit Size Problem.

Next, we prove two unconditional NP-hardness results for agnostic PAC learning. Building on ideas from [5], we prove that agnostic PAC-learning of polynomial-size boolean circuits is NP-hard in the “semi-proper” setting of learning size- $s(n)$ circuits by size- $s(n) \cdot n^{1/(\log \log n)^{O(1)}}$

circuits. We also prove NP-hardness of nearly improper learning in an agnostic “oracle-PAC” model that we define here, in which an algorithm is explicitly given the polynomial-length truth-table of a randomly sampled oracle function \mathcal{O} and is asked to learn with respect to \mathcal{O} -oracle circuits.

Lastly, we give some consequences of our results for the possibility of private- and public-key cryptography. Improving a main result of [3], we show that if improper agnostic PAC learning is NP-hard under a randomized non-adaptive reduction, then $\text{NP} \not\subseteq \text{ioBPP}$ implies the existence of one-way functions. Assuming a half-Levin reduction from an NP-complete language to CGL, we show that $\text{NP} \not\subseteq \text{ioBPP}$ implies the existence of public-key encryption. Along the way, we obtain: if $\text{NP} \not\subseteq \text{ioBPP}$, then witness encryption for NP implies public-key encryption.¹

References

- 1 Sanjam Garg, Craig Gentry, Amit Sahai, Brent Waters. Witness encryption and its applications. Symposium on Theory of Computing Conference (STOC), pp.467–476, 2013. 10.1145/2488608.2488667
- 2 Leslie G. Valiant. A Theory of the Learnable. Comm. ACM, 27(11) pp.1134–1142, 1984. 10.1145/1968.1972
- 3 Benny Applebaum, Boaz Barak, David Xiao. On Basing Lower-Bounds for Learning on Worst-Case Assumptions. Symposium on Foundations of Computer Science (FOCS), pp.211–220, 2008. 10.1109/FOCS.2008.35
- 4 Ilan Komargodski, Tal Moran, Moni Naor, Rafael Pass, Alon Rosen, Eylon Yogev. One-Way Functions and (Im)Perfect Obfuscation. Symposium on Foundations of Computer Science (FOCS), pp.374–383, 2014. 10.1109/FOCS.2014.47
- 5 Shuichi Hirahara. NP-Hardness of Learning Programs and Partial MCSP. Symposium on Foundations of Computer Science (FOCS), pp.968–979, 2022. 10.1109/FOCS54457.2022.00095
- 6 Shuichi Hirahara, Mikito Nanashima. One-Way Functions and Zero Knowledge. Symposium on Theory of Computing (STOC), pp.1731–1738, 2024. 10.1145/3618260.3649701
- 7 Yanyi Liu, Noam Mazon, Rafael Pass. A Note on Zero-Knowledge for NP and One-Way Functions. Electron. Colloquium Comput. Complex. (ECCC), TR24-095, 2024. <https://eccc.weizmann.ac.il/report/2024/095>

3.13 Stronger Cell Probe Lower Bounds via Local PRGs

Oliver Korten (Columbia University – New York, US)

License  Creative Commons BY 4.0 International license
© Oliver Korten

Joint work of Oliver Korten, Toni Pitassi, Russell Impagliazzo

Main reference Oliver Korten, Toniann Pitassi, Russell Impagliazzo: “Stronger Cell Probe Lower Bounds via Local PRGs”, Electron. Colloquium Comput. Complex., Vol. TR25-030, 2025.

URL <https://eccc.weizmann.ac.il/report/2025/030>

In this work, we develop a new method for proving lower bounds for static data structures in the classical cell probe model of Yao. Our methods give the strongest known lower bounds for any explicit problem in this model (quadratically stronger for space as a function of time)

¹ We believe that this last statement follows from a combination of techniques used in prior work ([1, 4, 6]; see [7]), but we have not seen the uniform version stated. In any case, we offer an alternative proof that does not rely on properties of statistical zero knowledge arguments.

and break a barrier which has stood for a few decades. Our lower bounds are based on a connection we establish between the static cell probe model and NC^0 generators, which have been studied extensively in cryptography and more recently in the context of “range avoidance.” With this connection in mind, we analyze the best known cryptographic attacks on NC^0 PRGs, which in turn are based on semirandom CSP refutation, and apply a similar family of arguments to analyze the cell probe model.

3.14 Collapsing Catalytic Classes

Michal Koucký (Charles University – Prague, CZ)

License © Creative Commons BY 4.0 International license
© Michal Koucký

Joint work of Michal Koucký, Ian Mertz, Edward Pyne, Sasha Sami

Main reference Michal Koucký, Ian Mertz, Edward Pyne, Sasha Sami: “Collapsing Catalytic Classes”, Electron. Colloquium Comput. Complex., Vol. TR25-019, 2025.

URL <https://eccc.weizmann.ac.il/report/2025/019>

A catalytic machine is a space-bounded Turing machine with additional access to a second, much larger work tape, with the caveat that this tape is full, and its contents must be preserved by the computation. Catalytic machines were defined by Buhrman et al. (STOC 2014), who, alongside many follow-up works, exhibited the power of catalytic space (CSPACE) and in particular catalytic logspace machines (CL) beyond that of traditional space-bounded machines.

Several variants of CL have been proposed, including non-deterministic and co-non-deterministic catalytic computation by Buhrman et al. (STACS 2016) and randomized catalytic computation by Datta et al. (CSR 2020). These and other works proposed several questions, such as catalytic analogues of the theorems of Savitch and Immerman and Szelepcsényi. Catalytic computation was recently derandomized by Cook et al. (STOC 2025), but only in certain parameter regimes.

We settle almost all questions regarding randomized and non-deterministic catalytic computation, by giving an optimal reduction from catalytic space with additional resources to the corresponding non-catalytic space classes. One main consequence of this is $\text{CL} = \text{CNL}$ i.e. with access to a large filled hard-drive, non-determinism provides no additional power.

Our results build on the compress-or-compute framework of Cook et al. (STOC 2025). Despite proving broader and stronger results, our framework is simpler and more modular.

3.15 High Dimensional Expanders for Error-correcting Codes

Siqi Liu (Institute for Advanced Study – Princeton, US)

License © Creative Commons BY 4.0 International license
© Siqi Liu

Joint work of Siqi Liu, Huy Tuan Pham, Irit Dinur, Rachel Zhang

Main reference Irit Dinur, Siqi Liu, Rachel Yun Zhang: “New Codes on High Dimensional Expanders”, CoRR, Vol. abs/2308.15563, 2023.

URL <https://doi.org/10.48550/ARXIV.2308.15563>

Expanders are well-connected graphs that have been extensively studied and have numerous applications in computer science, including error-correcting codes. High-dimensional expanders (HDXs) generalize expanders to hypergraphs and have the powerful local-to-global

property. Roughly speaking, this property states that the expansion of an HDX can be certified by the expansion of certain local structures. This property has made HDXs crucial in the recent breakthrough on locally testable codes (LTCs) [Dinur et al.'22]. These LTCs simultaneously achieve constant rate, constant relative distance, and constant query complexity. However, despite these desirable properties, these LTCs have yet to find applications in proof systems, as they lack the crucial multiplication property present in widely used polynomial codes. A major open question is: Do there exist LTCs with the multiplication property that achieve the same rate, distance, and query complexity as those constructed by Dinur et al.?

In this talk, I will provide intuition behind the connection between HDXs and LTCs, explain why the LTCs by Dinur et al. lack the multiplication property, and discuss my recent and ongoing work on constructing LTCs with the multiplication property. This talk is based on joint works with Irit Dinur, Rachel Zhang, and Huy Tuan Pham.

3.16 On the Complexity of Avoiding Heavy Elements

Zhenjian Lu (University of Warwick – Coventry, GB)

License  Creative Commons BY 4.0 International license
© Zhenjian Lu

Joint work of Zhenjian Lu, Igor C. Oliveira, Hanlin Ren, Rahul Santhanam

Main reference Zhenjian Lu, Igor C. Oliveira, Hanlin Ren, Rahul Santhanam: “On the Complexity of Avoiding Heavy Elements”, in Proc. of the 65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024, pp. 2403–2412, IEEE, 2024.

URL <https://doi.org/10.1109/FOCS61266.2024.00140>

We introduce and study the following natural total search problem, which we call the heavy element avoidance (Heavy Avoid) problem: for a distribution on N bits specified by a Boolean circuit sampling it, and for some parameter $\delta(N) \geq 1/\text{poly}(N)$ fixed in advance, output an N -bit string that has probability less than $\delta(N)$. We show that the complexity of Heavy Avoid is closely tied to frontier open questions in complexity theory about uniform randomized lower bounds and derandomization.

3.17 Simulating Time with Square Root Space

Ian Mertz (Charles University – Prague, CZ)

License  Creative Commons BY 4.0 International license
© Ian Mertz

Joint work of Ian Mertz, Ryan Williams

Main reference R. Ryan Williams: “Simulating Time with Square-Root Space”, in Proc. of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025, pp. 13–23, ACM, 2025.

URL <https://doi.org/10.1145/3717823.3718225>

We will cover a recent breakthrough result by Williams [3] showing that $\text{TIME}[t]$ is contained in $\text{SPACE}[(t \log t)^{1/2}]$ for all $t \geq n$. We give an overview of the technique, which combines a decomposition of $\text{TIME}[t]$ (given by Hopcroft, Paul, and Valiant [2]) with a recent space-efficient algorithm for solving Tree Evaluation (given by Cook and Mertz [1]). Finally we analyze both ideas and barriers with regards to further progress, as well as potential other directions.

References

- 1 James Cook, Ian Mertz. *Tree Evaluation is in Space $O(\log n \log \log n)$* . Symposium on the Theory of Computing (STOC), 2024.
- 2 John E. Hopcroft, Wolfgang J. Paul, Leslie G. Valiant. *On Time vs Space*. Journal of the ACM (J.ACM), 1977.
- 3 Ryan Williams. *Simulating Time with Square Root Space*. Symposium on the Theory of Computing (STOC) (to appear), 2025.

3.18 Truly Supercritical Trade-offs for Resolution, Cutting Planes, Monotone Circuits, and Weisfeiler–Leman

Jakob Nordström (University of Copenhagen, DK & Lund University, SE)

License © Creative Commons BY 4.0 International license
 © Jakob Nordström
Joint work of Susanna F. de Rezende, Noah Fleming, Duri Andrea Janett, Jakob Nordström, Shuo Pang
Main reference Susanna F. de Rezende, Noah Fleming, Duri Andrea Janett, Jakob Nordström, Shuo Pang: “Truly Supercritical Trade-Offs for Resolution, Cutting Planes, Monotone Circuits, and Weisfeiler–Leman”, in Proc. of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23–27, 2025, pp. 1371–1382, ACM, 2025.
URL <https://doi.org/10.1145/3717823.3718271>

We exhibit supercritical trade-offs for monotone circuits, showing that there are functions computable by small circuits for which any small circuit must have depth super-linear or even super-polynomial in the number of variables, far exceeding the linear worst-case upper bound. We obtain similar trade-offs in proof complexity, where we establish the first size-depth trade-offs for cutting planes and resolution that are truly supercritical, i.e., in terms of formula size rather than number of variables, and we also show supercritical trade-offs between width and size for treelike resolution.

Our results build on a new supercritical depth-width trade-off for resolution, obtained by refining and strengthening the compression scheme for the cop-robber game in [Grohe, Lichter, Neuen & Schweitzer 2023]. This yields robust supercritical trade-offs for dimension versus iteration number in the Weisfeiler–Leman algorithm. Our other results follow from improved lifting theorems that might be of independent interest.

3.19 Non-malleable affine extractors

Pavel Pudlák (The Czech Academy of Sciences – Prague, CZ)

License © Creative Commons BY 4.0 International license
 © Pavel Pudlák
Joint work of Svyatoslav Gryaznov, Pavel Pudlák, Navid Talebanfar


I will prove an exponential lower bound on bottom regular read-once linear branching programs computing non-malleable affine disperser. This is an improvement of our result [1], where we proved an exponential lower bound on branching programs satisfying a stronger condition.

References

- 1 S. Gryaznov, P. Pudlák, N. Talebanfar: Linear Branching Programs and Directional Affine Extractors. Proc. Computational Complexity Conference 2022, Pages 4:1–4:16.

3.20 Recent development in the construction of efficient t -wise independent permutations and unitary designs

Makrand Sinha (*University of Illinois – Urbana-Champaign, US*)

License  Creative Commons BY 4.0 International license

© Makrand Sinha

Joint work of Tony Metger, Alexander Poremba, Makrand Sinha, Henry Yuen

Main reference Tony Metger, Alexander Poremba, Makrand Sinha, Henry Yuen: “Simple Constructions of Linear-Depth t -Designs and Pseudorandom Unitaries”, in Proc. of the 65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024, pp. 485–492, IEEE, 2024.

URL <https://doi.org/10.1109/FOCS61266.2024.00038>

How can we efficiently construct a t -wise independent permutation from local permutation gates that act only on a constant number of bits? This question, originally studied by Gowers in 1996, turns out to be linked to an important object in quantum information theory called t -unitary designs. These designs are pseudorandom unitaries that information-theoretically reproduce the first t moments of the Haar measure on the unitary group. An important recent line of work in quantum computing concerns efficiently constructing such t -unitary designs from local unitary gates that act on a constant number of qubits.


This talk presents a survey of recent developments toward efficient construction of such objects. The talk will mainly be based on my work on the “PFC ensemble” [1], but will also discuss some subsequent followup works by other researchers.

References

- 1 T. Metger, A. Poremba, M. Sinha, and H. Yuen, “Simple Constructions of Linear-Depth t -Designs and Pseudorandom Unitaries,” in *65th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, Chicago, IL, USA, 2024, pp. 485–492. doi: 10.1109/FOCS61266.2024.00038.

3.21 Lifting Barriers: towards query-to-communication lifting with smaller gadgets

Avishay Tal (*University of California – Berkeley, US*)

License  Creative Commons BY 4.0 International license

© Avishay Tal

Joint work of Avishay Tal, Xinyu Wu

Query-to-communication lifting is a powerful method for transferring lower bounds from the query (or decision-tree) model to the communication model. A landmark result by Göös, Pitassi, and Watson (FOCS 2017, SICOMP 2020) demonstrated how to lift randomized query complexity bounds to randomized communication complexity of a related problem, obtained by replacing each input bit with a small “gadget”. A key lemma in their work is the uniform marginals lemma, whose proof is the most technical component of their paper.

We present a new, simpler proof for this lemma. We also discuss limitations of the lemma and, more broadly, of lifting results with the Index gadget, suggesting a modified gadget to address these limitations.

References

- 1 Göös, M., Pitassi, T. and Watson, T., 2017, October. Query-to-communication lifting for BPP. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (pp. 132-143). IEEE.

3.22 When Connectivity is Hard, Random Walks are Easy

Roei Tell (*University of Toronto, CA*)

License © Creative Commons BY 4.0 International license
© Roei Tell

Joint work of Dean Doron, Ted Pyne, Roei Tell, Ryan Williams

Classical PRGs are coupled with a reconstruction argument, asserting that if an adversary can break the PRG, then the adversary can also compute the underlying hard function. Classical reconstruction procedures are randomized, but a recent research effort developed reconstruction procedures for various PRGs that are deterministic, when considering limited types of adversaries.

This talk will present a recent result within this research effort. We construct a pair of deterministic low-space algorithms such that on every input graph, at least one of these algorithms solves a classical problem significantly better than the state-of-the-art: either s - t connectivity is solved, or random walk probabilities are estimated. Consequently, we'll see how to connect the $BPL = L$ question to the question of improving on Savitch's theorem.

3.23 Graph Rigidity

Thomas Thierauf (*Hochschule Aalen, DE*)

License © Creative Commons BY 4.0 International license
© Thomas Thierauf

Joint work of Rohit Gurjar, Kilian Rothmund, Thomas Thierauf

We give an introduction to graph rigidity. Similarly as the perfect matching problem it is related to many other algorithmic problems. In particular, minimal graph rigidity reduces to bipartite perfect matching, which puts it in quasi-NC. Our results are that minimal graph rigidity for planar graphs is in NC, as well as for $K_{3,3}$ -free and K_5 -free graphs.

3.24 Explicit Vertex Expanders Beyond the Spectral Barrier

Rachel Zhang (*MIT – Cambridge, US*)

License © Creative Commons BY 4.0 International license
© Rachel Zhang

Joint work of Jun-Ting Hsieh, Ting-Chun Lin, Sidhanth Mohanty, Ryan O'Donnell, Rachel Zhang

Main reference Jun-Ting Hsieh, Ting-Chun Lin, Sidhanth Mohanty, Ryan O'Donnell, Rachel Yun Zhang: "Explicit Two-Sided Vertex Expanders beyond the Spectral Barrier", in Proc. of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025, pp. 833–842, ACM, 2025.

URL <https://doi.org/10.1145/3717823.3718241>

We give the first explicit constructions of vertex expanders that pass the spectral barrier.


Previously, the strongest known explicit vertex expanders were those given by d -regular Ramanujan graphs, whose spectral properties imply that every small set S of vertices has at least $0.5d|S|$ distinct neighbors. However, it is possible to construct Ramanujan graphs containing a small set S that has no more than $0.5d|S|$ distinct neighbors. In fact, no explicit construction was known to beat the 0.5 barrier.

In this talk, I will discuss how we construct vertex expanders for which every small set expands by a factor of $0.6d$. In fact, our construction satisfies an even stronger property: small sets actually have $0.6d|S|$ *unique neighbors*.

4 Open problems

4.1 Can Sherali–Adams prove the totality of $\text{rwPHP}(\text{PLS})$ in low degree?

Hanlin Ren (*University of Oxford, GB*)

License  Creative Commons BY 4.0 International license
© Hanlin Ren

Joint work of Jiawei Li, Yuhao Li, Hanlin Ren

Main reference Jiawei Li, Yuhao Li, Hanlin Ren: “Metamathematics of Resolution Lower Bounds: A TFNP Perspective”, CoRR, Vol. abs/2411.15515, 2024.

URL <https://doi.org/10.48550/arXiv.2411.15515>

It is known that degree-polylog(n) Sherali–Adams can prove the retraction weak pigeonhole principle (rwPHP) as well as the totality of PLS [1]. The class $\text{rwPHP}(\text{PLS})$ is a combination of the above two classes, recently introduced in [2] to capture the complexity of proving resolution size lower bounds. Can Sherali–Adams prove the totality of $\text{rwPHP}(\text{PLS})$ in degree polylog(n)?


Either a Yes answer or a No answer to the above question would be very interesting. If the answer is Yes, then low-degree Sherali–Adams would be able to prove a large family of resolution size lower bounds (including those for random k -CNFs [3, 2]). On the other hand, a No answer would imply the NP-hardness of automating Sherali–Adams [4].

References

- 1 Mika Göös, Alexandros Hollender, Siddhartha Jain, Gilbert Maystre, William Pires, Robert Robere, Ran Tao. *Separations in Proof Complexity and TFNP*. Journal of the ACM 71(4), 1-45.
- 2 Jiawei Li, Yuhao Li, Hanlin Ren. *Metamathematics of Resolution Lower Bounds: A TFNP Perspective*. arXiv preprint arXiv:2411.15515 (2024).
- 3 Vašek Chvátal, Endre Szemerédi. *Many hard examples for resolution*. Journal of the ACM (JACM), 35(4), 759-768.
- 4 Susanna F. de Rezende, Mika Göös, Jakob Nordström, Toniann Pitassi, Robert Robere, Dmitry Sokolov. *Automating algebraic proof systems is NP-hard*. In STOC’21 (pp. 209-222).

4.2 Improving SPACE versus NSPACE via Tree Evaluation

Ian Mertz (*Charles University – Prague, CZ*)

License  Creative Commons BY 4.0 International license
© Ian Mertz

Savitch’s Theorem [2], which states that $\text{NSPACE}[s]$ is contained in $\text{SPACE}[s^2]$, has stood as a benchmark result in complexity theory for over fifty years. We propose that its tree-like structure can be exploited in conjunction with recent work of Cook and Mertz [1] to show

that $\text{NSPACE}[s] \subseteq \text{SPACE}[o(s^2)]$. This can be achieved by taking the classic NC^2 algorithm implicit in [2] and improving its height by an $\omega(\log \log n)$ factor at the expense of increasing the alphabet size of the wires and functions from $\{0, 1\}$ to $\{0, 1\}^{o(\log^2 n)}$.

References

- 1 James Cook, Ian Mertz. *Tree Evaluation is in Space $O(\log n \log \log n)$* . Symposium on the Theory of Computing (STOC), 2024.
- 2 Walter J. Savitch. *Relationships between nondeterministic and deterministic tape complexities*. Journal of Computer and System Sciences (JCSS), 1970.

Participants

- Olaf Beyersdorff
Friedrich-Schiller-Universität
Jena, DE
- Amey Bhangale
University of California –
Riverside, US
- Igor Carboni Oliveira
University of Warwick –
Coventry, GB
- Amit Chakrabarti
Dartmouth College –
Hanover, US
- Sourav Chakraborty
Indian Statistical Institute –
Kolkata, IN
- Arkadev Chattopadhyay
TIFR – Mumbai, IN
- Eshan Chattopadhyay
Cornell University – Ithaca, US
- Gil Cohen
Tel Aviv University, IL
- Radu Curticapean
Universität Regensburg, DE
- Yogesh Dahiya
The Institute of Mathematical
Sciences – Chennai, IN
- Susanna de Rezende
Lund University, SE
- Yuval Filmus
Technion – Haifa, IL
- Anna Gál
University of Texas – Austin, US
- Sumegha Garg
Rutgers University – New
Brunswick, US
- Mika Göös
EPFL Lausanne, CH
- Alexander Golovnev
Georgetown University –
Washington, DC, US
- Prahladh Harsha
TIFR – Mumbai, IN
- Johan Hastad
KTH Royal Institute of
Technology – Stockholm, SE
- Shuichi Hirahara
National Institute of Informatics –
Tokyo, JP
- Kaave Hosseini
University of Rochester, US
- Rahul Ilango
MIT – Cambridge, US
- Valentine Kabanets
Simon Fraser University –
Burnaby, CA
- Gillat Kol
Princeton University, US
- Antonina Kolokolova
Memorial University of
Newfoundland – St. John's, CA
- Swastik Kopparty
University of Toronto, CA
- Oliver Korten
Columbia University –
New York, US
- Michal Koucký
Charles University – Prague, CZ
- Sophie Laplante
Université Paris Cité, FR
- Nutan Limaye
IT University of
Copenhagen, DK
- Siqi Liu
Institute for Advanced Study –
Princeton, US
- Zhenjian Lu
University of Warwick –
Coventry, GB
- Meena Mahajan
The Institute of Mathematical
Sciences & HBNI – Chennai, IN
- Ian Mertz
Charles University – Prague, CZ
- Jakob Nordström
University of Copenhagen, DK &
Lund University, SE
- Pavel Pudlák
The Czech Academy of Sciences –
Prague, CZ
- Rüdiger Reischuk
Universität zu Lübeck, DE
- Hanlin Ren
University of Oxford, GB
- Robert Robere
McGill University –
Montréal, CA
- Noga Ron-Zewi
University of Haifa, IL
- Michael E. Saks
Rutgers University –
Piscataway, US
- Rahul Santhanam
University of Oxford, GB
- Makrand Sinha
University of Illinois –
Urbana-Champaign, US
- Amnon Ta-Shma
Tel Aviv University, IL
- Avishay Tal
University of California –
Berkeley, US
- Roei Tell
University of Toronto, CA
- Thomas Thierauf
Hochschule Aalen, DE
- Jacobo Torán
Universität Ulm, DE
- Rachel Zhang
MIT – Cambridge, US



PETs and AI: Privacy Washing and the Need for a PETs Evaluation Framework

Emiliano De Cristofaro^{*1}, Kris Shrishak^{*2}, Thorsten Strufe^{*3},
Carmela Troncoso^{*4}, and Felix Morsbach^{†5}

- 1 University of California – Riverside, US. emilianodc@cs.ucr.edu
- 2 Irish Council for Civil Liberties – Dublin, IE. kris.shrishak@iccl.ie
- 3 KIT – Karlsruher Institut für Technologie, DE. thorsten.strufe@kit.edu
- 4 MPI-SP – Bochum, DE. carmela.troncoso@mpi-sp.org
- 5 KIT – Karlsruher Institut für Technologie, DE. felix.morsbach@kit.edu

Abstract

As public awareness of data collection practices and regulatory frameworks grows, privacy-enhancing technologies (PETs) have emerged as a promising approach to reconciling data utility with individual privacy rights. PETs underpin privacy-preserving machine learning (PPML), integrating tools like differential privacy, homomorphic encryption, and secure multiparty computation to safeguard data throughout the AI lifecycle. However, despite significant technical progress, PETs face critical policy and governance challenges. Recent works have raised concerns about efficacy and deployment of PETs, observing that fundamental rights of people are continually being harmed, including, paradoxically, privacy. PETs have been used in surveillance applications and as a privacy washing tool. Current approaches often fail to address broader harms beyond data protection, highlighting the need for a more comprehensive privacy evaluation framework. This Dagstuhl Seminar brought together scholars in computer science and law, along with policymakers, regulators, and industry leaders, to discuss privacy washing and the challenges of detecting privacy washing through PETs and explored pathways toward a framework to address these challenges.

Seminar March 9–14, 2025 – <https://www.dagstuhl.de/25112>

2012 ACM Subject Classification Security and privacy → Privacy protections; Security and privacy → Privacy-preserving protocols; Security and privacy → Social aspects of security and privacy

Keywords and phrases Privacy Enhancing Technologies (PET), Privacy Evaluation, Privacy Harm, Privacy Threats, Privacy Washing

Digital Object Identifier 10.4230/DagRep.15.3.77

1 Executive Summary

Emiliano De Cristofaro (University of California – Riverside, US)

Kris Shrishak (Irish Council for Civil Liberties – Dublin, IE)

Thorsten Strufe (KIT – Karlsruher Institut für Technologie, DE)

Carmela Troncoso (MPI-SP – Bochum, DE)

License © Creative Commons BY 4.0 International license
© Emiliano De Cristofaro, Kris Shrishak, Thorsten Strufe, and Carmela Troncoso

Privacy is a fundamental human right. Article 12 of the Universal Declaration of Human Rights (UDHR) states that everyone has the right to protection from interference with their privacy. One part of protecting people’s privacy is data protection. Laws such as the EU’s

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

PETs and AI: Privacy Washing and the Need for a PETs Evaluation Framework, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 77–93

Editors: Emiliano De Cristofaro, Kris Shrishak, Thorsten Strufe, Carmela Troncoso, and Felix Morsbach



Dagstuhl Reports

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

General Data Protection Regulation (GDPR) have been drafted to protect personal data, which can be exploited to interfere with people's private life. Numerous countries around the world have adopted laws similar to the GDPR. These laws along with an increased awareness of personal data collection have contributed to the appeal of technological solutions known broadly as privacy enhancing technologies (PETs).

The premise of PETs is that these techniques allow data processing while protecting the underlying data from being revealed unnecessarily. PETs make it possible to analyse data from multiple sources without having to see the data. There are two major kinds of PETs: one that offers input privacy and another that offers output privacy. Input privacy allows different people to pour-in their individual data to combine and generate an insight, while no one learns anyone else's individual data. For example, a group of friends can learn who earns the highest without revealing their individual salary to each other. Techniques such as homomorphic encryption and secure multiparty computation (SMPC) fall into this category. These powerful techniques allow two or more entities to compute an agreed upon function on encrypted data. They are useful when the participating entities do not trust each other with their private inputs, but see mutual benefit in the output of the function. Output privacy allows for the release of aggregate data and statistical information while preventing the identification of individuals. Techniques such as differential privacy fall into this category. In many practical use cases, both input and output privacy is desired, and these techniques are combined.

One such use case is AI and in particular machine learning (ML). A huge volume of data is a key component of some of the machine learning techniques, especially those relying on deep neural networks. Personal data and those with sensitive attributes are also used to develop AI models. However, this has contributed to privacy risks. There are a range of attacks in the literature that aim to extract personal data from trained models. PETs have been proposed as the way to protect the functionality of AI while protecting against these privacy attacks. In fact, an entire research field known as privacy-preserving machine learning (PPML) has been formed. PPML incorporates various PETs techniques at various stages of the machine learning to (a) train over encrypted data (e.g., with homomorphic encryption or SMPC), (b) anonymize training process (e.g., DP-SGD), and (c) protect the outputs using differential privacy.

Despite the abundance of works in the area of PETs, AI, and their intersection, there are many remaining challenges. Addressing these challenges is crucial to understand the drawbacks and to reap the benefits of PETs. A range of research questions in Computer Science (protocol design, privacy guarantees, feasibility, scalability, efficiency, etc.) need to be addressed. There are also questions that are interdisciplinary and require expertise from NGOs, ethicists, policy making, law, and regulators. And these research questions are not merely to satisfy academic curiosity but have practical ramifications. They could affect policy making and the work of regulators.

In this Dagstuhl Seminar, a multidisciplinary group of computer science and legal academics and practitioners from industry, human rights groups, and regulators discussed two challenges:

1. **Privacy washing through PETs:** In the recent years, PETs have been used in surveillance applications as in the case of Apple's proposed (and then retracted) approach to scan images on people's phones when uploading photos to iCloud. They have also been used in applications where the personal data is seemingly protected but the privacy threats faced by people are amplified, for example in targeted advertising. Such applications show that PETs can be used for "privacy washing". At the heart of the issue is that

most works fail to protect against the interference with privacy as laid down in Article 12 of the UDHR. These works are agnostic to the application context or too generic or limited to the cryptographic protocol without considering the privacy threats due to the system where it is embedded. The imbalances and asymmetries of power between the stakeholders, the role of infrastructures and their providers, and the control of the computing infrastructure are not accounted for. Technical measures to protect data are discussed as being equivalent to privacy, when they are not. Privacy violations can take many other forms including economic and discrimination harms. When the goal of the application is to harm privacy, such technical measures to protect data cannot protect the interference with privacy. The threat models in the literature are inadequate, and thus, systems designed under such models continue to cause privacy harms.

2. **Evaluation framework to detect privacy washing:** If PETs are to protect against interference with privacy, as laid down in the UDHR, then we require standard evaluation methods and frameworks that allow us to compare the degree of protection. While the literature is filled with ways to measure PETs, they are hard to compare. Limitations of PETs should be well documented so that privacy washing through PETs is stopped. A lack of an independent evaluation framework allows privacy washing. Addressing this challenge is timely and this seminar took the initial steps towards an evaluation framework.

Seminar Structure

Since the participants came from diverse backgrounds ranging from different topics in computer science to legal and regulatory work, the seminar began with several introductory talks and two panel discussions to bring everyone up to speed. Then, we brainstormed in small groups about all the aspects that could influence whether the deployment of a PET could be considered privacy washing. We subsequently grouped these aspects into four topics: Functionality and Framing, Infrastructure for PETs, Accountability, and Detection of Fake PETs. We split the group into four subgroups to discuss these aspects further and develop criteria by which to evaluate the deployment of a PET leading to a vast catalogue of factors that influence the efficacy of PET deployments. During the plenary meetings after group discussions, the rapporteurs from each group shared the progress made during the group discussions. Finally, we spent the remaining time to merge the results of the four subgroups into a draft for a position paper. The position paper describes what privacy washing is, who is involved in its deployment, who can be affected by it, and the considerations that help to detect privacy washing in deployed systems.

2 Table of Contents

Executive Summary

Emiliano De Cristofaro, Kris Shrishak, Thorsten Strufe, and Carmela Troncoso . . . 77

Overview of Talks

Introduction to Differential Privacy and Federated Learning	
<i>Aurélien Bellet</i>	82
Agency Protection: Organisms & Institutions	
<i>Robin Berjon</i>	82
Purpose formulations as a weak link in data protection	
<i>Asia Biega</i>	82
A regulator's perspective on data protection and privacy	
<i>Paul Comerford</i>	83
Anonymisation: Introduction and Perspectives	
<i>Ana-Maria Cretu</i>	83
Attacks on privacy-preserving systems	
<i>Yves-Alexandre de Montjoye</i>	83
Understanding and addressing fairwashing in machine learning	
<i>Sébastien Gambs</i>	84
The PET Paradox – The case of Amazon Sidewalk	
<i>Seda F. Gürses</i>	84
PETs Intro: Multiparty Computation and Homomorphic Encryption	
<i>Bailey Kacsmar</i>	85
What I believe privacy engineering is and some missing pieces	
<i>Carmela Troncoso</i>	85
You Still See Me	
<i>Rui-Jie Yew</i>	85

Panel discussions

AI models, data protection and privacy washing	
<i>Aurélien Bellet, Asia Biega, Paul Comerford, Yves-Alexandre de Montjoye, Carmela Troncoso, and Rui-Jie Yew</i>	86
User Perspective of PETs	
<i>Bailey Kacsmar, Robin Berjon, Emiliano De Cristofaro, Lucy Qin, and Carmela Troncoso</i>	86

Working groups

Detecting Fake PETs	
<i>Frederik Armknecht, Aurélien Bellet, Ana-Maria Cretu, Yves-Alexandre de Montjoye, Georgi Ganev, Patricia Guerra-Balboa, Felix Morsbach, and Thorsten Strufe</i>	87
Infrastructure & PETs	
<i>Robin Berjon, Seda F. Gürses, Lucy Qin, Michael Veale, and Rui-Jie Yew</i>	88

Functionality and Framing
Asia Biega, Johanna Gunawan, Hinako Sugiyama, Vanessa Teague, and Carmela Troncoso 90


Accountability
Bailey Kacsmar, Paul Comerford, Sébastien Gambs, and Kris Shrishak 92

Participants 93

3 Overview of Talks

3.1 Introduction to Differential Privacy and Federated Learning

Aurélien Bellet (INRIA – Montpellier, FR)

License  Creative Commons BY 4.0 International license
© Aurélien Bellet

In an era of AI-driven applications, balancing data utility with user privacy is more important than ever. This talk provides a high level introduction to two key approaches addressing this challenge: federated learning and differential privacy. Federated learning enables collaborative model training without sharing raw data, while differential privacy provides strong guarantees against individual data leakage. This talk discusses the fundamental ideas behind these techniques, their real-world applications, and some challenges that remain.

3.2 Agency Protection: Organisms & Institutions

Robin Berjon (Princeton, US)

License  Creative Commons BY 4.0 International license
© Robin Berjon

It's always tempting to cut things at what seem like logical joints so as to make thinking about the individual component easier. In a sense, that's what we've done with privacy, focusing primarily on various forms of data processing. But the world is rarely as orthogonal as we model it to be. This brief talk situated privacy in a wider institutional framework and suggests that we may use an institutional grammar to evaluate the role and effectiveness of privacy decisions in a broader context.

3.3 Purpose formulations as a weak link in data protection

Asia Biega (MPI-SP – Bochum, DE)

License  Creative Commons BY 4.0 International license
© Asia Biega

Purpose limitation is one of the requirements under the GDPR. User data has to be processed for specific, explicit, and legitimate purposes. Purpose formulations specify and describe these purposes. In this talk, I presented four examples that, over time, convinced me that these formulations are a weak link in data protection: and thus become a tool for privacy washing.

3.4 A regulator's perspective on data protection and privacy

Paul Comerford (Information Commissioner's Office – Wilmslow, GB)

License © Creative Commons BY 4.0 International license
© Paul Comerford

Paul Comerford (Principal Technology Adviser at the ICO) discussed the role of the technology and innovation directorate at the ICO. The talk focused on the ICOs work on PETs across multiple domains and its upcoming guidance on anonymisation and pseudonymisation. He also discussed our recent work on AI and PETs.

3.5 Anonymisation: Introduction and Perspectives

Ana-Maria Cretu (EPFL – Lausanne, CH)

License © Creative Commons BY 4.0 International license
© Ana-Maria Cretu
Main reference Andrea Gadotti, Luc Rocher, Florimond Houssiau, Ana-Maria Cretu, Yves-Alexandre de Montjoye: “Anonymization: The imperfect science of using data while preserving privacy”, *Science Advances*, Vol. 10(29), p. eadn7053, 2024.
URL <https://doi.org/10.1126/sciadv.adn7053>

Anonymisation is the main legal paradigm for sharing data while protecting people's right to privacy. In spite of decades of research, robust anonymisation (“de-identifying”) of individual-level datasets remains an elusive goal. Numerous re-identification attacks have indeed shown how adversaries can use auxiliary information about individuals to single them out in supposedly anonymous datasets. One solution to the data sharing problem is aggregation, whereby data owners share with third parties the results of a computation across all records, while retaining control over the individual-level data. Aggregation solutions include summary statistics, interactive queries over the data, synthetic data, and machine learning. But aggregation does not, on its own, protect privacy, and evaluating the privacy of these solutions is far from trivial. This talk described the two main approaches for this: (1) designing and evaluating privacy attacks and (2) formal methods based on differential privacy, with their advantages and their challenges, together with my perspective on the field.

3.6 Attacks on privacy-preserving systems


Yves-Alexandre de Montjoye (Imperial College London, GB)

License © Creative Commons BY 4.0 International license
© Yves-Alexandre de Montjoye

Companies and governments are increasingly relying on privacy-preserving techniques to collect and process sensitive data. In this talk, I will discuss our efforts to red team deployed systems and argue that red teaming is essential to protect privacy in practice. I will first shortly describe how traditional de-identification techniques fail in today's world. I will then show how implementation choices and trade-offs have enabled attacks against real-world systems, from query-based systems to differential privacy mechanisms and synthetic data. I will conclude by discussing how this applies to modern AI systems.

3.7 Understanding and addressing fairwashing in machine learning


Sébastien Gambs (UQAM – Montreal, CA)

License  Creative Commons BY 4.0 International license
© Sébastien Gambs

Fairwashing refers to the risk that an unfair black-box model can be explained by a fairer model through post-hoc explanation manipulation. In this talk, I will first discuss how fairwashing attacks can transfer across black-box models, meaning that other black-box models can perform fairwashing without explicitly using their predictions. This generalization and transferability of fairwashing attacks imply that their detection will be difficult in practice. Finally, I will nonetheless review some possible avenues of research on how to limit the potential for fairwashing.

3.8 The PET Paradox – The case of Amazon Sidewalk

Seda F. Gürses (TU Delft, NL)

License  Creative Commons BY 4.0 International license
© Seda F. Gürses

Main reference Thijmen van Gend, Donald Jay Bertulfo, Seda F. Gürses: “The PET Paradox: How Amazon Instrumentalises PETs in Sidewalk to Entrench Its Infrastructural Power”, CoRR, Vol. abs/2412.09994, 2024.

URL <https://doi.org/10.48550/ARXIV.2412.09994>

Recent applications of Privacy Enhancing Technologies (PETs) reveal a paradox. PETs aim to alleviate power asymmetries, but can actually entrench the infrastructural power of companies implementing them vis-à-vis other public and private organizations. We investigate whether and how this contradiction manifests with an empirical study of Amazon’s cloud connectivity service called Sidewalk. In 2021, Amazon remotely updated Echo and Ring devices in consumers’ homes, to transform them into Sidewalk “gateways”. Compatible Internet of Things (IoT) devices, called “endpoints”, can connect to an associated “Application Server” in Amazon Web Services (AWS) through these gateways. We find that Sidewalk is not just a connectivity service, but an extension of Amazon’s cloud infrastructure as a software production environment for IoT manufacturers. PETs play a prominent role in this pursuit: we observe a two-faceted PET paradox. First, suppressing some information flows allows Amazon to promise narrow privacy guarantees to owners of Echo and Ring devices when “flipping” them into gateways. Once flipped, these gateways constitute a crowdsourced connectivity infrastructure that covers 90% of the US population and expands their AWS offerings. We show how novel information flows, enabled by Sidewalk connectivity, raise greater surveillance and competition concerns. Second, Amazon governs the implementation of these PETs, requiring manufacturers to adjust their device hardware, operating system and software; cloud use; factory lines; and organizational processes. Together, these changes turn manufacturers’ endpoints into accessories of Amazon’s computational infrastructure; further entrenching Amazon’s infrastructural power. We discuss similarities and differences between previous strategic uses of PETs by Google and Apple to expand their infrastructural offerings to third parties. Accordingly, we argue that power analyses undergirding PET designs should go beyond analyzing information flows. We propose future steps for policy and tech research.

3.9 PETs Intro: Multiparty Computation and Homomorphic Encryption

Bailey Kacsmar (University of Alberta – Edmonton, CA)

License © Creative Commons BY 4.0 International license
© Bailey Kacsmar

In this session we provided an overview on what multiparty computation (MPC) is and how we can think about its variants. We similarly discussed homomorphic encryption (HE). The goal with this session was to establish the breadth of the areas and provide attendees with a common language to think about the way privacy enhancing technologies (PETs) that employ MPC and HE can vary; allowing us to better evaluate the implications of these technologies for privacy and artificial intelligence. We concluded with an overview of some of what is currently possible, in terms of applications, that employ MPC and HE.

3.10 What I believe privacy engineering is and some missing pieces

Carmela Troncoso (MPI-SP – Bochum, DE)

License © Creative Commons BY 4.0 International license
© Carmela Troncoso

In this talk we revisited previous definitions of privacy engineering, showing that data or trust minimization do not necessarily minimize harms. We then argue that purpose minimization is the design goal that helps in this respect. Purpose-oriented thinking additionally has a benefit that it enables to identify fundamental purposes and harms that derive from the goal of the system and have to be assumed as a risk should the system be deployed. We then discussed some missing definitions that would allow to capture harms associated to function creep.

3.11 You Still See Me

Rui-Jie Yew (Brown University – Providence, US)

License © Creative Commons BY 4.0 International license
© Rui-Jie Yew

Main reference Rui-Jie Yew, Lucy Qin, Suresh Venkatasubramanian: “You Still See Me: How Data Protection Supports the Architecture of ML Surveillance”, CoRR, Vol. abs/2402.06609, 2024.

URL <https://doi.org/10.48550/ARXIV.2402.06609>

Data forms the backbone of artificial intelligence (AI). Privacy and data protection laws thus have strong bearing on AI systems. Shielded by the rhetoric of compliance with data protection and privacy regulations, privacy-preserving techniques have enabled the extraction of more and new forms of data. In this talk, I illustrate how the application of privacy-preserving techniques in the development of AI systems—from private set intersection as part of dataset curation to homomorphic encryption and federated learning as part of model computation—can further support surveillance infrastructure under the guise of regulatory permissibility. Finally, I propose technology and policy strategies to evaluate privacy-preserving techniques in light of the protections they actually confer. I conclude by highlighting the role that technologists can play in devising policies that combat surveillance AI technologies.

4 Panel discussions

4.1 AI models, data protection and privacy washing

Aurélien Bellet (INRIA – Montpellier, FR), Asia Biega (MPI-SP – Bochum, DE), Paul Comerford (Information Commissioner’s Office – Wilmslow, GB), Yves-Alexandre de Montjoye (Imperial College London, GB), Carmela Troncoso (MPI-SP – Bochum, DE), and Rui-Jie Yew (Brown University – Providence, US)

License © Creative Commons BY 4.0 International license

© Aurélien Bellet, Asia Biega, Paul Comerford, Yves-Alexandre de Montjoye, Carmela Troncoso, and Rui-Jie Yew

The panel explored how privacy-enhancing technologies (PETs) and regulatory tools are increasingly used for privacy washing – creating a surface-level appearance of compliance while sidestepping real accountability. Sandboxes and red teaming were called out as processes that can be used for legitimizing privacy-invasive systems without addressing underlying risks. Technologies like differential privacy, synthetic data generation and federated learning were highlighted as particularly vulnerable to misuse, especially when their implementation details are obscured or when their guarantees are undermined through practices like budget resetting or general misconfigurations. A key point raised was that evaluations should prioritize the actual impact on individuals and society, not just technical compliance or claimed adherence to norms.

The conversation also focused on the role and limits of transparency. While transparency was broadly supported as essential for accountability, it was acknowledged that legal barriers like trade secrets and competition law often prevent meaningful oversight. There was an agreement that transparency should go beyond abstract metrics and provide explanations that are intelligible to non-experts. At the same time, concerns were raised that transparency alone can also be co-opted as another form of privacy washing if not paired with enforcement and verification. The discussion underscored the need for enforceable standards, empowered regulators, and a shift away from over-optimizing technical frameworks toward addressing broader systemic and structural issues in privacy governance.

4.2 User Perspective of PETs

Bailey Kacsmar (University of Alberta – Edmonton, CA), Robin Berjon (Princeton, US), Emiliano De Cristofaro (University of California – Riverside, US), Lucy Qin (Georgetown University – Washington, DC, US), and Carmela Troncoso (MPI-SP – Bochum, DE)

License © Creative Commons BY 4.0 International license

© Bailey Kacsmar, Robin Berjon, Emiliano De Cristofaro, Lucy Qin, and Carmela Troncoso

The panel discussed the gap between how privacy-enhancing technologies (PETs) are developed and how real users understand, need, or experience them. There was a recurring argument that users often lack the language, awareness, or mental models to demand privacy – much like people once lacked the concept of clean tap water – yet that doesn’t mean privacy isn’t essential. PETs should be designed to be invisible and default, not something users must consciously engage with. Communication breakdowns between researchers, usability experts, and end-users were identified as major barriers. There was also a push to broaden the definition of “users” to include software engineers and institutional actors, since engineers

are often key decision-makers and operate much closer to the tools in practice. The lack of actionable usability research and the assumption of a clearly defined privacy “problem” were cited as weaknesses in the current ecosystem.

Beyond end-users, the conversation highlighted the role of high-risk populations, NGOs, policymakers, and businesses in the PETs landscape. Messaging must be tailored – migrants, for instance, face urgent harms that don’t always register as “privacy” risks. While PETs can support collective systems like digital commons, structural components and effective messaging are missing. Some companies adopt PETs reactively (e.g. post-GDPR), while others see them as a branding opportunity – but distinguishing meaningful implementations from superficial ones remains difficult. There’s also underused potential in inter-organizational PET deployments and in rethinking how to engage businesses without falling into technical “impossibility” traps. A key takeaway: users shouldn’t bear the burden of privacy, and communicating harm – especially to those at risk – must be better informed, more targeted, and more pragmatic.

5 Working groups

5.1 Detecting Fake PETs

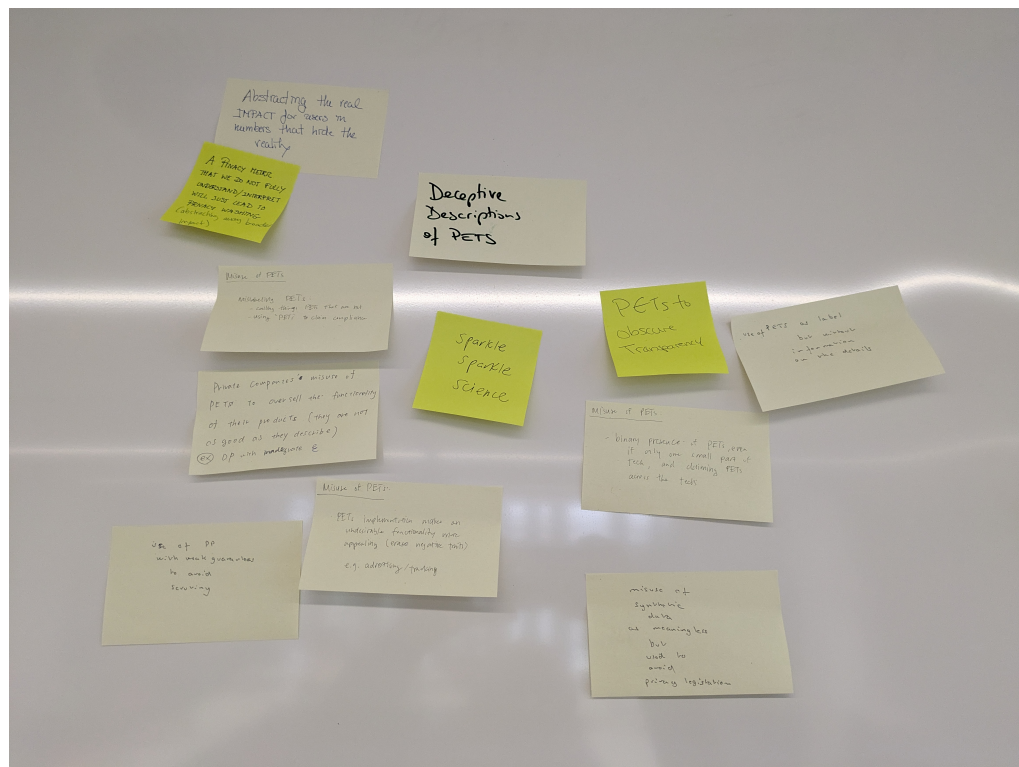
Frederik Armknecht (Universität Mannheim, DE), Aurélien Bellet (INRIA – Montpellier, FR), Ana-Maria Cretu (EPFL – Lausanne, CH), Yves-Alexandre de Montjoye (Imperial College London, GB), Georgi Ganey (University College London, GB), Patricia Guerra-Balboa (KIT – Karlsruher Institut für Technologie, DE), Felix Morsbach (KIT – Karlsruher Institut für Technologie, DE), and Thorsten Strufe (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 4.0 International license

© Frederik Armknecht, Aurélien Bellet, Ana-Maria Cretu, Yves-Alexandre de Montjoye, Georgi Ganey, Patricia Guerra-Balboa, Felix Morsbach, and Thorsten Strufe

The group focused on defining a structured approach to detect fake PETs – privacy-enhancing technologies that mislead through exaggerated claims, poor implementation, or misconfiguration. A central proposal was to create a standardized transparency tool, akin to model cards or data sheets, tentatively referred to as a “privacy card.” This would contain minimal but essential information to assess whether a system is making valid privacy claims. The discussion outlined four key failure categories in PETs: mismatch (between claims and actual threat mitigation), overestimation (inflated protection claims), wrong implementation, and wrong configuration. A foundational requirement is that any privacy claim must specify the threat model, the PET used, and the degree of mitigation. Vague claims without a clear adversarial context were identified as a red flag and sufficient grounds for labeling the system a fake PET.

Each failure mode was elaborated with practical evaluation steps. For mismatch and overestimation, the group emphasized decomposing systems into discrete threat-mitigation claims and validating them with formal or empirical evidence. Identifying implementation failures requires either open-source access or a reproducible protocol, with particular scrutiny on subtleties like side channels, flawed randomness, or deviations from trusted primitives. Configuration errors, such as excessive ϵ values in differential privacy or undersized encryption keys, must be contextualized within both the system’s technical parameters and its deployment environment. The group stressed that privacy guarantees are only meaningful when technical claims are precise, verifiable, and aligned with real-world adversary models – making transparent, auditable documentation a practical necessity to prevent privacy washing.



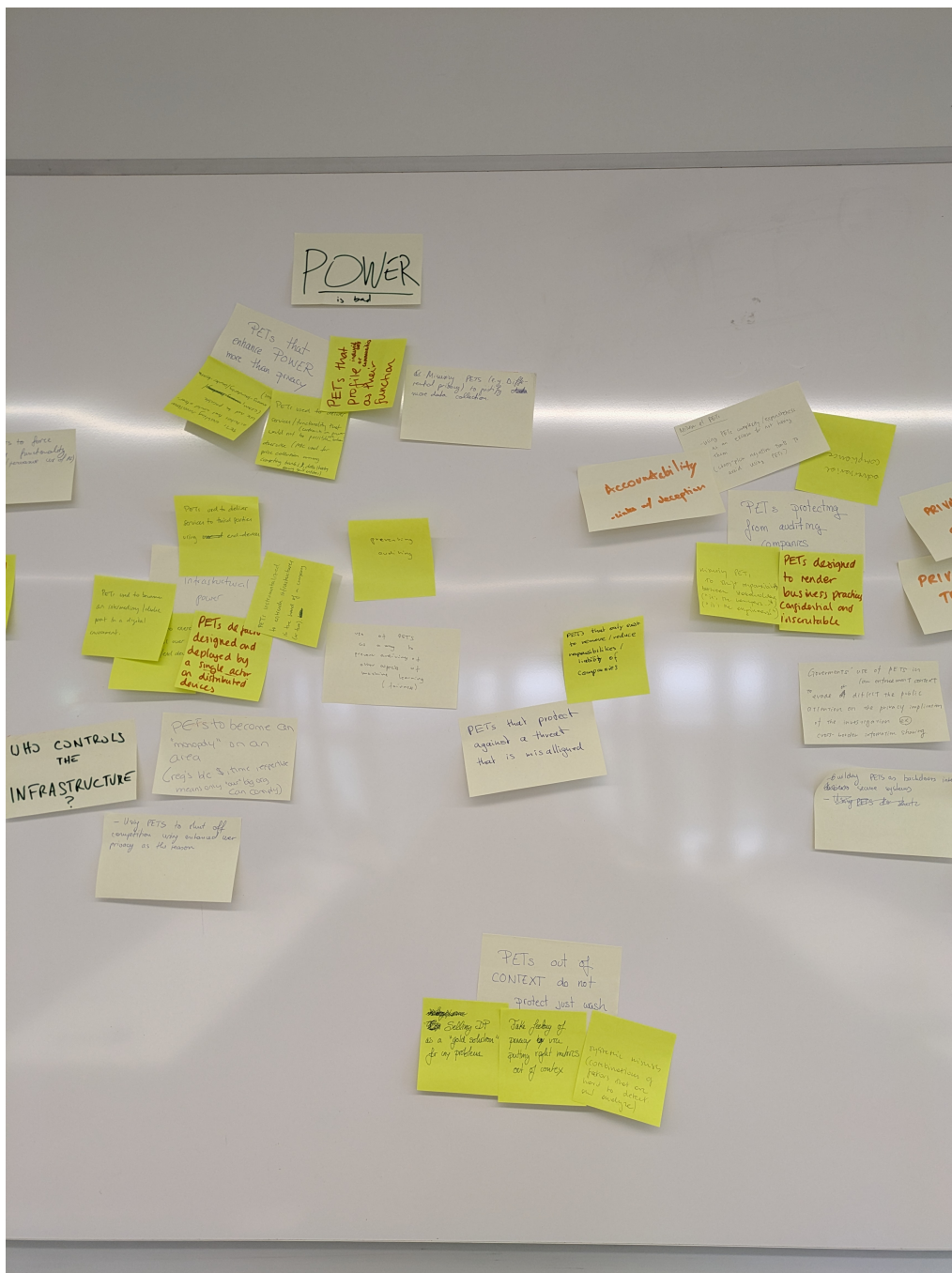
5.2 Infrastructure & PETs

Robin Berjon (Princeton, US), Seda F. Gürses (TU Delft, NL), Lucy Qin (Georgetown University – Washington, DC, US), Michael Veale (University College London, GB), and Rui-Jie Yew (Brown University – Providence, US)

License © Creative Commons BY 4.0 International license
© Robin Berjon, Seda F. Gürses, Lucy Qin, Michael Veale, and Rui-Jie Yew

The discussion highlighted that evaluating Privacy-Enhancing Technologies (PETs) cannot be isolated from the computational infrastructure they rely on, which often embodies extractive or privacy-compromising characteristics. A key challenge identified is the “stack problem”: PETs depend on underlying infrastructures that may themselves lack privacy protections, making truly independent PETs difficult to build and sustain without relying on PET-compatible infrastructure, governance, and funding. This dynamic concentrates power among well-resourced entities capable of controlling infrastructure, raising concerns about exclusionary effects on who can develop or maintain PETs based on existing economic and political incentives.

The group further emphasized that the production environment and infrastructure shape PET design, deployment, and sustainability, often introducing trust relationships and operational vulnerabilities. Coordination among infrastructure providers, deployers, and users creates new power relations, sometimes consolidating rather than distributing it. Ultimately, privacy-washing occurs when infrastructural dependencies and power asymmetries are overlooked, leading to overstated claims about PET’s protections while entrenching systemic privacy risks. Effective evaluation frameworks must therefore assess PETs in a full-stack context, including the socio-technical and governance layers that support or constrain them.



5.3 Functionality and Framing

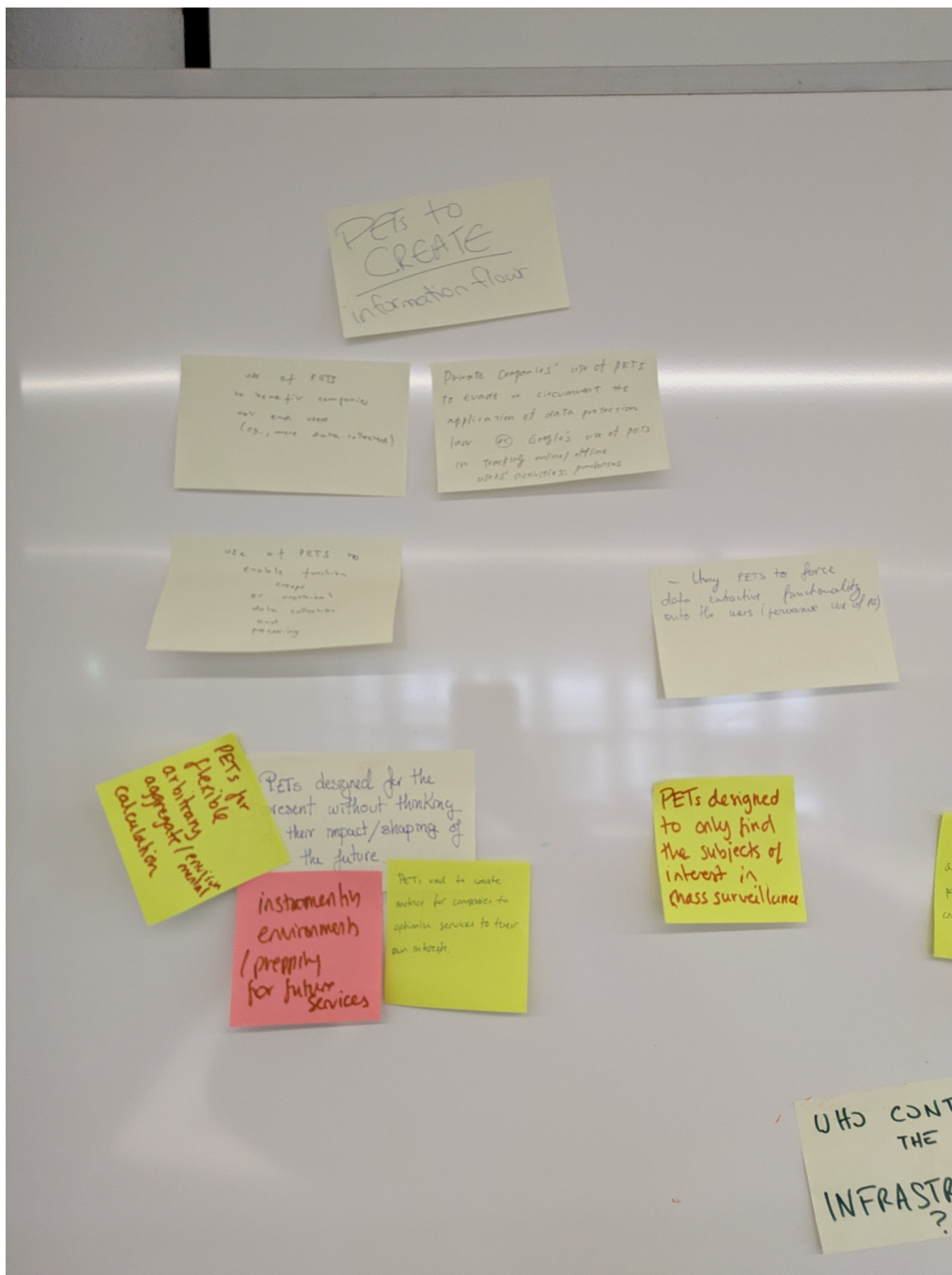
Asia Biega (MPI-SP – Bochum, DE), Johanna Gunawan (Maastricht University, NL), Hinako Sugiyama (University of California – Irvine, US), Vanessa Teague (Australian National University – Acton, AU), and Carmela Troncoso (MPI-SP – Bochum, DE)

License © Creative Commons BY 4.0 International license

© Asia Biega, Johanna Gunawan, Hinako Sugiyama, Vanessa Teague, and Carmela Troncoso

The group explored how privacy-enhancing technologies (PETs) can be co-opted to obscure harm rather than mitigate it, calling for a taxonomy of privacy-washing methods to support clearer evaluation, from the system’s purpose, to its implementation, communication, and resulting consequences. Three key forms of privacy washing were identified: first, when PETs are layered over systems whose underlying purpose is harmful or objectionable, PETs cannot fix this inherent harm; second, when the system’s implementation is harmful, even if its purpose is legitimate, and PETs are used to mask this; and third, when misleading communication about PETs causes harm – such as falsely marketing systems as end-to-end encrypted. In all three cases, PETs risk being used as decorative compliance tools, deflecting attention from structural issues or enabling more sophisticated forms of manipulation, profiling, or opacity.

The group proposed analyzing PETs through the full lifecycle of a system: from purpose, to technical implementation, to communication, and finally to consequences. A system-wide framing was emphasized – assessing not just whether a PET works, but whether it genuinely addresses the privacy risks tied to the system’s function and context. PETs that enable or justify harmful practices and information flows were flagged as particularly concerning. The group cautioned against starting privacy assessments too late in the process (e.g., at the DPIA stage), noting that foundational design choices may already predetermine harm. A meaningful framework, they argued, must account for both intentional and structural misuses of PETs, especially as these tools are increasingly used to manage – not eliminate – power imbalances and information asymmetries.



5.4 Accountability

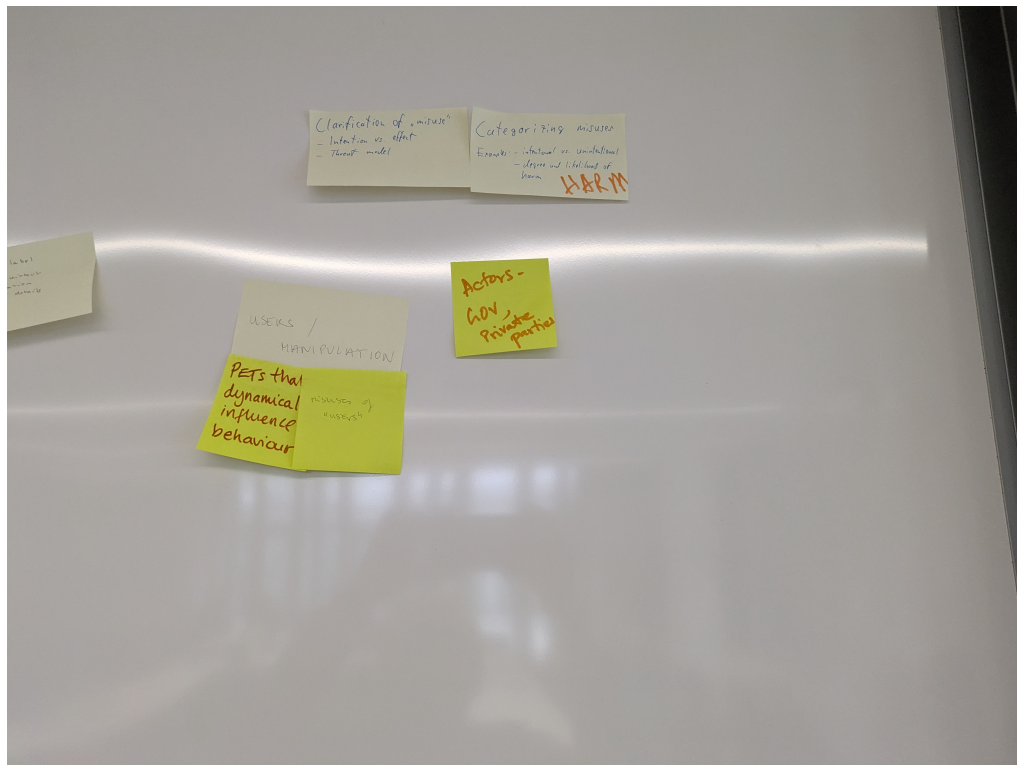
Bailey Kacsmar (University of Alberta – Edmonton, CA), Paul Comerford (Information Commissioner's Office – Wilmslow, GB), Sébastien Gambs (UQAM – Montreal, CA), and Kris Shrishak (Irish Council for Civil Liberties – Dublin, IE)

License © Creative Commons BY 4.0 International license

© Bailey Kacsmar, Paul Comerford, Sébastien Gambs, and Kris Shrishak

The group discussed how PETs can be misused to block accountability, particularly by preventing audits, obscuring system behavior, or undermining data access rights. Rather than supporting transparency, some PETs are designed – or framed – as privacy solutions while enabling organizations to evade scrutiny. For example, this includes using PETs to justify blocking data subject's access rights under the GDPR, offloading privacy responsibilities to third-party service providers, or deploying unverifiable systems that require trust without oversight. The group emphasized the importance of designing PETs with auditability and verifiability in mind to counter these failures and ensure they contribute to, rather than hinder, accountability.

In the context of AI, similar dynamics emerge. Techniques like federated learning – while often cited as privacy-preserving – can be used to obscure data processing practices and resist evaluation due to their complexity. Participants noted that organizations may deploy PETs to legitimize questionable practices or bypass legal requirements, especially in the private sector where business incentives dominate. Overall, the discussion called for evaluation frameworks that address how PETs are used in practice – focusing not just on their technical properties, but also on their role in enabling or obstructing rights, oversight, accountability, and public trust.



Participants

- Frederik Armknecht
Universität Mannheim, DE
- Aurélien Bellet
INRIA – Montpellier, FR
- Robin Berjon
Princeton, US
- Asia Biega
MPI-SP – Bochum, DE
- Paul Comerford
Information Commissioner's
Office – Wilmslow, GB
- Ana-Maria Cretu
EPFL – Lausanne, CH
- Emiliano De Cristofaro
University of California –
Riverside, US
- Yves-Alexandre de Montjoye
Imperial College London, GB
- Sébastien Gambs
UQAM – Montreal, CA
- Georgi Ganev
University College London, GB
- Patricia Guerra-Balboa
KIT – Karlsruher Institut für
Technologie, DE
- Seda F. Gürses
TU Delft, NL
- Johanna Gunawan
Maastricht University, NL
- Bailey Kacsmar
University of Alberta –
Edmonton, CA
- Felix Morsbach
KIT – Karlsruher Institut für
Technologie, DE
- Lucy Qin
Georgetown University –
Washington, DC, US
- Kris Shrishak
Irish Council for Civil Liberties –
Dublin, IE
- Thorsten Strufe
KIT – Karlsruher Institut für
Technologie, DE
- Hinako Sugiyama
University of California –
Irvine, US
- Vanessa Teague
Australian National University –
Acton, AU
- Carmela Troncoso
MPI-SP – Bochum, DE
- Michael Veale
University College London, GB
- Rui-Jie Yew
Brown University –
Providence, US



Scheduling

Claire Mathieu^{*1}, Nicole Megow^{*2}, Benjamin J. Moseley^{*3},
Frits C. R. Spijksma^{*4}, and Alexander Lindermayr^{†5}

1 CNRS, Paris, FR. clairemathieu@gmail.com

2 University of Bremen, DE. nicole.megow@uni-bremen.de

3 Carnegie Mellon University, Pittsburgh, USA. moseleyb@andrew.cmu.edu

4 TU Eindhoven, NL. f.c.r.spijksma@tue.nl

5 Universität Bremen, DE. linderal@uni-bremen.de

Abstract

This report documents the program and outcomes of Dagstuhl Seminar 25121, “Scheduling”. The seminar focused on bridging traditional algorithmic scheduling with the emerging field of fairness in resource allocation. Scheduling is a longstanding research area that has been studied from both practical and theoretical perspectives in computer science, mathematical optimization, and operations research for over 70 years. Fairness has become a key concern in recent years, particularly in the context of resource allocation and scheduling, where it naturally arises in applications such as kidney exchange, school choice, and political districting. The seminar centered on three main themes: (1) fair allocation, (2) fairness versus quality of service, and (3) modeling aspects of fairness in scheduling.

Seminar March 16–21, 2025 – <https://www.dagstuhl.de/25121>

2012 ACM Subject Classification Theory of computation → Scheduling algorithms; Mathematics of computing → combinatorial optimization; Theory of computation → Approximation algorithms analysis

Keywords and phrases scheduling, fairness, mathematical optimization, algorithms and complexity, uncertainty

Digital Object Identifier 10.4230/DagRep.15.3.94

1 Executive Summary

Claire Mathieu (CNRS, Paris, FR)

Nicole Megow (University of Bremen, DE)

Benjamin J. Moseley (Carnegie Mellon University, Pittsburgh, USA)

Frits C. R. Spijksma (TU Eindhoven, NL)

License  Creative Commons BY 4.0 International license

© Claire Mathieu, Nicole Megow, Benjamin J. Moseley, and Frits C. R. Spijksma

This Dagstuhl Seminar was number 8 in a series of Dagstuhl “Scheduling” seminars (since 2008). Scheduling is a major research field that is studied from a practical and theoretical perspective in computer science, mathematical optimization, and operations research. Applications range from traditional production scheduling and project planning to the newly arising resource management tasks in the advent of internet technology and shared resources. While there has been remarkable progress on algorithmic theory for fundamental scheduling problems, leading to insights for other fields as well, scheduling has proven to be an inspirational ground for new questions.

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Scheduling, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 94–112

Editors: Claire Mathieu, Nicole Megow, Benjamin J. Moseley, Frits C. R. Spijksma, and Alexander Lindermayr



Dagstuhl Reports

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

At this meeting, we have focussed on emerging models for **fairness in scheduling and resource allocation**. Traditionally, scheduling theory has focused on how to allocate resources to optimize quality of service guarantees, throughput, or efficiency. However, these objectives do not consider fairness to the underlying agents or entities.

An example of fairness considerations in government resource allocation can be observed in the distribution of healthcare resources, especially during times of crises like pandemic. A government may have a limited amount of resources available to distribute. One could distribute these to the areas affected most by the outbreak. It may also be important though to consider trade-offs between areas that traditionally have had disparities and are underfunded, ensuring vulnerable populations are not neglected. A government must balance immediate needs with long-term equability. Other examples of fairness in resource allocation and scheduling naturally arise in kidney exchange, school choice, tournament design, as well as political districting. These exciting and socially important problems demand to be better understood.

This seminar focused on three complementary themes.

- *Fair Allocation*. Fair allocation has taken center stage in multi-agent systems and economics over the past decade due to its significance both industrially and socially. Essentially, it addresses how to distribute items, whether they be goods or tasks, to agents in a way that leaves each content with their share. Notably, when dealing with indivisible items, perfect fairness metrics like envy-freeness and proportionality aren't always achievable. Recent research endeavors focus on creating algorithms that approximate these fairness standards. On the other hand, game theory delves into the challenge of fairly dividing resources among individuals with entitlements, a dilemma found in numerous real-world scenarios, from inheritance divisions to electronic frequency allocations. Fundamental to fair division is the belief that the involved parties, perhaps with the aid of a mediator, should carry out the allocation, as they best understand their value assessments. A classic example of a fair division method is the “divide and choose” algorithm, which ensures that two participants each feel they have received the most favorable portion. The vast landscape of fair division research extends this principle to more intricate contexts, adapting to varying goods, fairness criteria, player characteristics, and other evaluation standards. Fair allocation, resource allocation and scheduling are fields that build on one another as often algorithmic and analysis techniques in one find uses in the others.
- *Balancing Fairness and Quality of Service*. In the algorithms community, striking a balance between fairness and quality of service (QoS) is a pressing concern. While algorithms, particularly in sectors like finance, healthcare, and social networking, play a pivotal role in decision-making, ensuring equitable outcomes without compromising efficiency or performance is challenging. Fairness ensures that no group or individual is unfairly disadvantaged or discriminated against by algorithmic decisions, and it aims to create an even playing field across diverse sets of users or stakeholders. On the other hand, quality of service emphasizes responsiveness, reliability, and overall user satisfaction. Balancing these two elements is challenging analytically. The area requires a deep understanding of how to model the trade-offs and algorithmically balance quality of service and fairness.
- *Modeling Fairness*. Modeling fairness in scheduling and resource allocation presents a plethora of challenges. Scheduling and allocating resources inherently involves making decisions that prioritize certain tasks, individuals, or groups over others, which can inadvertently introduce biases or create disparities. One fundamental challenge lies in

defining what “fairness” actually means in varied contexts, as it can be subjective and differ across stakeholders. Even when fairness is well-defined, achieving it can sometimes conflict with optimizing for efficiency or maximum resource utilization. Additionally, when dealing with diverse sets of resources and stakeholders with distinct needs and preferences, ensuring equitable distribution becomes complex. There is also the issue of unseen biases in historical data, which, when used to train algorithms, can perpetuate past inequities. Furthermore, there is a constant need to balance immediate and long-term fairness, especially when resource availability fluctuates. Navigating these intricacies requires a deep understanding of real-world challenges to develop sound models for scheduling and resource allocation problems.

Organization of the Seminar. The seminar brought together 42 researchers from theoretical computer science, mathematical optimization, and operations research. The participants consisted of both senior and junior researchers, including a number of postdocs and advanced PhD students. During the five days of the seminar, 29 talks of different lengths took place. Five keynote speakers gave an overview of the state-of-the art of the respective area or presented recent highlight results in 60 minutes:

- Adrian Vetta: Six Candidates Suffice to Win a Voter Majority
- Swati Gupta: Fair Resource Allocation from Theory to Practice
- Lars Rohwedder: The Santa Claus Problem: Three Perspectives
- Kavitha Telikepalli: Fair solutions to the house allocation problem
- Ulrike Schmidt-Kraepelin: Proportional Representation in Budget Allocation.

The remaining slots were filled with shorter talks of 30 minutes on various topics related to the intersection of fairness, social choice, and scheduling.

Outcome. Organizers and participants regard the seminar as a great success. The seminar achieved the goal to bring together the related communities, share the state-of-the art research and discuss the current major challenges. The talks were excellent and stimulating; participants actively met in working groups in the afternoon and evenings. It was remarked positively that a significant number of younger researchers (postdocs and PhD students) participated and integrated well.

Acknowledgements. The organizers wish to express their gratitude towards the Scientific Directorate and the administration of the Dagstuhl Center for their great support for this seminar.

2 Table of Contents

Executive Summary

Claire Mathieu, Nicole Megow, Benjamin J. Moseley, and Frits C. R. Spieksma . . . 94

Overview of Talks

Lossless Robustification of Packet Scheduling Algorithms <i>Yossi Azar</i>	99
Fair Strategic Facility Location with Predictions <i>Eric Balkanski</i>	99
Lift-and-Project Integrality Gaps for Santa Claus <i>Etienne Bamas</i>	100
Minimax Group Fairness in Strategic Classification <i>Emily Diana</i>	100
A Tight $(3/2 + \epsilon)$ -Approximation Algorithm for Demand Strip Packing <i>Franziska Eberle</i>	101
Students in highly competitive markets: the case of New York City specialized high schools <i>Yuri Faenza</i>	101
Fair Resource Allocation: From Theory to Practice <i>Swati Gupta</i>	102
Online Scheduling via Gradient Descent <i>Sungjin Im</i>	102
Fair solutions to the house allocation problem <i>Telikepalli Kavitha</i>	103
Supermodular Approximation of Norms and Applications <i>Thomas Kesselheim</i>	103
FPT Algorithms using Minimal Parameters for a Generalized Version of Maximin Shares <i>Alexandra Lassota</i>	104
A Little Clairvoyance Is All You Need <i>Alexander Lindermayr</i>	104
The Power of Proportional Fairness and Unifying Scheduling Algorithms for Group Completion Times <i>Nicole Megow</i>	105
Minimum Cost Adaptive Submodular Cover <i>Viswanath Nagarajan</i>	105
Near-Optimal PCM Wear-Leveling Under Adversarial Attacks <i>Seffi Naor</i>	106
Robust Gittins for Stochastic Scheduling <i>Heather Newman</i>	107
Fair Caching <i>Debmalya Panigrahi</i>	107

The Santa Claus Problem – Three Perspectives	
<i>Lars Rohwedder</i>	108
Optimal Online Discrepancy Minimization	
<i>Thomas Rothvoss</i>	108
Stochastic scheduling with Bernoulli-type jobs through policy stratification	
<i>Kevin Schewior</i>	108
Proportional Representation in Budget Allocation	
<i>Ulrike Schmidt-Kraepelin</i>	109
A new deterministic approximation for graph burning	
<i>Jiri Sgall</i>	109
A Simple Algorithm for Dynamic Carpooling with Recourse	
<i>Cliff Stein</i>	110
Six Candidates Suffice to Win a Voter Majority	
<i>Adrian Vetta</i>	110
The Power of Migrations in Dynamic Bin Packing	
<i>Rudy Zhou</i>	110
Participants	112

3 Overview of Talks

3.1 Lossless Robustification of Packet Scheduling Algorithms

Yossi Azar (*Tel Aviv University, IL*)

License © Creative Commons BY 4.0 International license
© Yossi Azar

Joint work of Yossi Azar, Or Vardi

Heuristics on what online algorithms should do at any given time can give large improvements to the performance of the algorithm. Today, such heuristics are mostly generated by some machine learning algorithm that was trained on what is hoped to be a similar input. We consider the online packets scheduling problem where unit size packets arrive over time, each is associated with a value and a deadline. The goal is to schedule the packets to maximize the value of the packets transmitted by their deadline. We consider an arbitrary algorithm (heuristic) and robustify it without loss. Specifically, we provide an algorithm that is at least as good as the heuristic for any input, while proving $O(1)$ competitiveness no matter how bad the heuristic is. For subclass of certain algorithms (called prediction upon arrival heuristic), we even provide a better robustness bound that provably cannot be achieved for general heuristics. Finally, we show that it is not possible to be as good as the prediction and remain $O(1)$ competitive if we consider the asynchronous model.

3.2 Fair Strategic Facility Location with Predictions

Eric Balkanski (*Columbia University – New York, US*)

License © Creative Commons BY 4.0 International license
© Eric Balkanski

Joint work of Priyank Agrawal, Eric Balkanski, Vasilis Gkatzelis, Tingting Ou, Golnoosh Shahkarami, Xizhi Tan

Main reference Priyank Agrawal, Eric Balkanski, Vasilis Gkatzelis, Tingting Ou, Xizhi Tan: “Learning-Augmented Mechanism Design: Leveraging Predictions for Facility Location”, *Math. Oper. Res.*, Vol. 49(4), pp. 2626–2651, 2024.

URL <https://doi.org/10.1287/MOOR.2022.0225>

Main reference Eric Balkanski, Vasilis Gkatzelis, Golnoosh Shahkarami: “Randomized Strategic Facility Location with Predictions”, in *Proc. of the Advances in Neural Information Processing Systems*, Vol. 37, pp. 35639–35664, Curran Associates, Inc., 2024.

URL https://proceedings.neurips.cc/paper_files/paper/2024/file/3ec7806669b4048cdba4d1defc76ace3-Paper-Conference.pdf

In the strategic facility location problem, a set of agents report their locations in a metric space and the goal is to use these reports to open a new facility, minimizing an aggregate distance measure from the agents to the facility. However, agents are strategic and may misreport their locations to influence the facility’s placement in their favor. The aim is to design truthful mechanisms, ensuring agents cannot gain by misreporting. This problem was recently revisited through the learning-augmented framework, aiming to move beyond worst-case analysis and design truthful mechanisms that are augmented with (machine-learned) predictions. In this talk, I will focus on recent results for the egalitarian social cost objective, where the goal is to minimize the distance between the facility and the location of the agent who is the farthest from the facility.

3.3 Lift-and-Project Integrality Gaps for Santa Claus

Etienne Bamas (ETH Zürich, CH)

License © Creative Commons BY 4.0 International license
© Etienne Bamas

Main reference Étienne Bamas: “Lift-and-Project Integrality Gaps for Santa Claus”, in Proc. of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2025, New Orleans, LA, USA, January 12-15, 2025, pp. 572–615, SIAM, 2025.

URL <https://doi.org/10.1137/1.9781611978322.18>

In this talk, I will focus on the MaxMinDegree Arborescence (MMDA) problem in layered directed graphs of depth $\ell \leq O(\log n / \log \log n)$, which is a key special case of the Santa Claus problem. The only way we have to solve the MMDA problem within a polylogarithmic factor is via an elegant recursive rounding of the $(\ell - 1)^{th}$ level of the Sherali-Adams hierarchy. However, it remains plausible that one could obtain a polylogarithmic approximation in polynomial time by using the same rounding with only 1 round of the Sherali-Adams hierarchy. As a main result, we rule out this possibility by constructing an MMDA instance of depth 3 for which a polynomial integrality gap survives 1 round of the Sherali-Adams hierarchy. This result is tight since it is known that after only 2 rounds the gap is at most polylogarithmic on depth-3 graphs. I will conclude the talk by related open problems.

3.4 Minimax Group Fairness in Strategic Classification

Emily Diana (TTIC – Chicago, US)

License © Creative Commons BY 4.0 International license
© Emily Diana

Joint work of Emily Diana, Saeed Sharifi-Malvajerdi, Ali Vakilian

Main reference Emily Diana, Saeed Sharifi-Malvajerdi, Ali Vakilian: “Minimax Group Fairness in Strategic Classification”, in Proc. of the IEEE Conference on Secure and Trustworthy Machine Learning, SaTML 2025, Copenhagen, Denmark, April 9-11, 2025, pp. 753–772, IEEE, 2025.

URL <https://doi.org/10.1109/SaTML64287.2025.00047>

In strategic classification, agents manipulate their features, at a cost, to receive a positive classification outcome from the learner’s classifier. The goal of the learner in such settings is to learn a classifier that is robust to strategic manipulations. While the majority of works in this domain consider accuracy as the primary objective of the learner, in this work, we consider learning objectives that have group fairness guarantees in addition to accuracy guarantees. We work with the minimax group fairness notion that asks for minimizing the maximal group error rate across population groups. Motivating examples will be focused on situations where agents are competing for resources and the classification decision influences allocation policies.

3.5 A Tight $(3/2 + \varepsilon)$ -Approximation Algorithm for Demand Strip Packing

Franziska Eberle (TU Berlin, DE)

License © Creative Commons BY 4.0 International license
 © Franziska Eberle
Joint work of Franziska Eberle, Felix Hommelsheim, Malin Rau, Stefan Walzer
Main reference Franziska Eberle, Felix Hommelsheim, Malin Rau, Stefan Walzer: “A Tight $(3/2 + \varepsilon)$ -Approximation Algorithm for Demand Strip Packing”, in Proc. of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2025, New Orleans, LA, USA, January 12-15, 2025, pp. 641–699, SIAM, 2025.
URL <https://doi.org/10.1137/1.9781611978322.20>

We consider the Demand Strip Packing problem (DSP), in which we are given a set of jobs, each specified by a processing time and a demand. The task is to schedule all jobs such that they are finished before some deadline D while minimizing the peak demand, i.e., the maximum total demand of tasks executed at any point in time. DSP is closely related to the Strip Packing problem (SP), in which we are given a set of axis-aligned rectangles that must be packed into a strip of fixed width while minimizing the maximum height. DSP and SP are known to be NP-hard to approximate to within a factor below $\frac{3}{2}$.

To achieve the essentially best possible approximation guarantee, we prove a structural result. Any instance admits a solution with peak demand at most $(\frac{3}{2} + \varepsilon)\text{OPT}$ satisfying one of two properties. Either (i) the solution leaves a gap for a job with demand OPT and processing time $\mathcal{O}(\varepsilon D)$ or (ii) all jobs with demand greater than $\frac{\text{OPT}}{2}$ appear sorted by demand in immediate succession. We then provide two efficient algorithms that find a solution with maximum demand at most $(\frac{3}{2} + \varepsilon)\text{OPT}$ in the respective case. A central observation, which sets our approach apart from previous ones for DSP, is that the properties (i) and (ii) need not be efficiently decidable: We can simply run both algorithms and use whichever solution is the better one.

3.6 Students in highly competitive markets: the case of New York City specialized high schools

Yuri Faenza (Columbia University – New York, US)

License © Creative Commons BY 4.0 International license
 © Yuri Faenza
Joint work of Yuri Faenza, Swati Gupta, Xuan Zhang
Main reference Yuri Faenza, Swati Gupta, Xuan Zhang: “Discovering Opportunities in New York City’s Discovery Program: Disadvantaged Students in Highly Competitive Markets”, in Proc. of the 24th ACM Conference on Economics and Computation, EC 2023, London, United Kingdom, July 9-12, 2023, p. 585, ACM, 2023.
URL <https://doi.org/10.1145/3580507.3597762>

Eight among the most competitive high schools of the New York City Department of Education (NYC DOE) admit students only based on their score on a test, called SHSAT. 20% of these seats are reserved for students that the NYC DOE classifies, mostly following economic criteria, as disadvantaged. We show that the mechanism currently employed by the NYC DOE to assign these reserved seats creates a significant incentive for disadvantaged students to underperform, and we study alternatives. In particular, we highlight the superiority of one such alternative under the new ex-post hypothesis of High competitiveness (HC) of the market. We also give sufficient ex-ante conditions under which the HC hypothesis is satisfied with high probability. To prove such results, we rely on generalizations of Gale and Shapley’s

marriage model involving choice functions, and on the classical occupancy problem. Using 12 years of data, we show that the NYC DOE market that originated our work satisfies the HC hypothesis.

3.7 Fair Resource Allocation: From Theory to Practice

Swati Gupta (MIT – Cambridge, US)

License © Creative Commons BY 4.0 International license
© Swati Gupta

Joint work of Swati Gupta, Jai Moondra, Mohit Singh, Cheol Woo Kim, Shresth Verma, Madeleine Pollack, Lingkai Kong, Milind Tambe

Fairness in resource allocation is a fundamental problem that arises in a variety of domains, including healthcare, hiring, admissions, infrastructure development, recommendation systems, disaster management, and emergency response. Different ethical theories provide distinct lenses through which fairness can be understood and operationalized. In this talk, I will discuss (i) what it means to be fair in static and dynamic settings, depending on the application context, (ii) theoretical models for understanding noise and bias in data, and (iii) connections with law and policy. Through some of my recent work, I will discuss challenges related to differences in fairness objectives (e.g., how to find some “good” enough solutions across all objectives), navigating the space of human-AI collaboration (e.g., what should AI optimize?), and deviations from theoretical assumptions (e.g., of clean group memberships, discrimination models, etc).

3.8 Online Scheduling via Gradient Descent

Sungjin Im (University of California at Santa Cruz, US)

License © Creative Commons BY 4.0 International license
© Sungjin Im

Joint work of Qingyun Chen, Sungjin Im, Aditya Petety

Main reference Qingyun Chen, Sungjin Im, Aditya Petety: “Online Scheduling via Gradient Descent for Weighted Flow Time Minimization”, in Proc. of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2025, New Orleans, LA, USA, January 12-15, 2025, pp. 3802–3841, SIAM, 2025.

URL <https://doi.org/10.1137/1.9781611978322.128>

In this talk, I will show how a generalization of the shortest remaining time first (SRPT) scheduling algorithm can be effectively used for various scheduling problems to minimize total weighted flow time. Essentially, SRPT can be interpreted as gradient descent on an estimate of the remaining jobs’ cost. In particular, we show that gradient descent is effective when the residual estimate possesses supermodularity, and that this supermodularity can be achieved when the scheduling constraints induce gross substitute valuations in the Walrasian Market.

3.9 Fair solutions to the house allocation problem

Telikepalli Kavitha (Tata Institute of Fundamental Research – Mumbai, IN)

License © Creative Commons BY 4.0 International license
© Telikepalli Kavitha
Joint work of Tamas Kiraly, Jannik Matuschke, Ildiko Schlotter, Ulrike Schmidt-Kraepelin
Main reference Telikepalli Kavitha, Tamás Király, Jannik Matuschke, Ildikó Schlotter, Ulrike Schmidt-Kraepelin: “The popular assignment problem: when cardinality is more important than popularity”, in Proc. of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 – 12, 2022, pp. 103–123, SIAM, 2022.
URL <https://doi.org/10.1137/1.9781611977073.6>

Matching problems with one-sided preferences are seen in many applications such as campus housing allocation in universities. Popularity is a well-studied notion of fairness that captures collective welfare. This talk will be on some simple algorithms to find popular solutions for matching problems in this model.

3.10 Supermodular Approximation of Norms and Applications

Thomas Kesselheim (Universität Bonn, DE)

License © Creative Commons BY 4.0 International license
© Thomas Kesselheim
Joint work of Thomas Kesselheim, Marco Molinaro, Sahil Singla
Main reference Thomas Kesselheim, Marco Molinaro, Sahil Singla: “Supermodular Approximation of Norms and Applications”, in Proc. of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24–28, 2024, pp. 1841–1852, ACM, 2024.
URL <https://doi.org/10.1145/3618260.3649734>

Many classic scheduling problems can be understood as minimizing a norm objective: Most prominently, the Makespan is nothing but the ℓ_∞ norm of the vector of machine loads. Every additive objective, like for example in Set Cover, can also be understood as an ℓ_1 norm. Over the years, a lot of results have been generalized to ℓ_p norms.

In this talk, we discuss techniques and results to go beyond ℓ_p norms. With a particular focus on online problems, we identify supermodularity—often reserved for combinatorial set functions and characterized by monotone gradients—as a defining feature. Every ℓ_p -norm is p -supermodular, meaning that its p^{th} power function exhibits supermodularity. The association of supermodularity with norms offers a new lens through which to view and construct algorithms.

For a large class of problems p -supermodularity is a sufficient criterion for developing good algorithms. Moreover, we show that every symmetric norm can be $O(\log m)$ -approximated by an $O(\log m)$ -supermodular norm, resulting in $O(\text{poly } \log m)$ -competitive algorithms for load balancing and covering with respect to an arbitrary monotone symmetric norm.

3.11 FPT Algorithms using Minimal Parameters for a Generalized Version of Maximin Shares

Alexandra Lassota (TU Eindhoven, NL)

License © Creative Commons BY 4.0 International license

© Alexandra Lassota

Joint work of Klaus Jansen, Alexandra Lassota, Malte Tutas, Adrian Vetta

Main reference Klaus Jansen, Alexandra Lassota, Malte Tutas, Adrian Vetta: “FPT Algorithms using Minimal Parameters for a Generalized Version of Maximin Shares”, CoRR, Vol. abs/2409.04225, 2024.

URL <https://doi.org/10.48550/ARXIV.2409.04225>

We study the computational complexity of fairly allocating indivisible, mixed-manna items. For basic measures of fairness, this problem is hard in general. The paradigm of fixed-parameter tractability (FPT) has led to new insights and improved algorithms for a variety of fair allocation problems. Our focus is designing FPT time algorithms for finding a best solution w.r.t. the fairness measure maximin shares (MMS). Furthermore, our techniques extend to finding allocations that optimize alternative objectives, such as minimizing the additive approximation, and maximizing some variants of global welfare. Our algorithms are actually designed for a more general MMS problem in machine scheduling. Here, each mixed-manna item (job) must be assigned to an agent (machine) and has a processing time and a deadline.

3.12 A Little Clairvoyance Is All You Need

Alexander Lindermayr (Universität Bremen, DE)

License © Creative Commons BY 4.0 International license

© Alexander Lindermayr

Joint work of Anupam Gupta, Haim Kaplan, Alexander Lindermayr, Jens Schlöter, Sorrachai Yingchareonthawornchai

We revisit the classical problem of minimizing the total *flow time* of jobs on a single machine in the online setting where jobs arrive over time. It has long been known that the Shortest Remaining Processing Time (SRPT) algorithm is optimal (i.e., 1-competitive) when the job sizes are known up-front [Schrage, 1968]. But in the non-clairvoyant setting where job sizes are revealed only when the job finishes, no algorithm can be constant-competitive [Motwani, Phillips, and Torng, 1994].

We consider the ε -clairvoyant setting, where $\varepsilon \in [0, 1]$, and each job’s processing time becomes known once its remaining processing time equals an ε fraction of its processing time. This captures settings where the system user uses the initial $(1 - \varepsilon)$ fraction of a job’s processing time to learn its true length, which it can then reveal to the algorithm. The model was proposed by Yingchareonthawornchai and Torng (2017), and it smoothly interpolates between the clairvoyant setting (when $\varepsilon = 1$) and the non-clairvoyant setting (when $\varepsilon = 0$). In a concrete sense, we are asking: *how much knowledge is required to circumvent the hardness of this problem?*

We show that *a little knowledge is enough*, and that a constant competitive algorithm exists for every constant $\varepsilon > 0$. More precisely, for all $\varepsilon \in (0, 1)$, we present an deterministic $\lceil 1/\varepsilon \rceil$ -competitive algorithm, which is optimal for deterministic algorithms. We also present a matching lower bound (up to a constant factor) for randomized algorithms.

Our algorithm to achieve this bound is remarkably simple and applies the “optimism in the face of uncertainty” principle. For each job, we form an optimistic estimate of its length, based on the information revealed thus far and run SRPT on these optimistic estimates. The

proof relies on maintaining a matching between the jobs in OPT’s queue and the algorithm’s queue, with small prefix expansion. We achieve this by carefully choosing a set of jobs *to arrive earlier than their release times* without changing the algorithm, and possibly helping the adversary. These early arrivals allow us to maintain structural properties inductively, giving us the tight guarantee.

3.13 The Power of Proportional Fairness and Unifying Scheduling Algorithms for Group Completion Times

Nicole Megow (*Universität Bremen, DE*)

License © Creative Commons BY 4.0 International license
© Nicole Megow

Joint work of Sven Jäger, Alexander Lindermayr, Zhenwei Liu, Nicole Megow

Main reference Emily Diana, Saeed Sharifi-Malvajerdi, Ali Vakilian: “Minimax Group Fairness in Strategic Classification”, in Proc. of the IEEE Conference on Secure and Trustworthy Machine Learning, SaTML 2025, Copenhagen, Denmark, April 9-11, 2025, pp. 753–772, IEEE, 2025.

URL <https://doi.org/10.1109/SaTML64287.2025.00047>

Main reference Sven Jäger, Alexander Lindermayr, Nicole Megow: “The Power of Proportional Fairness for Non-Clairvoyant Scheduling under Polyhedral Constraints”, in Proc. of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2025, New Orleans, LA, USA, January 12-15, 2025, pp. 3901–3930, SIAM, 2025.

URL <https://doi.org/10.1137/1.9781611978322.132>

We propose new abstract problems that unify a collection of scheduling and graph coloring problems with general min-sum objectives. Specifically, we consider the weighted sum of completion times over groups of entities (jobs, vertices, or edges), which generalizes two important objectives in scheduling: makespan and sum of weighted completion times.

We study these problems in both online and offline settings. In the non-clairvoyant online setting, we give a novel $O(\log g)$ -competitive algorithm, where g is the size of the largest group. This is the first non-trivial competitive bound for many problems with group completion time objective, and it is an exponential improvement over previous results for non-clairvoyant coflow scheduling. Notably, this bound is asymptotically best-possible. For offline scheduling, we provide powerful meta-frameworks that lead to new or stronger approximation algorithms for our new abstract problems and for previously well-studied special cases. In particular, we improve the approximation ratio from 13.5 to 10.874 for non-preemptive related machine scheduling and from $4 + \varepsilon$ to $2 + \varepsilon$ for preemptive unrelated machine scheduling (MOR 2012), and we improve the approximation ratio for sum coloring problems from 10.874 to 5.437 for perfect graphs and from 11.273 to 10.874 for interval graphs (TALG 2008).

3.14 Minimum Cost Adaptive Submodular Cover

Viswanath Nagarajan (*University of Michigan – Ann Arbor, US*)

License © Creative Commons BY 4.0 International license
© Viswanath Nagarajan

Joint work of Hessa Al-Thani, Yubing Cui, Blake Harris, Viswanath Nagarajan

Adaptive submodularity is a fundamental concept in stochastic optimization, with numerous applications such as sensor placement, hypothesis identification and viral marketing. We consider the problem of covering an adaptive-submodular function at minimum expected cost, which generalizes the classic set cover and submodular cover problems to the stochastic setting.

We show that the natural greedy policy has an approximation ratio of $4 \cdot (1 + \ln Q)$, where Q is the goal value. In fact, we consider a significantly more general objective of minimizing the p^{th} moment of the coverage cost, and show that the greedy policy *simultaneously* achieves a $(p + 1)^{p+1} \cdot (\ln Q + 1)^p$ approximation guarantee for all $p \geq 1$. All our approximation ratios are best possible up to constant factors (assuming $P \neq NP$). We also show that the greedy policy for minimizing expected cost has an approximation ratio at least $1.3 \cdot (1 + \ln Q)$ even when $Q = 1$, which invalidates a prior result on adaptive submodular cover. Moreover, our results extend to the setting where one wants to cover *multiple* adaptive-submodular functions, for which we obtain the same approximation guarantees.

3.15 Near-Optimal PCM Wear-Leveling Under Adversarial Attacks

Seffi Naor (Technion – Israel Institute of Technology – Haifa, IL)

License © Creative Commons BY 4.0 International license
© Seffi Naor

Joint work of Tomer Lange, Seffi Naor, Gala Yadgar

Main reference Tomer Lange, Joseph (Seffi) Naor, Gala Yadgar: “SSD Wear Leveling with Optimal Guarantees”, in Proc. of the 2024 Symposium on Simplicity in Algorithms, SOSA 2024, Alexandria, VA, USA, January 8-10, 2024, pp. 306–320, SIAM, 2024.

URL <https://doi.org/10.1137/1.9781611977936.28>

Phase change memory (PCM) is a promising memory technology known for its speed, high density, and durability. However, each PCM cell can endure only a limited number of erase and subsequent write operations before failing, and the failure of a single cell can limit the lifespan of the entire device. This vulnerability makes PCM particularly susceptible to adversarial attacks that induce excessive writes to accelerate device failure. To counter this, wear-leveling techniques aim to distribute write operations evenly across PCM cells.

In this paper we study the *online PCM utilization problem*, which seeks to maximize the number of write requests served before any cell reaches the erase limit. While extensively studied in the systems and architecture communities, this problem remains largely unexplored from a theoretical perspective. We bridge this gap by presenting a novel algorithm that leverages cell wear information to optimize PCM utilization. We prove that our algorithm achieves near-optimal worst-case guarantees and outperforms state-of-the-art practical solutions both theoretically and empirically, providing an efficient approach to prolonging PCM lifespan.

3.16 Robust Gittins for Stochastic Scheduling

Heather Newman (Carnegie Mellon University – Pittsburgh, US)

License © Creative Commons BY 4.0 International license
© Heather Newman

Joint work of Benjamin Moseley, Heather Newman, Kirk Pruhs, Rudy Zhou

Main reference Benjamin Moseley, Heather Newman, Kirk Pruhs, Rudy Zhou: “Robust Gittins for Stochastic Scheduling”, in Proc. of the Abstracts of the 2025 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS 2025, Stony Brook, NY, USA, June 9-13, 2025, pp. 166–168, ACM, 2025.

URL <https://doi.org/10.1145/3726854.3727315>

A common theme in stochastic optimization problems is that, theoretically, stochastic algorithms need to “know” relatively rich information about the underlying distributions. This is at odds with most applications, where distributions are rough predictions based on historical data. Thus, commonly, stochastic algorithms are making decisions using imperfect predicted distributions, while trying to optimize over some unknown true distributions.

We consider the fundamental problem of scheduling stochastic jobs preemptively on a single machine to minimize expected mean completion time in the setting where *the scheduler is only given imperfect predicted job size distributions*. If the predicted distributions are perfect, then it is known that this problem can be solved optimally by the Gittins index policy.

The goal of our work is to design a scheduling policy that is robust in the sense that it produces nearly optimal schedules even if there are modest discrepancies between the predicted distributions and the underlying real distributions. Our main contributions are:

- We show that the standard Gittins index policy is *not robust* in this sense. If the true distributions are perturbed by even an arbitrarily small amount, then running the Gittins index policy using the perturbed distributions can lead to an unbounded increase in mean completion time.
- We explain how to modify the Gittins index policy to make it robust, that is, to produce nearly optimal schedules, where the approximation depends on a new measure of error between the true and predicted distributions that we define.

Looking forward, the approach we develop here can be applied more broadly to many other stochastic optimization problems to better understand the impact of mispredictions, and lead to the development of new algorithms that are robust against such mispredictions.

3.17 Fair Caching

Debmalya Panigrahi (Duke University – Durham, US)

License © Creative Commons BY 4.0 International license
© Debmalya Panigrahi


Joint work of Anupam Gupta, Amit Kumar, Debmalya Panigrahi

Online convex paging models a broad class of cost functions for the classical paging problem. In particular, it naturally captures fairness constraints: e.g., that no specific page (or groups of pages) suffers an unfairly high number of evictions by considering ℓ_p norms of eviction vectors for $p > 1$. The case of the ℓ_∞ norm has also been of special interest, and is called min-max paging.

In this talk, I will discuss tight upper and lower bounds for the convex paging problem for a broad class of convex functions. Prior to our work, only fractional algorithms were known for this general setting. Moreover, our new results settle the competitive ratio for min-max paging and ℓ_p -norm paging for all values of $p > 1$.

3.18 The Santa Claus Problem – Three Perspectives

Lars Rohwedder (Maastricht University, NL)

License  Creative Commons BY 4.0 International license
© Lars Rohwedder

Santa Claus cannot accept that even a single child is unhappy on Christmas. Therefore, when he distributes his gifts, he maximizes the total value of gifts that the least happy child gets. This is a non-trivial task, especially when each gift j has a different value v_{ij} for each child i . This very natural problem, sometimes under the more serious name of max-min fair allocation, has seen significant attention in the last two decades. Yet, many questions about it remain widely open. We will survey developments on the problem using three different perspectives that demonstrate its versatile nature: First, we view it as a fair allocation problem, then as a scheduling problem, and finally as a network design problem.

3.19 Optimal Online Discrepancy Minimization

Thomas Rothvoss (University of Washington – Seattle, US)


License  Creative Commons BY 4.0 International license
© Thomas Rothvoss

Joint work of Thomas Rothvoss, Janardhan Kulkarni, Victor Reis
Main reference Janardhan Kulkarni, Victor Reis, Thomas Rothvoss: “Optimal Online Discrepancy Minimization”, CoRR, Vol. abs/2308.01406, 2023.
URL <https://doi.org/10.48550/ARXIV.2308.01406>

We prove that there exists an online algorithm that for any sequence of vectors $v_1, \dots, v_T \in \mathbb{R}^n$ with $\|v_i\|_2 \leq 1$, arriving one at a time, decides random signs $x_1, \dots, x_T \in \{-1, 1\}$ so that for every $t \leq T$, the prefix sum $\sum_{i=1}^t x_i v_i$ is $O(1)$ -subgaussian. This improves over the work of Alweiss, Liu and Sawhney who kept prefix sums $O(\sqrt{\log(nT)})$ -subgaussian. Our proof combines a generalization of Banaszczyk’s prefix balancing result to trees with a cloning argument to find distributions rather than single colorings.

3.20 Stochastic scheduling with Bernoulli-type jobs through policy stratification

Kevin Schewior (Universität zu Köln, DE)

License  Creative Commons BY 4.0 International license
© Kevin Schewior

Joint work of Antonios Antoniadis, Ruben Hoeksma, Kevin Schewior, Marc Uetz

This paper addresses the problem of computing a scheduling policy that minimizes the total expected completion time of a set of N jobs with stochastic processing times on m parallel identical machines. When all processing times follow Bernoulli-type distributions, Gupta et al. (SODA ’23) exhibited approximation algorithms with an approximation guarantee $\tilde{O}(\sqrt{m})$, where m is the number of machines and $\tilde{O}(\cdot)$ suppresses polylogarithmic factors in N , improving upon an earlier $O(m)$ approximation by Eberle et al. (OR Letters ’19) for a special case. The present paper shows that, quite unexpectedly, the problem with Bernoulli-type jobs admits a PTAS whenever the number of different job-size parameters is bounded by a constant. The result is based on a series of transformations of an optimal

scheduling policy to a “stratified” policy that makes scheduling decisions at specific points in time only, while losing only a negligible factor in expected cost. An optimal stratified policy is computed using dynamic programming. Two technical issues are solved, namely (i) to ensure that, with at most a slight delay, the stratified policy has an information advantage over the optimal policy, allowing it to simulate its decisions, and (ii) to ensure that the delays do not accumulate, thus solving the trade-off between the complexity of the scheduling policy and its expected cost. Our results also imply a quasi-polynomial $O(\log N)$ -approximation for the case with an arbitrary number of job sizes.

3.21 Proportional Representation in Budget Allocation

Ulrike Schmidt-Kraepelin (TU Eindhoven, NL)


License  Creative Commons BY 4.0 International license
© Ulrike Schmidt-Kraepelin

The ideal of proportional representation in social choice theory is easy to state yet challenging to formalize: any α -fraction of the population should have a say in determining an α -fraction of the outcome. This principle has gained significant attention in recent years and is arguably the most studied fairness notion in social choice theory today.

This talk explores proportional representation in the context of budget allocation—a broad framework capturing various models with wide-ranging applications, including apportionment, participatory budgeting, and committee elections. We will examine several formalizations of proportionality, introduce algorithms designed to achieve proportional outcomes, and highlight key open questions in the field. Beyond that, I hope to inspire discussion on how proportional representation might be relevant in settings beyond social choice theory.

3.22 A new deterministic approximation for graph burning

Jiri Sgall (Charles University – Prague, CZ)


License  Creative Commons BY 4.0 International license
© Jiri Sgall
Joint work of Matej Lieskovsky, Jiri Sgall

Graph Burning models information spreading in a given graph as a process such that in each step one node is infected (informed) and also the infection spreads to all neighbors of previously infected nodes. Formally, given a graph $G = (V, E)$, possibly with edge lengths, the burning number $b(G)$ is the minimum number g such that there exist nodes $v_0, \dots, v_{g-1} \in V$ satisfying the property that for each $u \in V$ there exists $i \in \{0, \dots, g-1\}$ so that the distance between u and v_i is at most i .

We present a simple deterministic 2.314-approximation algorithm for computing the burning number of a general graph, even with arbitrary edge lengths. This complements our previous more complicated randomized algorithm with the same approximation ratio.

3.23 A Simple Algorithm for Dynamic Carpooling with Recourse

Cliff Stein (Columbia University – New York, US)

License  Creative Commons BY 4.0 International license
© Cliff Stein

Joint work of Cliff Stein, Shyamal Patel, Yuval Efron

Main reference Yuval Efron, Shyamal Patel, Cliff Stein: “A Simple Algorithm for Dynamic Carpooling with Recourse”, in Proc. of the 2025 Symposium on Simplicity in Algorithms, SOSA 2025, New Orleans, LA, USA, January 13-15, 2025, pp. 196–201, SIAM, 2025.


URL <https://doi.org/10.1137/1.9781611978315.15>

We give an algorithm for the fully-dynamic carpooling problem with recourse: Edges arrive and depart online from a graph G with n nodes according to an adaptive adversary. Our goal is to maintain an orientation H of G that keeps the discrepancy, defined as $\max_{v \in V} |\deg_H^+(v) - \deg_H^-(v)|$, small at all times.

We present a simple algorithm and analysis for this problem with recourse based on cycles that simplifies and improves on a result of Gupta et al. [SODA '22].

3.24 Six Candidates Suffice to Win a Voter Majority

Adrian Vetta (McGill University – Montreal, CA)

License  Creative Commons BY 4.0 International license
© Adrian Vetta

Joint work of Moses Charikar, Alexandra Lassota, Prasanna Ramakrishnan, Adrian Vetta, Kangning Wang


Main reference Moses Charikar, Alexandra Lassota, Prasanna Ramakrishnan, Adrian Vetta, Kangning Wang: “Six Candidates Suffice to Win a Voter Majority”, CoRR, Vol. abs/2411.03390, 2024.

URL <https://doi.org/10.48550/ARXIV.2411.03390>

Given an election of n voters with preference lists over m candidates, Elkind, Lang, and Saffidine (2011) defined a Condorcet winning set to be a collection of candidates that the majority of voters prefer over any individual candidate. Condorcet winning sets of cardinality one (a Condorcet winner) or cardinality two need not exist. We prove however that a Condorcet winning set of cardinality at most six exists in any election.

3.25 The Power of Migrations in Dynamic Bin Packing

Rudy Zhou (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 4.0 International license
© Rudy Zhou

Joint work of Konstantina Mellou, Marco Molinaro, Rudy Zhou

Main reference Konstantina Mellou, Marco Molinaro, Rudy Zhou: “The Power of Migrations in Dynamic Bin Packing”, Proc. ACM Meas. Anal. Comput. Syst., Vol. 8(3), pp. 45:1–45:28, 2024.

URL <https://doi.org/10.1145/3700435>

In the *Dynamic Bin Packing* problem, n items arrive and depart the system in an online manner, and the goal is to maintain a good packing throughout. We consider the objective of minimizing the total active time, i.e., the sum of the number of open bins over all times. An important tool for maintaining an efficient packing in many applications is the use of *migrations*; e.g., transferring computing jobs across different machines. However, there are large gaps in our understanding of the approximability of dynamic bin packing with migrations. Prior work has covered the power of no migrations and $> n$ migrations, but we ask the question: What is the power of limited ($\leq n$) migrations?

Our first result is a dichotomy between no migrations and linear migrations: Using a sublinear number of migrations is asymptotically equivalent to doing *zero* migrations, where the competitive ratio grows with μ , the ratio of the largest to smallest item duration. On the other hand, we prove that for every $\alpha \in (0, 1]$, there is an algorithm that does $\approx \alpha n$ migrations and achieves competitive ratio $\approx 1/\alpha$ (in particular, independent of μ); we also show that this tradeoff is essentially best possible. This fills in the gap between zero migrations and $> n$ migrations in Dynamic Bin Packing.

Finally, in light of the above impossibility results, we introduce a new model that more directly captures the impact of migrations. Instead of limiting the number of migrations, each migration adds a delay of C time units to the item's duration; this commonly appears in settings where a blackout or set-up time is required before the item can restart its execution in the new bin. In this new model, we prove a $O(\min(\sqrt{C}, \mu))$ -approximation, and an almost matching lower bound. We also present preliminary experiments that indicate that our theoretical results are predictive of the practical performance of our algorithms.

Participants

- Antonios Antoniadis
University of Twente, NL
- Yossi Azar
Tel Aviv University, IL
- Eric Balkanski
Columbia University –
New York, US
- Etienne Bamas
ETH Zürich, CH
- Sanjoy Baruah
Washington University –
St. Louis, US
- Emily Diana
TTIC – Chicago, US
- Franziska Eberle
TU Berlin, DE
- Yuri Faenza
Columbia University –
New York, US
- Naveen Garg
Indian Institute of Technology –
New Delhi, IN
- Swati Gupta
MIT – Cambridge, US
- Sungjin Im
University of California –
Santa Cruz, US
- Thomas Kesselheim
Universität Bonn, DE
- Samir Khuller
Northwestern University –
Evanston, US
- Alexandra Lassota
TU Eindhoven, NL
- Alexander Lindermayr
Universität Bremen, DE
- Alberto Marchetti-Spaccamela
Sapienza University of Rome, IT
- Claire Mathieu
CNRS – Paris, FR
- Nicole Megow
Universität Bremen, DE
- Benjamin J. Moseley
Carnegie Mellon University –
Pittsburgh, US
- Viswanath Nagarajan
University of Michigan –
Ann Arbor, US
- Seffi Naor
Technion – Haifa, IL
- Heather Newman
Carnegie Mellon University –
Pittsburgh, US
- Debmalya Panigrahi
Duke University – Durham, US
- Kirk Pruhs
University of Pittsburgh, US
- Lars Rohwedder
Maastricht University, NL
- Thomas Rothvoss
University of Washington –
Seattle, US
- Kevin Schewior
Universität Köln, DE
- Ulrike Schmidt-Kraepelin
TU Eindhoven, NL
- Jiri Sgall
Charles University – Prague, CZ
- David Shmoys
Cornell University – Ithaca, US
- Martin Skutella
TU Berlin, DE
- Frits C. R. Spieksma
TU Eindhoven, NL
- Clifford Stein
Columbia University –
New York, US
- Leen Stougie
CWI – Amsterdam, NL
- Ola Svensson
EPFL – Lausanne, CH
- Kavitha Teliakapalli
TIFR Mumbai, IN
- Marc Uetz
University of Twente –
Enschede, NL
- Adrian Vetta
McGill University –
Montreal, CA
- Tjark Vredeveld
Maastricht Univ. School of
Business & Economics, NL
- Andreas Wiese
TU München, DE
- Hang Zhou
Ecole Polytechnique –
Palaiseau, FR
- Rudy Zhou
Carnegie Mellon University –
Pittsburgh, US



Climate Change: What is Computing's Responsibility?

Vicki Hanson^{*1}, and Bran Knowles^{*2}

1 ACM – New York, US. vlh@acm.org

2 Lancaster University, GB. b.h.knowles1@lancaster.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 25122 “Climate Change: What is Computing's Responsibility?” The workshop brought together global experts from computing, environmental science, and policy to explore the detrimental impacts of computing technologies on the environment, particularly with respect to climate change. These harms were considered alongside possibilities for computing technologies to facilitate climate mitigation and adaptation, as well as on balance with the social benefits delivered by computing technologies. Key topics of discussion included the role of computing in enabling a safe and just transition to a sustainable society, methodological challenges in estimating environmental impacts (beneficial and detrimental; direct and indirect), and matters of accountability and governance. Through discussions, participants converged on a vision for a paradigm shift that would align computing with climate goals, and detailed fundamental premises and commitments by computing professionals within this new paradigm.

Seminar March 16–19, 2025 – <https://www.dagstuhl.de/25122>

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Human-centered computing → Human computer interaction (HCI); Software and its engineering

Keywords and phrases sustainability, climate change, efficiency, supply chain management, climate modelling

Digital Object Identifier 10.4230/DagRep.15.3.113

1 Executive Summary

Vicki Hanson (ACM – New York, US)

Bran Knowles (Lancaster University, GB)

License  Creative Commons BY 4.0 International license
© Vicki Hanson and Bran Knowles

The Dagstuhl Perspectives Workshop 25122, held 16–19 March 2025, convened global experts from computing, environmental science, and policy to address computing's role and responsibility in the climate crisis. Participants discussed the environmental impacts of computing technologies alongside vaunted possibilities for climate mitigation and adaptation. Through these discussions, participants converged on a shared vision of the responsibility of computing professionals within the present reality of climate crisis. This vision was articulated in a Manifesto outlining core recognitions and commitments.

Key Themes and Insights

- Computing's Role in Climate Change:
 - Computing technologies have a growing negative impact on the environment; these impacts deserve more serious consideration at all levels.

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Climate Change: What is Computing's Responsibility?, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 113–124

Editors: Vicki Hanson and Bran Knowles



Dagstuhl Reports

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

- While there are positive environmental use cases for computing, rhetoric on the potential of computing to mitigate climate change is unsubstantiated and likely overly optimistic.
- Computing technologies tend to amplify and accelerate; so in a society on a path towards exceeding climate targets, computing technologies are likely getting us there faster.
- Responsibility and Ethics:
 - Discussions emphasised the need for computing professionals to embrace a broader ethical responsibility that includes not only minimising harm but also prioritising sustainability and promoting systemic change.
 - Responsible computing must at the same time promote a socially just transition to a sustainable future
- Measurement and Accountability:
 - Participant experts highlighted the methodological challenges in assessing computing's environmental footprint, including direct and indirect impacts (e.g. rebound effects).
 - Calls were made for stronger professional standards, improved methodologies for assessing impacts, and greater transparency.
- Additional leverage points:
 - Participants discussed the global technology policy landscape and the need for computing professionals to actively influence regulation.
 - Promoting a paradigm shift will require a revolution in computing education.
- Producing a Manifesto:
 - The workshop culminated in the co-development of a Manifesto outlining shared commitments and principles for sustainable computing.
 - The Manifesto is intended to serve as both a guiding document for practitioners and a tool for influencing policy, advocating for transparency, repairable and reusable technologies, and prioritising the public good over profit.

Conclusions

The workshop underscored that computing is at a crossroads: it can continue contributing to the climate crisis or become a powerful agent of change. Achieving the latter requires a collective shift in values, rigorous evaluation of impacts, systemic policy engagement, and a strong ethical framework guiding innovation.

Acknowledgements

This workshop was initiated and partly funded by the ACM Europe Council [1], which promotes dialogue and the exchange of ideas on technology and computing policy issues with the European Commission and other governmental bodies in Europe.

References

- 1 ACM. Acm europe council, 2025.

2 Table of Contents

Executive Summary
 Vicki Hanson and Bran Knowles 113

Day 1: Understanding Computing’s Responsibility
 Keynote 1: Mike Berners-Lee 116
 Keynote 2: Vint Cerf 117
 Discussion 118

Day 2: Effecting Positive Change
 Keynote 3: Vlad C. Coroamă 119
 Keynote 4: Emma Strubell 120
 Keynote 5: Tom Romanoff 121
 Discussion 122

Day 3: 19 March 2025
 Manifesto Development and Long-Term Vision for Computing 123

Participants 124

3 Day 1: Understanding Computing’s Responsibility

On the first day, participants worked towards establishing a shared vision of computing’s responsibility in the face of climate change. In the morning, the group explored this from a climate-focused perspective. Prompted by a keynote from Mike Berners-Lee, we considered questions such as:

- *What does the climate emergency require in terms of societal change?*
- *How is computing hindering progress in addressing this emergency?* and
- *What role can computing play in realising a positive vision for the future?*

The afternoon explored the workshop’s topic from a technology-focused perspective. A keynote by Vint Cerf provided a springboard for exploring questions such as:

- *How might computing critically assess new technologies?*
- *How should computing balance the drive for innovation with long-term environmental goals?* and
- *What are some levers for effecting positive change in computing?*

3.1 Keynote 1: Mike Berners-Lee

Mike Berners-Lee is the founder and director of Small World Consulting, which provides supply chain carbon metrics and management for its many clients. He has also authored a number of best-selling books including *How Bad Are Bananas?*, *The Burning Question*, and *There Is No Planet B*. His new book, *A Climate of Truth*, explains how decades of climate discussions and international conferences have failed to curb rising greenhouse gas emissions, and what this means about addressing the root causes of the climate emergency. Drawing from this latest book [1], Berners-Lee’s talk outlined his understanding of why progress on climate has been limited and how society can begin to implement real solutions.

Berners-Lee presented data showing roughly exponential growth in greenhouse gas emissions over 60 years with no discernible indication that 30 years of climate COPs have had any impact on this trend. Record breaking temperatures year-on-year indicate a faster acceleration of planetary warming than had been predicted, in part due to feedbacks such as melting permafrost releasing methane. He urged participants to understand the climate crisis as part of a broader “polycrisis” that includes biodiversity loss, food insecurity, pollution, and global inequality. He highlighted plastic pollution as particularly insidious, with microplastics devastating human and ecological health.

According to Berners-Lee, we have arrived at the brink of a critical tipping point where human technological power exceeds the resilience of Earth’s natural systems. Despite the urgency of this crisis, the global policy response has been ineffectual, signaling the need to approach this problem differently. A transition to renewable energy, for example, is necessary but insufficient without also drastically curbing energy demand. Reducing energy demand, in turn, requires rethinking a traditional focus on GDP growth as a measure of prosperity, which could instead be measured in terms of improved well-being, resilience, and sustainability. In this light, the implementation of a global carbon price to drive down fossil fuel consumption, shifting towards a circular economy, and redesigning products to be more durable and reusable could help arrest climate change while improving prosperity. Without these systemic economic changes, efforts to mitigate climate catastrophe are destined to fail.

The keynote also emphasised the urgent need for political and cultural change, and the role technology must play in this transition. Berners-Lee called for a renewed societal commitment to truth and accountability, in particular countering the spread of misinformation

through social media and enabling increased transparency in journalism. He also stressed the importance of citizens engaging politically to push for systemic changes, and that computing technologies could facilitate this engagement.

The keynote ended by reiterating that solutions to the climate emergency exist and can be implemented so long as humanity is willing to acknowledge the truth and act decisively. Berners-Lee's call to action is for computing to foster a culture that values truth, sustainability, and responsible decision-making at every level of society.

References

- 1 Mike Berners-Lee. *A Climate of Truth: Why We Need It and How To Get It*. Cambridge University Press, 2025.

3.2 Keynote 2: Vint Cerf

Vint Cerf is a Vice President and Chief Internet Evangelist for Google. He is known as one of the “Fathers of the Internet” for his work on TCP/IP protocols and the architecture of the Internet. For this and his continuing work on Internet development and his efforts to increase access to the Internet for everyone, he has received numerous accolades worldwide. Notably, he is a recipient of the ACM Alan M. Turing Award. His keynote focused on the impact of developing technologies both in contributing to the climate crisis and, potentially, to providing solutions.

Cerf discussed the role of machine learning in understanding climate change through detecting complex patterns (e.g. “atmospheric rivers”) and forecasting future trends. He advocated for extensive data collection on temperature changes, atmospheric conditions, and other climate variables to improve predictions and develop more effective interventions. He further stressed the need to refine theoretical models to improve their predictive power and, thus, help policymakers craft effective, targeted responses.

Like Berners-Lee, Cerf raised alarm regarding the urgency of the climate crisis. He noted, however, that even if changes were implemented immediately to drastically reduce emissions, the climate has already altered in ways that require serious consideration of climate adaptation strategies. He suggested that computing could be particularly useful here in modeling risks and benefits of novel interventions through large-scale simulations, thereby helping to avoid unintended negative consequences. He also noted the potential for advanced computational tools to assess regional risks and guide planning.

Cerf also echoed Berners-Lee in calling for solutions to combat misinformation. Chief among these is the need for increased public education to be able to evaluate online sources and verify data. Computing could further promote fact-based climate discourse by using digital signatures to authenticate images and reports, labeling AI-generated content, and flagging informational inaccuracies.

Cerf noted the key role of industry in driving change. Acknowledging that responding proactively to climate change contributes to the long-term profitability of companies, he urged businesses to integrate sustainability into their operations (e.g. pursuing improved hardware efficiency and the use of renewable energy) before such changes are legally mandated. He also underscored the need to develop robust and adaptable technologies that can withstand disruptions that might arise from environmental or political instability to avoid catastrophic failures in critical infrastructures.

The keynote concluded by emphasising that developing effective solutions will require collaboration between environmental scientists, economists, technologists, and policymakers, and that computing can be the bridge between these disciplines through modeling, predicting, and assessing synergistic solutions. In this way, computing can help with navigating the complexities of climate change while simultaneously building the requisite infrastructure for a more resilient, sustainable future.

3.3 Discussion

The first day involved considerable discussion of computing professionals' responsibilities with respect to the climate crisis. One dimension of this responsibility is ensuring that the environmental footprint of computing technologies themselves are minimised through actively considering these impacts throughout the development process. Responsible innovation frameworks were seen as useful in guiding development of new technologies in line with environmental considerations. There was some debate as to whether existing codes of ethics, such as those put out by ACM [1] and IEEE [2], adequately address the environmental impacts of computing.

Resilient system design emerged as an important challenge for the sector. Many agreed that the field places too much emphasis on rapid iteration, often leading to unsustainable infrastructure. Participants emphasised the need to create systems that can function reliably over long periods without requiring frequent maintenance. They also stressed that overly complex systems tend to be resource-intensive, and suggested that change in the direction of simplicity would be environmentally beneficial.

Regulation was another important theme of the day's discussions. There was a suggestion that computing professionals should adopt regulatory structures similar to those in the medical field, where ethical review processes are a prerequisite for new developments. While this could ensure greater accountability for the long-term impact of technological advancements, there were as yet unresolved questions regarding the practical implications of this suggestion. There was agreement, however, about the importance of incorporating environmental and social impact assessments into all technological developments.

There was also debate about stronger regulation of specific classes of computing technologies. Technologies such as blockchain, cryptocurrencies, and generative AI were debated extensively due to their outsized environmental impacts. Discussion highlighted that impacts, both positive and negative (environmental and otherwise), are highly dependent on their specific applications and implementations, complicating any potential regulation.

Participants also discussed the importance of systemic, industry-wide change driven by policy reforms. Ensuring accountability was noted as a particular challenge, and participants explored how companies could be prevented from bypassing ethical responsibilities, as they might do through strategic public relations efforts.

References

- 1 ACM Code 2018 Task Force. ACM Code of Ethics and Professional Conduct. Professional code, Association for Computing Machinery, 6 2018. Adopted by the ACM Council on June 22nd, 2018.
- 2 ACM Code 2018 Task Force. IEEE Code of Ethics and Professional Conduct – 7.8 IEEE Code of Ethics. Professional code, IEEE, 2018.

4 Day 2: Effecting Positive Change

The second day of the workshop focused on the practical matters of effecting the kinds of positive changes that were articulated the previous day. Discussions centred considerations of accountability and governance, the role of researchers and policymakers, and the influence of professional bodies such as the Association for Computing Machinery (ACM).

The morning session focused on methodologies for measuring the environmental footprint of computing technologies and some of their strengths and weaknesses. Keynotes by Vlad C. Coroamă and Emma Strubell provided a basis for discussion on what oversight might look like if the computing sector were to coordinate efforts towards improved climate responsibility. The afternoon session began with a keynote by Tom Romanoff which prompted discussion on opportunities for influencing global technology policy.

4.1 Keynote 3: Vlad C. Coroamă

Vlad C. Coroamă is the founder of the Roegen Centre for Sustainability (Zurich, Switzerland) and affiliated researcher and lecturer with the TU Berlin, Germany. His research revolves around the relation between computing and the environment. He contributed both methodologically and with concrete assessments to understanding the environmental impact of ICT. Today, his main research interest lies in exploring the mechanisms through which computing can save resources, energy and emissions, and how to best exploit this potential, while understanding and avoiding the counteracting rebound effects.

The environmental effects of computing are numerous, multifaceted and intertwined. Various taxonomies for their conceptualisation have been proposed. This presentation used one of the simplest conceptualisations, which distinguishes between direct effects, beneficial indirect effects and detrimental indirect ones [1].

Focusing on indirect effects, the presentation started by presenting the typical bottom-up estimation process, as deployed in several of the current assessment methodologies. For the example of teleworking [2], these steps are:

- First, identifying impact avoidance mechanisms (e.g., less commute, less office energy).
- Then, the baseline impact is estimated in a counterfactual without the computing service (i.e., the environmental impact of traditional commuting).
- The third and fourth step estimate the savings per usage for each mechanism identified in step 1 and the adoption rate of the service, respectively.
- Finally, the overall beneficial effect is computed as the sum over all mechanisms of the mechanisms effect per instance times the adoption rate.
- Usually as a mere afterthought, rebound effects (and the negative indirect effects they trigger) are mentioned but not assessed.

In its second part, the presentation discussed some of the flaws and limitations of this paradigm:

1. The ontologically uncertain set of mechanisms yielding indirect effects, and the epistemically uncertain assessment of those that are known.
2. The “chronic potentialitis” [3] of such assessments, which typically lie in the future and their occurrence is almost never validated in hindsight.
3. The plethora of different types of rebound effects that exist and can outweigh the positive indirect effects.

4. The difficulty in estimating the hypothetical baseline/counterfactual, often leading to its overstatement, which consequently also yields an overstated positive effect.
5. Possible time boundaries for indirect effects: When they become part of the socio-technical regime [4], should these effects no longer be considered additional?
6. The possibly difficult boundary between rebound effects and economic growth: Are rebound effects merely one mechanism of economic growth, and if so, should they be counted as indirect effects of computing at all?

To address some of the first 4 limitations, the final part of the presentation argued in favor of top-down assessments such as quantitative systems dynamics or input-output analyses. As opposed to bottom-up assessments, they can set the system boundary arbitrarily wide and thus account for the subtle and hard-to-grasp mechanisms as well. For top-down assessments, however, causal links are hard to establish, so they miss some of the explanatory power of bottom-up analyses. A hybrid approach deploying both might thus be called for.

References

- 1 Christina Bremer, George Kamiya, Pernilla Bergmark, Vlad C Coroama, Eric Masanet, and Reid Lifset. Assessing energy and climate effects of digitalization: Methodological challenges and key recommendations. *nDEE Framing Paper Series*, 2023.
- 2 Jan CT Bieser, Vlad C Coroamă, Pernilla Bergmark, and Matthias Stürmer. The greenhouse gas (GHG) reduction potential of ICT: A critical review of telecommunication companies’ GHG enablement assessments. *Journal of Industrial Ecology*, 28(5):1132–1146, 2024.
- 3 Vlad C. Coroamă. The chronic potentialitis of digital enablement, 2024.
- 4 Frank W Geels. The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental innovation and societal transitions*, 1(1):24–40, 2011.

4.2 Keynote 4: Emma Strubell

Emma Strubell is the Raj Reddy Assistant Professor in the Language Technologies Institute (within the School of Computer Science) at Carnegie Mellon University. Strubell’s research focuses on advancing machine learning and natural language processing methodology and measurement in order to promote environmentally sustainable development and deployment of AI. work has been recognised with a Madrona AI Impact Award, best paper awards at ACL and EMNLP, and in 2024 they were named one of the most powerful people in AI by Business Insider.

Modern AI approaches, powered by deep learning and large language models (LLMs), have the potential to accelerate progress by augmenting human intelligence in our efforts to overcome urgent societal challenges such as climate change. At the same time, training and deploying these increasingly capable models comes at an increasingly high computational cost, with corresponding energy demands and environmental impacts. This presentation characterises the complex relationship between AI and the environment through the lens of LLMs, with a focus on describing what we know about AI’s direct environmental impacts.

The presentation begins by establishing some high level metrics for how much we know that AI is emitting in terms of direct GHG emissions, versus how much we should in fact be curbing those emissions. While there exist many potential environmentally beneficial applications of AI, such as energy optimisation, materials discovery, and policy analysis, many these benefits have yet to be demonstrated in practice while the negative environmental impacts are already quite clear. At the same time, AI’s energy consumption is growing, despite targets to reduce emissions due to ICT (including AI data centre emissions) by 50%

by 2030 [1]. Self-reported GHG emissions from major tech companies are rising significantly due to the increasing development and deployment of generative AI such as LLMs [3, 2, 4], which are resource intensive during model development (training) and use (inference). While training is the most well-studied phase of the AI model lifecycle where data is most readily available, inference likely makes up the majority of AI energy use, and rapidly growing due to recent methodological trends such as DeepSeek.

The presentation then shows that estimates of data centre energy usage vary widely, and argues that this stems from a lack of available data. There is a need for greater transparency from major tech companies regarding their energy use and emissions to improve assessments of direct impacts. There is also a need for better benchmarking tools, following from first steps via the AI Energy Score project [5], to measure and compare the energy efficiency of different models, and to better understand alignment between AI methodology, capability, hardware, and energy requirements.

The talk explores in more detail in what ways AI directly impacts the environment (primarily: GHG emissions, water consumption, and waste production), and how those impacts arise (examples: fossil-fuel based energy powering data centres, evaporative cooling to cool hardware in data centres, and mineral extraction for hardware.)

The presentation advocates for the importance of mitigating AI's direct impacts. The current, unsustainable, trajectory is a consequence of prioritising computational scaling over improved efficiency, or environmental sustainability. This could be spurred through enforcement of stricter emissions and energy-use guidelines, such as a carbon tax. The talk ends by calling for collaboration between policymakers, researchers, and industry leaders to put AI on a sustainable trajectory.

References

- 1 IEA. Net Zero Emissions by 2050: A Roadmap for the Global Energy Sector. Technical report, IEA, 2023.
- 2 Microsoft. Microsoft 2024 Environmental Sustainability Report. Technical report, Microsoft, 2024.
- 3 Baidu. Baidu 2024 Environmental, Social and Governance Report. Technical report, Baidu, 2024.
- 4 Google. Google 2024 Environmental Report. Technical report, Google, 2024.
- 5 AI energy score project.

4.3 Keynote 5: Tom Romanoff

Tom Romanoff is the ACM Director of Policy. In his career he has led AI policy research and issued recommendations adopted in NIST governance, White House Executive Order requirements, and US legislation. He launched AI101.org to educate and inform congressional staff on AI background and issues. He has also directed research and developed recommendations on topics including cybersecurity, privacy, technology's role in climate change mitigation, content moderation, data privacy, digital divide issues, and competition in the technology sector.

Romanoff's keynote provided an overview of the global regulatory and legislative landscape, current policy challenges, and reflections on strategy for effective advocacy in technology policy discussions, e.g. by organisations such as ACM. He noted that governments are investing billions in AI research and development for strategic advantage. They are also taking different approaches to regulating AI in different regions, with the US taking a market-driven approach

with a slow roll-out of safeguards, the EU implementing stricter regulations such as the AI Act, and China prioritising AI efficiency and large-scale deployment, focusing on AI-driven infrastructure and national security applications. This has led to global regulatory fragmentation, hindering progress in curbing AI’s environmental impacts.

Romanoff identified several challenges in AI governance as it relates to climate change. The first, echoing Strubell, is the lack of transparency around AI models and their environmental impacts. Another is the influence of corporate interests in shaping AI policy and the resulting prioritisation of profits over ethical considerations. To address these challenges, he advocated mandating transparency and developing clearer, and more proactive (rather than reactionary), AI accountability frameworks; greater involvement of independent experts in influencing policy; and development of mechanisms to ensure compliance and assess long-term policy effectiveness.

The remainder of the keynote explored strategy for influencing technology policy. He recommended pursuing both insider and outsider strategies, i.e. establishing relationships with policymakers while simultaneously adding pressure through media influence. He also stressed that materials presented to policymakers must be backed by robust data to enhance credibility and framed in an accessible manner to be consumed by busy staffers; and that they are strategically timed to coincide with policy cycles and current priorities. This often means being ready to respond to calls for comment at short notice. He ended by providing some insights into the ways ACM currently works to influence policy, and how workshop participants could leverage these mechanisms to amplify the group’s manifesto.

4.4 Discussion

A major theme of the second day was the growing trend toward regulating of AI. The European Union’s AI Act was discussed as a key example of such efforts, with participants exploring its implications for technology companies inside Europe and beyond. Of concern was the apparent pattern of big tech firms using their lobbying power to shape regulations to suit their financial interests. A noted corporate strategy was framing AI as a necessary tool for progress, even progress on environmental issues, while downplaying its various negative impacts. The lack of transparency, e.g. regarding data usage and carbon emissions, was seen as enabling overly positive framings of AI. Aware of the challenges in measuring AI’s environmental impacts, participants proposed creating standardised benchmarking tools to help with assessment of energy consumption. Participants also explored various strategies for mitigating AI’s energy consumption, including developing more efficient algorithms, optimising model architectures, and exploring alternative computing paradigms such as neuromorphic computing. There was some discussion of incentives for industry leading a transformation along these lines, given that rising costs of AI infrastructure may make large-scale models financially unsustainable in the long run, even potentially catalysing a new “AI winter”.

The other major theme of the day was an exploration of the role of computing professionals in influencing policy. It was broadly agreed that researchers and industry experts should take a more proactive approach to influencing policy. This was seen as critical given that policymakers lack the technical knowledge needed to craft effective policies. Strategies for policy engagement were explored at length. Those noted as especially relevant were capitalising on opportunities for providing expert testimony and publishing research on AI’s impacts, but also included advocating for policies that promote transparency and sustainability, collaborating with advocacy groups, and educating the public on the long-term consequences of AI-driven energy consumption.

5 Day 3: 19 March 2025**5.1 Manifesto Development and Long-Term Vision for Computing**

On the final day, participants worked toward developing a Manifesto, outlining key values and responsibilities for the field. The session consisted largely of full-group discussion, with participants debating audience, structure, content, and specific wording for the Manifesto, with all involved in co-writing a draft.

One of the key debates was whether the Manifesto should primarily address computing professionals, policymakers, or both. Some suggested a tiered approach, including different recommendations for individual developers, corporations, and policymakers, thus allowing for a more nuanced and actionable set of guidelines. The group decided for a more general approach that represented the core commitments of those in attendance, but potentially building on the Manifesto in ways that targeted different audiences.

While not explicitly referenced in the final Manifesto, the concept of “doughnut economics” [1] was used as a reference point to guide conversations around boundaries. The consensus, in line with this economic model, was that technological development must operate within environmental boundaries while also meeting basic societal needs. As for environmental boundaries, participants stressed that the Manifesto should draw attention to the issue of electronic waste, discouraging planned obsolescence and encouraging the development of repairable and reusable technologies.

Another source of debate was the question of whether computing has more potential for harm or for good when it comes to the issue of climate change. A shared concern was the tension between the profit motives for big tech and issues of public interest (e.g. sustainability), and agreement that our Manifesto ought to advocate strongly for prioritising public interest.

The environmental impacts of large-scale AI continued to be an important focal point for discussion. There was significant unease amongst participants about the unchecked expansion of AI – not only in terms of the environmental impacts of this expansion but also in terms of social impacts, e.g. on labour markets. The notion of mandating that AI developers conduct environmental impact assessments before deploying new models was explored in terms of its practical implementation, e.g. who scrutinises these assessments, who might oversee approvals, and how would this be coordinated given geopolitical tensions? The group conceded significant challenges, while agreeing that the Manifesto should nonetheless endorse transparency.

Participants also stressed the importance of using research to inform policy and ensuring that professional organisations, such as ACM, actively advocate for ethical computing policies. This became an important theme within the resulting Manifesto.

The final half hour of the workshop was dedicated to exploring next steps, including avenues for amplifying the impact of the Manifesto or otherwise shaping technology policy. A number of practical steps were identified, including various outputs that could reiterate the Manifesto for different audiences.

References

- 1 Kate Raworth. *Doughnut economics: Seven ways to think like a 21st century economist*. Chelsea Green Publishing, 2018.

Participants

- Christoph Becker
University of Toronto, CA
- Mike Berners-Lee
Lancaster University, GB
- Vinton G. Cerf
Google – Reston, US
- Andrew A. Chien
University of Chicago, US
- Benoit Combemale
University of Rennes, FR
- Vlad Coroamă
Roegen Centre for Sustainability
– Zürich, CH
- Koen De Bosschere
Ghent University, BE
- Yi Ding
Purdue University –
West Lafayette, US
- Adrian Friday
Lancaster University, GB
- Boris Gamazaychikov
Salesforce – Paris, FR
- Vicki Hanson
ACM – New York, US
- Lynda Hardman
CWI – Amsterdam, NL &
Utrecht University, NL
- Simon Hinterholzer
Borderstep Institute – Berlin, DE
- Mattias Höjer
KTH Royal Institute of
Technology – Stockholm, SE
- Lynn Kaack
Hertie School of Governance –
Berlin, DE
- Bran Knowles
Lancaster University, GB
- Lenneke Kuijer
TU Eindhoven, NL
- Anne-Laure Ligozat
CNRS – Orsay, FR
- Jan Tobias Muehlberg
Free University of Brussels, BE
- Yunmook Nah
Dankook University –
Yongin-si, KR
- Thomas Olsson
University of Tampere, FI
- Anne-Cécile Orgerie
CNRS – IRISA – Rennes, FR
- Daniel Pargman
KTH Royal Institute of
Technology – Stockholm, SE
- Birgit Penzenstadler
Chalmers University of
Technology – Göteborg, SE
- Chris Preist
University of Bristol, GB
- Tom Romanoff
ACM – New York, US
- Emma Strubell
Carnegie Mellon University –
Pittsburgh, US
- Colin Venters
University of Limerick, IE
- Junhua Zhao
The Chinese University of Hong
Kong – Shenzhen, CN



Weihrauch Complexity: Structuring the Realm of Non-Computability

Vasco Brattka^{*1}, Alberto Marcone^{*2}, Arno Pauly^{*3},
Linda Westrick^{*4}, and Kenneth Gill^{†5}

1 Universität der Bundeswehr – München, DE. vasco.brattka@cca-net.de

2 University of Udine, IT. alberto.marcone@uniud.it

3 Swansea University, GB. arno.m.pauly@gmail.com

4 Pennsylvania State University – University Park, US. 1zw299@psu.edu

5 La Salle University - Philadelphia, US. gillmathpsu@posteo.net

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 25131 “Weihrauch Complexity: Structuring the Realm of Non-Computability”. It includes an abstract of every talk given during the seminar, as well as summaries of all presentations from the sessions on open problems and new research directions. At the end is the latest version of a bibliography on Weihrauch complexity which was originally started a decade ago at the first Dagstuhl Seminar on the topic (15392).

Seminar March 23–28, 2025 – <https://www.dagstuhl.de/25131>

2012 ACM Subject Classification Mathematics of computing → Mathematical analysis; Theory of computation → Computational complexity and cryptography; Theory of computation → Logic; Theory of computation → Models of computation

Keywords and phrases combinatorial problems, computability and complexity, computable analysis, reverse and constructive mathematics, Weihrauch reducibility and related reducibilities

Digital Object Identifier 10.4230/DagRep.15.3.125

1 Executive Summary

Vasco Brattka (Universität der Bundeswehr – München, DE)

Alberto Marcone (University of Udine, IT)

Arno Pauly (Swansea University, GB)

License  Creative Commons BY 4.0 International license
© Vasco Brattka, Alberto Marcone, and Arno Pauly

This Dagstuhl Seminar is dedicated to the investigation of two active areas of research, one in theoretical computer science, the other in mathematical logic. These are computable analysis on the one hand, and reverse mathematics and applied computability theory on the other. That there is a deep connection between these areas was first suggested by Gherardi and Marcone (2008) and later independently by Dorais, Dzhafarov, Hirst, Milet, and Shafer (2016) and Hirschfeldt and Jockusch (2016). The past decade has seen this connection blossom into a rich and productive area of research, with by now many papers and several Ph.D. theses dedicated to it. Results in this area fall into two intertwined groups: Some clarify the structure of the degrees of non-computability; some further our understanding of the precise nature of non-computability of particular computational tasks of interest. Grasping

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Weihrauch Complexity: Structuring the Realm of Non-Computability, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 125–158

Editors: Vasco Brattka, Alberto Marcone, Arno Pauly, Linda Westrick, and Kenneth Gill



Dagstuhl Reports

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

the nature of non-computability is a profound goal mirroring the quest to understand the nature of computation. Knowing the degree of non-computability of a computational task brings with it answers as to whether weaker or approximate versions of it might be solvable. This interdisciplinary development was fostered not least by the two precursor Dagstuhl Seminars on this topic.¹

The current seminar explored recent trends and results, open questions, and new directions of this fascinating field of research that has become known as Weihrauch complexity. The main part of each day was taken up by regular talks, with extra time set aside for two sessions devoted to open questions and new research directions, as well as plenty of opportunities for less structured socialization and collaboration. Although the ratio of number of talks to number of open questions (as represented in the sessions and this report) was nominally greater than in the previous seminar from 2018, a number of the talks themselves focused heavily on enumerating open questions and outlining future work, and indeed the field has only widened in the intervening years. To mention just a few highlights: investigations of the Weihrauch complexity of reverse-mathematical principles have continued to spur new developments, and this was reflected accordingly in many of the talks here, representing the study of “new” principles as well as new light still being shed on old ones. Important progress has also been made in our understanding of the properties of the Weihrauch lattice itself, such as the existence of uncountable chains and antichains and the density of the Weihrauch degrees above the identity map. Operators on Weihrauch degrees were a prominent theme during the seminar, featuring in the sessions on open problems and new research directions as well as being central to several talks. A few talks concerned recent work to place Weihrauch reducibility in context as an instance of a more general sort of object in category or topos theory.

Last but not least, underscoring the increasingly interdisciplinary interest in this subject, a well-attended joint evening session was spontaneously planned with the concurrent Dagstuhl Seminar² in which a speaker from each seminar gave an expository talk aimed at the other’s participants: Kevin Schewior spoke about approximate sampling algorithms for stochastic function evaluation, and Arno Pauly about the non-computability of finding Nash equilibria.

This report includes the abstracts of all talks and other presentations given during the seminar (except for the joint talks), along with the most recent version of a bibliography on Weihrauch complexity which was begun during the first Dagstuhl Seminar on the topic in 2015. Altogether, this report reflects the high degree of productivity of our seminar, and we would like to use this opportunity to thank all participants for their valuable contributions and the Dagstuhl staff for their excellent support!

¹ Seminars 15392 and 18361; see <https://doi.org/10.4230/DagRep.5.9.77> and <https://doi.org/10.4230/DagRep.8.9.1>.

² Approximation Algorithms for Stochastic Optimization (25132; see <https://www.dagstuhl.de/25132>).

2 Table of Contents

Executive Summary

<i>Vasco Brattka, Alberto Marcone, and Arno Pauly</i>	125
---	-----

Overview of Talks


A category-theoretic account of generalized Weihrauch degrees <i>Andrej Bauer</i>	129
Survey on Weihrauch Complexity: Scaffolding, Operators, Dichotomies <i>Vasco Brattka</i>	130
Effective Reducibility Notions with Transfinite Machine Models <i>Merlin Carl</i>	130
A well-quasi-order for continuous functions <i>Raphaël Carroy</i>	131
The category of quasi-Polish spaces as a represented space <i>Matthew de Brecht</i>	131
The tree pigeonhole principle in the Weihrauch degrees <i>Damir D. Dzhafarov</i>	131
No dilator characterizes Ramsey's theorem for pairs <i>Anton Freund</i>	132
Formalization of Weihrauch reducibility in second-order arithmetic between existence statements <i>Makoto Fujiwara</i>	132
Reverse Math of Regular Countable Second Countable Spaces <i>Giorgio G. Genovesi</i>	132
Pigeonhole principles for countable structures <i>Kenneth Gill</i>	133
Forests Describing Topological Weihrauch Degrees of Functions with Discrete Range <i>Peter Hertling</i>	133
Basis theorems: Reverse mathematics and Weihrauch reductions <i>Jeffrey L. Hirst</i>	133
Generalized Weihrauch reducibility <i>Takayuki Kihara</i>	134
Recent applications of proof mining to splitting algorithms <i>Ulrich Kohlenbach</i>	134
Better quasi-orders on labelled trees <i>Davide Manca</i>	135
The Galvin-Prikry theorem in the Weihrauch lattice <i>Alberto Marcone</i>	136
Indices, Computable Discontinuities and the Recursion Theorem <i>Daniel Mourad</i>	136
The equational theory of the Weihrauch degrees <i>Arno Pauly</i>	136

Weihrauch problems are containers	
The equational theory of slightly extended Weihrauch degrees with composition	
<i>Cécilia Pradic</i>	137
Principal Spaces	
<i>Matthias Schröder</i>	138
Old directions in degree theory	
<i>Mariya I. Soskova</i>	138
Weihrauch degrees without roots	
<i>Patrick Uftring</i>	139
An overview on the structure of the Weihrauch degrees	
<i>Manlio Valenti</i>	139
On the hierarchy above ATR in Weihrauch degrees and reverse mathematics	
<i>Keita Yokoyama</i>	140
Open problems	
What to do about all the other Weihrauch lattices?	
<i>Andrej Bauer</i>	141
Interior operators in the Weihrauch lattice	
<i>Jun Le Goh</i>	141
Question on the strength of the infinite loop closure	
<i>Takayuki Kihara</i>	142
Strong Weihrauch compositional product	
<i>Alberto Marcone</i>	142
A question about the the uniform content of index sets	
<i>Daniel Mourad</i>	143
On residual operators	
<i>Manlio Valenti</i>	143
Preservation results for well-quasiorders in the Weihrauch lattice	
<i>Arno Pauly</i>	143
A problem on the preservation of well-foundedness	
<i>Keita Yokoyama</i>	144
Bibliography on Weihrauch Complexity	145
Participants	158

3 Overview of Talks

3.1 A category-theoretic account of generalized Weihrauch degrees

Andrej Bauer (University of Ljubljana & Institute for Mathematics, Physics, and Mechanics – Ljubljana, SI)

License  Creative Commons BY 4.0 International license

© Andrej Bauer

Joint work of Danel Ahman, Andrej Bauer

In joint work with Danel Ahman [1] we developed and investigated a general theory of representations of second-order functionals, based on a notion of a right comodule for a monad on the category of containers. The theory can be used to give a type-theoretic account of instance reducibility [2] and, through their realizability interpretation, generalized Weihrauch degrees.

A *container* $A \triangleleft P$ is given by a type A and a type family $P: A \rightarrow \mathbf{Type}$. A morphism $f \triangleleft g: (A \triangleleft P) \rightarrow (B \triangleleft Q)$ is given by a map $f: A \rightarrow B$ and a map $g: \prod_{a:A} Q(f a) \rightarrow P a$. Containers have been studied extensively in type theory and functional programming.

A special case is a *propositional container* $A \triangleleft^P P$, which is given by a type A and a predicate $P: A \rightarrow \mathbf{Prop}$. A morphism of propositional containers $f: (A \triangleleft^P P) \rightarrow (B \triangleleft^P Q)$ is a map $f: A \rightarrow B$ such that

$$\forall a:A. Q(f a) \Rightarrow P a.$$

In terms of instance degrees, such a map f is a *functional instance reduction*. The majority of instance reductions seen in mathematical practice (both classical and constructive) are of this kind.

The notion of *instance reducibility*, which states that $A \triangleleft^P P$ is reducible to $B \triangleleft^P Q$ when

$$\forall a:A. \exists b:B. (Q b \Rightarrow P a),$$

can be accounted for in terms of the general theory of representations of second-order functionals. Namely, it corresponds to the preorder reflection of the Kleisli category for the inhabited powerset monad on the category of propositional containers [1, Prop. 8.5].


These observations open up the possibility for generalizations of Weihrauch degrees, and application of type-theoretic and category-theoretic techniques to the topic. They also show how Weihrauch reducibility is situated in the wider context of representations of second-order functionals.

References

- 1 Danel Ahman and Andrej Bauer. Comodule representations of second-order functionals. *Journal of Logical and Algebraic Methods in Programming*, 146:101071, 2025.
- 2 Andrej Bauer. Instance reducibility and Weihrauch degrees. *Logical Methods in Computer Science*, 18(3), 2022.

3.2 Survey on Weihrauch Complexity: Scaffolding, Operators, Dichotomies

Vasco Brattka (*Universität der Bundeswehr – München, DE*)

License  Creative Commons BY 4.0 International license
© Vasco Brattka


Main reference Vasco Brattka: “The discontinuity Problem”, *J. Symb. Log.*, Vol. 88(3), pp. 1191 – 1212, 2023.

URL <https://doi.org/10.1017/JSL.2021.106>

We give a survey on basic problems, operators and dichotomies in Weihrauch complexity. In particular, we describe how LPO and LLPO can be used together with operators such as jump, parallelization, diamond, first-order part, and deterministic part to generate a whole class of very basic and important Weihrauch degrees. We describe how these degrees give natural classes of computable problems and how they match with systems in reverse mathematics. We also briefly discuss the role of closure and interior operators. Finally, we show how some of these degrees also lead to dichotomies for continuous problems with respect to continuous Weihrauch reducibility and different codomains. We close with a brief demonstration of how such dichotomies can be de-uniformized with the help of parallelization in order to obtain dichotomies for computable reducibility.

3.3 Effective Reducibility Notions with Transfinite Machine Models

Merlin Carl (*Europa-Universität – Flensburg, DE*)

License  Creative Commons BY 4.0 International license
© Merlin Carl

Joint work of Merlin Carl, Lorenzo Galeotti, Robert Passmann

In recent years, various notions of effectivity and effective reducibility, such as Weihrauch reducibility and realizability, have been adapted to work on sets of arbitrary size by replacing Turing computability with computability by transfinite machine models, such as Koepke’s Ordinal Turing Machines. In this talk, we will give an overview of this area with some of the central results, in particular concerning the mutual effective reducibility between the axioms and axiom schemes of ZFC usually regarded as non-constructive or impredicative, such as the power set axiom, the axiom of choice, and the schemes of separation and replacement.

References

- 1 Merlin Carl. Effectivity and reducibility with ordinal Turing machines. *Computability* 10(4) (2021), 289-304. doi: doi:10.3233/COM-210307.
- 2 Robert Passmann. The first-order logic of CZF is intuitionistic first-order logic. *Journal of Symbolic Logic* 89(1) (2022), 308-330. doi:10.1017/jsl.2022.51.
- 3 Merlin Carl, Lorenzo Galeotti, and Robert Passmann. Realisability for infinitary intuitionistic set theory. *Annals of Pure and Applied Logic* 174(6):103259 (2023). doi: doi:10.1016/j.apal.2023.103259.
- 4 Merlin Carl. Full generalized effective reducibility. Submitted (2025). arXiv: 2411.19386.

3.4 A well-quasi-order for continuous functions

Raphaël Carroy (University of Torino, IT)

License © Creative Commons BY 4.0 International license
 © Raphaël Carroy
Joint work of Raphaël Carroy, Yann Pequignot
Main reference Raphaël Carroy, Yann Pequignot: “A well-quasi-order for continuous functions”, CoRR, Vol. abs/2410.13150, 2024.
URL <https://arxiv.org/abs/2410.13150>

We prove that continuous reducibility – or topological strong Weihrauch reducibility – on continuous functions from a 0-dimensional analytic domain to a separable metrizable space is a well-quasi-order, or more precisely, a better-quasi-order. To do so, we introduce and describe the class of scattered continuous functions with a 0-dimensional domain.

3.5 The category of quasi-Polish spaces as a represented space

Matthew de Brecht (Kyoto University, JP)

License © Creative Commons BY 4.0 International license
 © Matthew de Brecht
Main reference Matthew de Brecht: “The category of quasi-Polish spaces as a represented space”, 2021
URL <https://www.mathsoc.jp/section/topology/topsymp/2021/ts2021Brecht.pdf>

We construct the category of quasi-Polish spaces as a represented space, which allows us to investigate the computability aspects of some category theoretical constructions, such as functors and limits, within the framework of Type-Two Theory of Effectivity. As an example, we demonstrate the computability of the lower, upper, double, and valuation powerspace endofunctors on the category of quasi-Polish spaces. (This talk was originally presented at the 68th Topology Seminar, August 2021: <https://www.mathsoc.jp/section/topology/topsymp.html>)

3.6 The tree pigeonhole principle in the Weihrauch degrees

Damir D. Dzhabarov (University of Connecticut – Storrs, US)

License © Creative Commons BY 4.0 International license
 © Damir D. Dzhabarov
Joint work of Damir D. Dzhabarov, Reed Solomon, Manlio Valenti
Main reference Damir D. Dzhabarov, Reed Solomon, Manlio Valenti: “The Tree Pigeonhole Principle In The Weihrauch Degrees”, The Journal of Symbolic Logic, p. 1–23, 2025.
URL <https://doi.org/10.1017/jsl.2025.11>

I will discuss recent work studying versions of the tree pigeonhole principle, TT^1 , in the context of Weihrauch-style computable analysis. The principle has previously been the subject of extensive research in reverse mathematics, an outstanding question of which investigation is whether TT^1 is Π_1^1 -conservative over the ordinary pigeonhole principle, RT^1 . Using the recently introduced notion of the first-order part of an instance-solution problem, we formulate the analogue of this question for Weihrauch reducibility, and give an affirmative answer. In combination with other results, we use this to show that unlike RT^1 , the problem TT^1 is not Weihrauch equivalent to any first-order problem. Our proofs develop new combinatorial machinery for constructing and understanding solutions to instances of TT^1 . This is joint work with Reed Solomon and Manlio Valenti.

3.7 No dilator characterizes Ramsey's theorem for pairs

Anton Freund (Universität Würzburg, DE)

License © Creative Commons BY 4.0 International license
© Anton Freund

Main reference Anton Freund: “Dilators and the reverse mathematics zoo”, Journal of Mathematical Logic, p. 2550010, 0.

URL <https://doi.org/10.1142/S0219061325500102>

Dilators are particularly uniform transformations of well-orders. Above ACA_0 , every Π_2^1 statement corresponds to a dilator, by a classical result of Girard. In contrast, we show that no dilator corresponds to Ramsey's theorem for pairs and two colours (and the same is true for many other principles from the reverse mathematics zoo). Our proof involves a new principle of slow transfinite Π_2^0 -induction, which admits a recursive counterexample but seems to lie below the Turing jump (though the latter is an open conjecture).

3.8 Formalization of Weihrauch reducibility in second-order arithmetic between existence statements

Makoto Fujiwara (Tokyo University of Science, JP)

License © Creative Commons BY 4.0 International license
© Makoto Fujiwara

Joint work of Makoto Fujiwara, Yudai Suzuki

Main reference Makoto Fujiwara and Yudai Suzuki. Formalization of Weihrauch reducibility in second-order arithmetic between existence statements. Accepted to *Computability*.

We formalize the notion of Weihrauch reducibility between existence statements in terms of second-order arithmetic [1], which is a standard framework of reverse mathematics. This formalization enables us to determine the strength of verification theories needed for Weihrauch reducibility between existence statements. As an example, we show that for any second-order theory T which is an extension of RCA_0 , weak König's lemma with a uniqueness hypothesis is Weihrauch reducible to the identity map in T if and only if T proves weak König's lemma. This is joint work with Yudai Suzuki.

References

- 1 S. G. Simpson. *Subsystems of Second Order Arithmetic*, 2nd ed. Cambridge University Press, 2009.

3.9 Reverse Math of Regular Countable Second Countable Spaces

Giorgio G. Genovesi (University of Leeds, GB)

License © Creative Commons BY 4.0 International license
© Giorgio G. Genovesi

Main reference Giorgio G. Genovesi: “Reverse mathematics of regular countable second countable spaces”, CoRR, Vol. abs/2410/22227, 2024.

URL <https://arxiv.org/abs/2410.22227>

One approach to studying theorems of general topology in second order arithmetic is to consider the countable second countable spaces, or CSC spaces. There are several classical theorems in general topology which characterize the regular CSC spaces. We go over the strength of some of these theorems in relation to the Big Five systems of second order arithmetic. We also outline how ATR_0 proves that regular Hausdorff CSC spaces are a well-quasi-order under embedding.

3.10 Pigeonhole principles for countable structures

Kenneth Gill (La Salle University – Philadelphia, US)

License © Creative Commons BY 4.0 International license
© Kenneth Gill

Joint work of Kenneth Gill, Damir Dzhafarov, Reed Solomon

Main reference Kenneth Gill: “Indivisibility and uniform computational strength”, *Log. Methods Comput. Sci.*, Vol. 21(2), 2025.

URL [https://doi.org/10.46298/LMCS-21\(2:22\)2025](https://doi.org/10.46298/LMCS-21(2:22)2025)

A countable structure is said to be indivisible if for every presentation and every bounded coloring of the presentation, there is a monochromatic substructure isomorphic to the whole structure. Examples include the natural numbers, Rado and Henson graphs, and nonscattered linear orders. This notion naturally gives rise to an instance-solution problem which outputs such a substructure given a presentation and coloring. We discuss the Weihrauch degrees of these problems in general and for some specific structures, surveying what is known and highlighting current investigations. This is (in part) joint ongoing work with Damir Dzhafarov and Reed Solomon.

3.11 Forests Describing Topological Weihrauch Degrees of Functions with Discrete Range

Peter Hertling (Universität der Bundeswehr – München, DE)

License © Creative Commons BY 4.0 International license
© Peter Hertling

We show that a certain initial segment of the degree structure of functions with discrete, possibly infinite, range under continuous Weihrauch reducibility is isomorphic to a hierarchy of labeled forests with respect to a suitable reducibility relation. We also present an explicit calculation of the degree structure of the topological Weihrauch degrees of functions of level of discontinuity at most 4.

References

- 1 Peter Hertling. *Unstetigkeitsgrade von Funktionen in der effektiven Analysis*. PhD thesis. Fachbereich Informatik, FernUniversität Hagen, 1996.
- 2 Peter Hertling. Forests describing Wadge degrees and topological Weihrauch degrees of certain classes of functions and relations. *Computability* 9 (2020), 249–307. doi:10.3233/COM-190255.

3.12 Basis theorems: Reverse mathematics and Weihrauch reductions

Jeffrey L. Hirst (Appalachian State University – Boone, US)

License © Creative Commons BY 4.0 International license
© Jeffrey L. Hirst

Joint work of Caleb Davis, Silva Keohulian, Brody Miller, and Jessica Ross, and separately, with Carl Mummert

Main reference Caleb Davis, Jeffrey Hirst, Silva Keohulian, Brody Miller, Jessica Ross: “Reverse mathematics of a pigeonhole basis theorem”. To appear in *Computability* (2025).

URL <https://hirstjl.github.io/bib/pdf/cb111024LargePrint.pdf>

There are a number of basis theorems that are equivalent to Σ_2^0 induction in the reverse mathematics framework. For example, the color basis theorem and the basis theorem for finite dimensional e-matroids are provably equivalent. They are not Weihrauch equivalent. See

[1] and [2]. Insights from Weihrauch analysis can motivate interesting reformulations of the reverse mathematics results. Other examples of statements equivalent to Σ_2^0 induction with various Weihrauch strengths can be found in the recent work of Pauly, Pradic, and Soldà [3].

References

- 1 Caleb Davis, Jeffrey Hirst, Silva Keohulian, Brody Miller, and Jessica Ross. Reverse mathematics of a pigeonhole basis theorem. To appear in *Computability* (2025).
- 2 Jeffrey Hirst and Carl Mummert. Reverse mathematics of matroids. In Adam Day et al. (editors), *Computability and Complexity*, Lecture Notes in Computer Science vol. 10010, 143–159. Cham: Springer, 2017. doi: doi:10.1007/978-3-319-50062-1_12.
- 3 Arno Pauly, Cécilia Pradic, and Giovanni Soldà. On the Weihrauch degree of the additive Ramsey theorem. *Computability* 13(3-4) (2024), 459–483. doi:10.3233/COM-230437.

3.13 Generalized Weihrauch reducibility

Takayuki Kihara (Nagoya University, JP)

License  Creative Commons BY 4.0 International license
© Takayuki Kihara

I will give an overview of generalized Weihrauch reducibility from the perspectives of computability theory, reverse mathematics, and realizability topos theory, with concrete examples and applications. This talk will cover the following topics: compositional product, reduction game, Weihrauch-oracle realizability, constructive reverse mathematics, realizability topos, Lawvere-Tierney topology, subtopos, and extended generalized Weihrauch reducibility.

References

- 1 Takayuki Kihara. Lawvere-Tierney topologies for computability theorists. *Trans. Amer. Math. Soc. Series B* 10 (2023), 48–85.
- 2 Takayuki Kihara. Rethinking the notion of oracle: A prequel to Lawvere-Tierney topologies for computability theorists. Preprint (2022). arXiv: 2202.00188.

3.14 Recent applications of proof mining to splitting algorithms

Ulrich Kohlenbach (TU Darmstadt, DE)

License  Creative Commons BY 4.0 International license
© Ulrich Kohlenbach

Splitting methods play a central role in nonsmooth optimization in the design of algorithms for the computation of zeros of maximally monotone set-valued operators in Hilbert spaces which can be written as the sum $A + B$ of two such operators. The main point here is to avoid the use of the resolvent of $A + B$ and to involve only the individual resolvents J_A, J_B of A and B respectively, which may be easier to compute (note that to compute the resolvents of an operator amounts to solving in inverse problem). The most well-studied such algorithms are (i) Tseng’s Splitting Algorithm, (ii) the Forward-Backward Splitting Algorithm, (iii) the Douglas-Rachford Splitting Algorithm and, as the limiting case of (iii), (iv) the Peaceman-Rachford Algorithm (see e.g. [1]).


In [7] and [5], the logic-based proof mining methodology ([2]) is used to extract rates of convergence in certain quantitative forms of uniform monotonicity which give rise to moduli of uniqueness and hence moduli of regularity in the sense of [6]. The existence of such moduli has been studied in terms of reverse mathematics and Weihrauch complexity in [3] and in terms of intuitionistic reverse mathematics recently in [4].

References

- 1 H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. CMS Books in Mathematics. Cham: Springer, 2017.
- 2 Ulrich Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Heidelberg-Berlin: Springer, 2008.
- 3 Ulrich Kohlenbach. On the reverse mathematics and Weihrauch complexity of moduli of regularity and uniqueness. *Computability* 8 (2019), 377-387.
- 4 Ulrich Kohlenbach. On the computational content of moduli of regularity and their logical strength. Submitted.
- 5 Ulrich Kohlenbach and Nicholas Pischke. Quantitative results for a Tseng-type primal-dual method for composite monotone inclusions. submitted.
- 6 Ulrich Kohlenbach, Genaro Lopéz-Acedo, and Adriana Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics* 232 (2019), 261-297.
- 7 Jacqueline Treusch and Ulrich Kohlenbach. Rates of convergence for splitting algorithms. To appear in *Israel Journal of Mathematics*.

3.15 Better quasi-orders on labelled trees

Davide Manca (Universität Würzburg, DE)

License  Creative Commons BY 4.0 International license
© Davide Manca

Main reference Davide Manca. *At the limits of predicativity: the reverse mathematics of ordering relations*. Ph.D. dissertation. To appear (2025).

Kruskal's theorem states that finite trees with labels in a well quasi-order (wqo) form a wqo under infima-preserving embeddings. Nash-Williams proved a version of that theorem for infinite trees, which relies on the stronger notion of better quasi-order [3] (see [1] for the result for labelled trees). That version has not yet been analyzed in an appropriate context, such as that of reverse mathematics. On the other hand, a number of weaker results about the structure of trees with labels in a better quasi-order have been studied, often in relation to open problems such as the strength of Fraïssé's conjecture [2]. We review the currently available results from the point of view of reverse mathematics and discuss some new ones, as well as some ideas for future research.

References

- 1 Richard Laver. On Fraïssé's order type conjecture. *Annals of Mathematics* 93(1) (1971).
- 2 Antonio Montalbán. Fraïssé's conjecture in Π_1^1 -comprehension. *Journal of Mathematical Logic* 17(2) (2017).
- 3 C. St. J. A. Nash-Williams. On well-quasi-ordering infinite trees. *Mathematical Proceedings of the Cambridge Philosophical Society* 61(3) (1965).

3.16 The Galvin-Prikry theorem in the Weihrauch lattice

Alberto Marcone (University of Udine, IT)

License  Creative Commons BY 4.0 International license
© Alberto Marcone

Joint work of Alberto Marcone, Gian Marco Osso

Main reference Alberto Marcone, Gian Marco Osso: “The Galvin-Prikry Theorem in the Weihrauch lattice”, CoRR, Vol. abs/2410/06928, 2024.

URL <https://doi.org/10.48550/arXiv.2410.06928>

We address the classification of different fragments of the Galvin-Prikry theorem in terms of their uniform computational content. We show that functions related to the Galvin-Prikry theorem for Borel sets of rank n are strictly between the $(n + 1)$ th and n th iterate of the hyperjump operator. To this end we establish the following result: a Turing jump ideal containing homogeneous sets for all $\Delta_{n+1}^0(X)$ sets must also contain the n th hyperjump of X . Similar results also hold for Borel sets of transfinite rank. These findings yield a partial refinement of previous results in the reverse mathematics of the Galvin-Prikry theorem. Moreover, in combination with previous results of Marcone and Valenti, they allow us to obtain a fairly complete picture of the Weihrauch degrees of the functions studied.

3.17 Indices, Computable Discontinuities and the Recursion Theorem

Daniel Mourad (Nanjing University, CN)

License  Creative Commons BY 4.0 International license
© Daniel Mourad

Consider a problem P with at least one computable instance. Let P' be the problem whose instances are indices n such that the n th computable partial function ϕ_n is an instance of P and such that $P'(n) = P(\phi_n)$. We investigate the relationship between discontinuity of P and computability of P' . We show that if P has a computable discontinuity (which we will define) then P' is not computable. This fact generalizes many applications of the recursion theorem, such as showing that P' is not computable when $P = \text{WKL}$ or $P = \text{RT}_1^1$. We also pose some questions about how having the index of a solution rather than the set that the index encodes influences Weihrauch reductions.

3.18 The equational theory of the Weihrauch degrees

Arno Pauly (Swansea University, GB)

License  Creative Commons BY 4.0 International license
© Arno Pauly

Joint work of Arno Pauly, Eike Neumann, Cécilia Pradic

Main reference Eike Neumann, Arno Pauly, Cécilia Pradic: “The equational theory of the Weihrauch lattice with multiplication”, CoRR, Vol. abs/2403.13975, 2024.

URL <https://doi.org/10.48550/ARXIV.2403.13975>

The algebraic structure of the Weihrauch degrees has long been a subject of study. It is linked to the “inherent logic of computability”. Identifying the Weihrauch degrees as an instance of a previously studied class of structures, in particular one with a logical flavour, could significantly advance our understanding.

Here we study the equational theory of the Weihrauch lattice with multiplication, meaning the collection of equations between terms built from variables, the lattice operations \sqcup and \sqcap , the product \times , and the finite parallelization $(\cdot)^*$ which are true however we substitute Weihrauch degrees for the variables. We provide a combinatorial description of these in terms of a reducibility between finite graphs, and moreover, show that deciding which equations are true in this sense is complete for the third level of the polynomial hierarchy. Pradic has similarly studied the equational structure of the Weihrauch lattice with composition.

References

- 1 Vasco Brattka and Arno Pauly. On the algebraic structure of Weihrauch degrees. *Logical Methods in Computer Science* 14(4) (2018). doi:10.23638/LMCS-14(4:4)2018.
- 2 Kojiro Higuchi and Arno Pauly. The degree structure of Weihrauch-reducibility. *Logical Methods in Computer Science* 9(2) (2011). doi:10.2168/LMCS-9(2:2)2013.
- 3 Cécilia Pradic. The equational theory of the Weihrauch lattice with (iterated) composition. Preprint (2024). arXiv: 2408.14999.

3.19 Weihrauch problems are containers The equational theory of slightly extended Weihrauch degrees with composition

Cécilia Pradic (Swansea University, GB)

License © Creative Commons BY 4.0 International license
© Cécilia Pradic

Joint work of Cécilia Pradic, Ian Price

Main reference Cécilia Pradic, Ian Price: “Weihrauch problems as containers”, CoRR, Vol. abs/2501.17250, 2025.
URL <https://doi.org/10.48550/ARXIV.2501.17250>

I’ll explain that Weihrauch problems can be regarded as containers over the category of subspaces of Baire spaces and computable maps and that Weihrauch reductions correspond exactly to container morphisms. Up to restricting to those containers that do not allow a problem not to answer a question, we get a clean equivalence. We can make similar observations and elaborations regarding extended/generalized/strong Weihrauch reducibility.


In the second part of the talk, I will discuss the equational theory of the Weihrauch lattice equipped with (iterated) composition. Terms in this theory can be translated to alternating automata, and reductions regarded as a somewhat weird kind of simulation. This leads to decidability and a complete axiomatization that includes a generalization of a result of Linda Westrick.

References

- 1 Cécilia Pradic. The equational theory of the Weihrauch lattice with (iterated) composition. Preprint (2025). arXiv: 2408.14999.
- 2 Cécilia Pradic and Ian Price. Weihrauch problems as containers. Preprint (2025). arXiv: 2501.17250.
- 3 Linda Westrick. A note on the diamond operator. *Computability* 10 (2023).

3.20 Principal Spaces

Matthias Schröder (TU Darmstadt, DE)

License  Creative Commons BY 4.0 International license
© Matthias Schröder

We introduce the class of principal topological spaces. Principal spaces have some bizarre properties which might be useful in Computability Theory. For example, they admit some automatic continuity properties.


Under the Axiom of Choice, principal spaces are very rare: no infinite Hausdorff space is principal under AC. By contrast, in Shelah's model of set theory and thus under the Axiom of Determinacy a big class of topological spaces relevant to Computable Analysis turn out to be principal, including all computable metric spaces and, more generally, all functionally Hausdorff qcb-spaces.

References

- 1 Eric Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, 1997,
- 2 Matthias Schröder. Admissibly represented spaces and qcb-spaces. In Vasco Brattka and Peter Hertling, editors, *Handbook of Computability and Complexity in Analysis*, 305-346. Cham: Springer, 2021. doi:10.1007/978-3-030-59234-9_9.

3.21 Old directions in degree theory

Mariya I. Soskova (University of Wisconsin – Madison, US)

License  Creative Commons BY 4.0 International license
© Mariya I. Soskova

I was asked to present a brief overview of aspects of degree theory that have been studied throughout the years. The intention was that researchers interested in the Weihrauch degrees may use this as a source for questions that they may pursue. I focused on the following aspects:

I discussed the complexity of the theory of Turing degrees and its fragments when restricted to statements of limited quantifier complexity. I proposed the following questions about the Weihrauch lattice: How complicated are the fragments of the theory of \mathcal{D}_W ? At what quantifier level does decidability break down? Are there upper cones of Weihrauch degrees with a decidable/less complicated theory? Specifically, what about the cone above the degree of id?

I discussed the larger structure of the enumeration degrees and ways in which studying the Turing degrees within this larger context has been illuminating. I introduced the enumeration-Weihrauch degrees and suggested the following questions: Can enumeration Weihrauch reducibility be defined entirely in terms of Weihrauch reducibility à la Selman's theorem? How do other operators on the Weihrauch degrees live inside the \leq_{eW} -degrees? Are the Weihrauch degrees definable in the \leq_{eW} -degrees? What is the relationship between problems represented in the \leq_{eW} -degrees and their total counterparts coming from the Weihrauch degrees?

I discussed local substructures such as the c.e. Turing degrees and ways in which working with them has expanded our toolbox (the priority method). I asked what local structures of the Weihrauch degrees arise naturally or determine the global structure.

Finally I discussed ways in which effective mathematics influences our view of degree structures and helps solve purely structural problems within and asked whether a similar phenomenon can be observed in the Weihrauch lattice.

3.22 Weihrauch degrees without roots

Patrick Uftring (Universität Würzburg, DE)

License © Creative Commons BY 4.0 International license
© Patrick Uftring

Main reference Patrick Uftring: “Weihrauch degrees without roots”, CoRR, Vol. abs/2308.01422, 2023.

URL <https://doi.org/10.48550/ARXIV.2308.01422>

We answer the following question by Arno Pauly ([1, Open Question 12]): “Is there a square-root operator on the Weihrauch degrees?” In fact, we show that there are uncountably many pairwise incomparable Weihrauch degrees without any roots. We also prove that the omniscience principles LPO and LLPO do not have roots.

References

- 1 Arno Pauly. An update on Weihrauch complexity, and some open questions. Preprint (2020). arXiv:2008.11168.

3.23 An overview on the structure of the Weihrauch degrees

Manlio Valenti (Swansea University, GB)

License © Creative Commons BY 4.0 International license
© Manlio Valenti

In this talk, I will provide an overview of what is currently known about the structural properties of the Weihrauch degrees, including some of the more recent results about the existence and properties of chains, antichains, intervals and minimal covers, strong minimal covers, minimal pairs, and embeddings. I will also highlight some open questions and research directions.


References

- 1 Uri Andrews, Steffen Lempp, Alberto Marcone, Joseph S. Miller, and Manlio Valenti. A jump operator on the Weihrauch degrees. To appear in *Computability*. arXiv: 2402.13163.
- 2 Vasco Brattka and Guido Gherardi. Weihrauch degrees, omniscience principles and weak computability. *The Journal of Symbolic Logic* 76(1) (2011), 143–176. doi:10.2178/jsl/1294170993. arXiv: 0905.4679.
- 3 Vasco Brattka, Guido Gherardi, and Arno Pauly. Weihrauch complexity in computable analysis. In Vasco Brattka and Peter Hertling (editors), *Handbook of Computability and Complexity in Analysis*, 367 – 417. Springer International Publishing, 2021. doi:10.1007/978-3-030-59234-9_11. arXiv: 1707.03202.
- 4 Elena Z. Dymont. On some properties of the Medvedev lattice. *Mathematics of the USSR-Sbornik* 30(3) (1976), 321 – 340. doi:10.1070/SM1976v030n03ABEH002277.
- 5 Kojiro Higuchi and Arno Pauly. The degree structure of Weihrauch reducibility. *Logical Methods in Computer Science* 9(2:02) (2013), 1 – 17. doi:10.2168/LMCS-9(2:02)2013. arXiv: 1101.0112.

- 6 Steffen Lempp, Alberto Marcone, and Manlio Valenti. Chains and antichains in the Weihrauch lattice. Preprint (2024). arXiv: 2411.07792.
- 7 Steffen Lempp, Joseph S. Miller, Arno Pauly, Mariya I. Soskova, and Manlio Valenti. Minimal covers in the Weihrauch degrees. *Proceedings of the American Mathematical Society* 152(11) (2024), 4893 – 4901. doi:10.1090/proc/16952. arXiv: 2311.12676.
- 8 Arno Pauly. On the (semi)lattices induced by continuous reducibilities. *Mathematical Logic Quarterly* 56(5) (2010), 488 – 502. doi:10.1002/malq.200910104. arXiv: 10.1002/malq.200910104.
- 9 Paul Shafer. *On the complexity of mathematical problems: Medvedev degrees and reverse mathematics*. Ph.D. thesis. Cornell University, 2011.
- 10 Andrea Sorbi. The Medvedev lattice of degrees of difficulty. In *Computability, Enumerability, Unsolvability*, London Math. Soc. Lecture Note Series vol. 224, 289 – 312. New York: Cambridge University Press, 1996. doi:10.1017/CBO9780511629167.015.
- 11 Sebastiaan A. Terwijn. On the Structure of the Medvedev Lattice. *The Journal of Symbolic Logic* 73(2) (2008), 543 – 558. doi:10.2178/jsl/1208359059. Available at <http://www.jstor.org/stable/27588647>.

3.24 On the hierarchy above ATR in Weihrauch degrees and reverse mathematics

Keita Yokoyama (Tohoku University, JP)

License  Creative Commons BY 4.0 International license
© Keita Yokoyama

Joint work of Keita Yokoyama, Yudai Suzuki

In the study of reverse mathematics, the gap between ATR_0 and $\Pi_1^1\text{-CA}_0$ is rather large, with many mathematical theorems falling in between. We focus on those theorems which are described by Π_2^1 -sentences and examine the hierarchy above arithmetical transfinite recursion in the context of Weihrauch degrees and reverse mathematics. This is joint work with Yudai Suzuki.


References

- 1 Yudai Suzuki and Keita Yokoyama. Searching problems above arithmetical transfinite recursion. *Annals of Pure and Applied Logic* 175 (2024), 31pp. doi:10.1016/j.apal.2024.103488.
- 2 Yudai Suzuki and Keita Yokoyama. On the Π_2^1 consequences of $\Pi_1^1\text{-CA}_0$. Preprint (2024). arXiv: 2402.07136.

4 Open problems

4.1 What to do about all the other Weihrauch lattices?

Andrej Bauer (University of Ljubljana & Institute for Mathematics, Physics, and Mechanics – Ljubljana, SI)

License  Creative Commons BY 4.0 International license
© Andrej Bauer

The lattice of (generalized) Weihrauch degrees arises as the lattice of instance degrees [1], interpreted in the Kleene-Vesley topos $\mathbf{RT}(\mathbb{N}^{\mathbb{N}}, (\mathbb{N}^{\mathbb{N}})_{\text{eff}})$, based on the function realizability model [2]. However, the instance degrees may be calculated in any topos to give many new variants of Weihrauch reduction. For example, in the relative realizability topos $\mathbf{RT}(\mathcal{P}\omega, (\mathcal{P}\omega)_{\text{eff}})$ based on Scott's graph model, we obtain the so-called *enumeration* Weihrauch lattice.

More generally, any partial combinatory algebra \mathbb{A} with an elementary subalgebra \mathbb{A}' begets a relative realizability topos $\mathbf{RT}(\mathbb{A}, \mathbb{A}')$, see [5], and thereby a Weihrauch-style reducibility lattice $\mathcal{W}_{\mathbb{A}, \mathbb{A}'}$. Of particular interest are examples of pcas \mathbb{A} that are also topological spaces, with \mathbb{A}' their effective parts. Among these are van Oosten's pca of sequential functionals, universal Scott domain \mathbb{U} , Plotkin's universal coherent domain \mathbf{T}^{ω} , and others.

We propose a new direction of research that studies the alternative Weihrauch lattices. We expect that John Longley's notion of *simulation* [3], also known as *applicative morphism*, and his analysis of topological pcas [4] will be of some help in establishing basic results, and in particular in relating the variants of Weihrauch lattices.

References

- 1 Andrej Bauer. Instance reducibility and Weihrauch degrees. *Logical Methods in Computer Science*, 18(3), 2022.
- 2 Stephen Cole Kleene and Richard Eugène Vesley. *The Foundations of Intuitionistic Mathematics, especially in relation to recursive functions*. North-Holland Publishing Company, 1965.
- 3 J. Longley. *Realizability Toposes and Language Semantics*. PhD thesis, Edinburgh University, 1995.
- 4 John Longley. On the ubiquity of certain total type structures. *Mathematical Structures in Computer Science*, 17(5):841 – 953, 2007.
- 5 Jaap van Oosten. *Realizability: An Introduction To Its Categorical Side*, volume 152 of *Studies in logic and the foundations of mathematics*. Elsevier, 2008.

4.2 Interior operators in the Weihrauch lattice

Jun Le Goh (National University of Singapore, SG)

License  Creative Commons BY 4.0 International license
© Jun Le Goh


Joint work of Jun Le Goh, Vasco Brattka, Damir Dzhafarov, Reed Solomon, Keita Yokoyama, Vittorio Cipriani, Arno Pauly

1. Brattka defined an interior operator on the Weihrauch degrees called the upper Turing cone version of a problem. This problem is induced by the closure operator on $\mathcal{P}(\mathbb{N}^{\mathbb{N}})$ given by upward closure under Turing reducibility.
Question: Which other interior operators can we form by considering closure operators on $\mathcal{P}(\mathbb{N}^{\mathbb{N}})$?

2. The first-order part of a problem f is the maximum Weihrauch degree of a problem g with codomain \mathbb{N} which reduces to f (Dzhafarov, Solomon, Yokoyama). The k -finitary part of a problem f is the maximum Weihrauch degree of a problem g with codomain \mathbf{k} which reduces to f (Cipriani, Pauly). Pauly observed during this Dagstuhl meeting that for each represented space \mathbf{X} and each problem f , the maximum Weihrauch degree among all problems with codomain \mathbf{X} which reduce to f exists.
 Question: For which other represented spaces is this maximum useful? How about Sierpinski space?

4.3 Question on the strength of the infinite loop closure

Takayuki Kihara (Nagoya University, JP)

License  Creative Commons BY 4.0 International license
 © Takayuki Kihara


Recently, Brattka introduced the notion of infinite loop operation on Weihrauch problems. Applying Yoshimura's unpublished theorem, one can see that the Weihrauch problem F is closed under the infinite loop operation (the inverse limit) if and only if F -relative realizability validates the axiom of dependent choice. Therefore, it is an important problem to investigate which Weihrauch problems are closed under the infinite loop operation. Here, we ask about the strength of the infinite loop closure of LLPO_k (all-or-counique choice on k).

Question: Is $\text{LLPO}_{k+1}^{\infty\infty\infty\cdots} <_{\text{W}} \text{LLPO}_k^{\infty\infty\infty\cdots}$?

This problem was solved by myself during the conference. That is, $\text{LLPO}_k^{\infty\infty\infty\cdots}$ is equivalent to DNR_k , and thus the problem is positively resolved.

4.4 Strong Weihrauch compositional product

Alberto Marcone (University of Udine, IT)

License  Creative Commons BY 4.0 International license
 © Alberto Marcone

Joint work of Alberto Marcone, Gian Marco Osso

Main reference Alberto Marcone, Gian Marco Osso: “The Galvin-Prikry Theorem in the Weihrauch lattice”, CoRR, Vol. abs/2410/06928 2024.

URL <https://arxiv.org/abs/2410.06928>

In the Weihrauch degrees we have an explicit definition of a multi-valued function $f \star g$ such that

$$f \star g \equiv_{\text{W}} \max_{\leq_{\text{W}}} \{h \circ k : h \leq_{\text{W}} f \wedge k \leq_{\text{W}} g\}.$$

In the paper with Gian Marco Osso we define a multi-valued function $f \tilde{\star} g$ such that if g is a cylinder then

$$f \tilde{\star} g \equiv_{\text{sW}} \max_{\leq_{\text{sW}}} \{h \circ k : h \leq_{\text{sW}} f \wedge k \leq_{\text{sW}} g\}.$$

$\tilde{\star}$ has some nice properties:


- $(f \tilde{\star} g) \tilde{\star} h \equiv_{\text{W}} f \tilde{\star} (g \tilde{\star} h)$;
- $(\text{id}_{\mathbb{N}} \times f) \tilde{\star} g \equiv_{\text{W}} f \star g$;
- if $g_0 \leq_{\text{W}} g_1$, then $f \tilde{\star} g_0 \leq_{\text{W}} f \tilde{\star} g_1$;
- if $f_0 \leq_{\text{sW}} f_1$ and $g_0 \leq_{\text{sW}} g_1$, then $f_0 \tilde{\star} g_0 \leq_{\text{sW}} f_1 \tilde{\star} g_1$.

However, we have examples where g is not a cylinder and $\max_{\leq_{\text{SW}}} \{h \circ k : h \leq_{\text{SW}} f \wedge k \leq_{\text{SW}} g\}$ either exists but is not represented by $f \star g$ or does not exist.

It would be interesting to characterize when $\max_{\leq_{\text{SW}}} \{h \circ k : h \leq_{\text{SW}} f \wedge k \leq_{\text{SW}} g\}$ exists and in those cases provide an explicit realizer of this strong Weihrauch degree.

4.5 A question about the the uniform content of index sets

Daniel Mourad (Nanjing University, CN)


License  Creative Commons BY 4.0 International license
© Daniel Mourad

Myhill showed that a set X is productive if and only if the complement of the halting set K is 1-reducible to X . It follows that the index sets Ind_n for partial computable functions are productive. Let IndRed be the problem which takes n and produces the graph of a 1-reduction from \overline{K} to Ind_n . Let IndProd be the problem which takes n and produces a graph of a productive set for Ind_n . Myhill's proofs are uniform in the index: $G \circ \text{IndRed}$ is Weihrauch equivalent to $G \circ \text{IndProd}$, where G is the Gödel function which takes a computable set to one of its indices. It turns out that one does not need the index to produce a graph in one of the directions: IndRed is Weihrauch reducible to IndProd .

Questions: Is IndRed Weihrauch equivalent to IndProd ? How about to $G \circ \text{IndRed}$?

4.6 On residual operators


Manlio Valenti (Swansea University, GB)

License  Creative Commons BY 4.0 International license
© Manlio Valenti

A residual lattice is a lattice equipped with a monoidal operator $*$ such that for every f and g there are maximum h and k such that $h * f \leq g$ and $f * k \leq g$. Given the large number of operators in the Weihrauch lattice, it is natural to ask what are the operators that make the Weihrauch lattice or its dual a residual lattice. Some of these questions have been already answered, but we still miss a complete picture. In particular, it is open whether there always exists a maximum h such that the compositional product $f * h$ is Weihrauch reducible to g .

4.7 Preservation results for well-quasiorders in the Weihrauch lattice

Arno Pauly (Swansea University, GB)

License  Creative Commons BY 4.0 International license
© Arno Pauly

Results of the form “If X is well-quasiordered, then so is $F(X)$ ” for various constructions of quasi-orders F have been a fruitful subject of study in reverse mathematics. Kruskal's and Higman's theorems are probably the most famous example, but already “If α is an ordinal, so is 2^α ” has non-trivial strength. At first glance, such results don't seem to have computational content per se. However, we can look at their contrapositives. The algorithmic task then becomes “Given a quasi-order X and a bad sequence in $F(X)$, find a bad sequence in X ”.


The task of finding a bad sequence in a quasi-ordered merely promised to be non-well was studied by Goh, Valenti and the author [1, 2]. By investigating how much the Weihrauch degree decreases if a bad sequence in $F(X)$ is provided as part of the input, we gain insight on how tightly the non-wqo-ness of $F(X)$ and X are linked in an effective way.

References

- 1 Jun Le Goh, Arno Pauly & Manlio Valenti. Finding descencing sequences through ill-founded linear orders. *Journal of Symbolic Logic* 86(2) (2021). doi:10.1017/jsl.2021.15.
- 2 Jun Le Goh, Arno Pauly & Manlio Valenti. The weakness of finding descending sequences in ill-founded linear orders. Preprint (2024). arXiv: 2401.11807.

4.8 A problem on the preservation of well-foundedness

Keita Yokoyama (Tohoku University, JP)

License  Creative Commons BY 4.0 International license
© Keita Yokoyama

Consider the following condition for a real $X \in 2^\omega$: (\dagger) if L is a computable linear order on ω with no computable infinite decreasing sequence, then X doesn't compute any infinite decreasing sequence for L .

Freund and Uftring [1] showed that if X is hyperimmune-free then X satisfies (\dagger) . Then, is the condition (\dagger) equivalent to being hyperimmune-free?

Joseph Miller answered this question. If X is 1-generic, then X satisfies (\dagger) , and thus (\dagger) is a strictly weaker notion than being hyperimmune-free.

References

- 1 Anton Freund and Patrick Uftring. More conservativity for weak König's lemma. Preprint (2024). arXiv: 2410.20591.

5 Bibliography on Weihrauch Complexity

For an always up-to-date version of this bibliography, see

<http://cca-net.de/publications/weibib.php>.

References

- 1 Nathanael Ackerman, Julian Asilis, Jieqi Di, Cameron Freer, and Jean-Baptiste Tristan. Computable PAC learning of continuous features. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '22*, New York, NY, USA, 2022. Association for Computing Machinery.
- 2 Nathanael L. Ackerman, Cameron E. Freer, and Daniel M. Roy. On computability and disintegration. *Mathematical Structures in Computer Science*, 27(8):1287–1314, 2017.
- 3 Djamel Eddine Amir. *Computability of Topological Spaces*. Ph.D. thesis, Université de Lorraine, 2023.
- 4 Djamel Eddine Amir and Mathieu Hoyrup. Comparing computability in two topologies. hal-03702999, 2022.
- 5 Djamel Eddine Amir and Mathieu Hoyrup. Strong computable type. arXiv 2210.08309, 2022.
- 6 Djamel Eddine Amir and Mathieu Hoyrup. Strong computable type. *Computability*, 12(3):227–269, 2023.
- 7 Djamel Eddine Amir and Mathieu Hoyrup. Comparing computability in two topologies. *Journal of Symbolic Logic*, 89(3):1232–1250, 2024.
- 8 Paul-Elliot Anglès d’Auriac. *Infinite Computations in Algorithmic Randomness and Reverse Mathematics*. Ph.D. thesis, Université Paris-Est, 2020.
- 9 Paul-Elliot Anglès d’Auriac and Takayuki Kihara. A comparison of various analytic choice principles. *The Journal of Symbolic Logic*, 86(4):1452–1485, 2021.
- 10 Eric P. Astor, Damir D. Dzharfarov, Reed Solomon, and Jacob Suggs. The uniform content of partial and linear orders. *Annals of Pure and Applied Logic*, 168(6):1153 – 1171, 2017.
- 11 Andrej Bauer. Instance reducibility and Weihrauch degrees. arXiv 2106.01734, 2021.
- 12 Andrej Bauer. Instance reducibility and Weihrauch degrees. *Logical Methods in Computer Science*, 18(3):20:1–20:18, August 2022.
- 13 Nikolay Bazhenov, Marta Fiori-Carones, Lu Liu, and Alexander Melnikov. Primitive recursive reverse mathematics. *Annals of Pure and Applied Logic*, 175(1, Part A):103354, 2024.
- 14 Zach BeMent, Jeffry Hirst, and Asuka Wallace. Reverse mathematics and Weihrauch analysis motivated by finite complexity theory. *Computability*, 10(4):343–354, 2021.
- 15 Zack BeMent, Jeffry Hirst, and Asuka Wallace. Reverse mathematics and Weihrauch analysis motivated by finite complexity theory. arXiv 2105.01719, 2021.
- 16 Laurent Bienvenu and Rutger Kuyper. Parallel and serial jumps of Weak König’s Lemma. In Adam Day, Michael Fellows, Noam Greenberg, Bakhadyr Khoussainov, Alexander Melnikov, and Frances Rosamond, editors, *Computability and Complexity: Essays Dedicated to Rodney G. Downey on the Occasion of His 60th Birthday*, volume 10010 of *Lecture Notes in Computer Science*, pages 201–217. Springer, Cham, 2017.
- 17 H. Boche and V. Pohl. The solvability complexity index of sampling-based Hilbert transform approximations. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pages 1–4, 2019.
- 18 Vasco Brattka. Computable invariance. In Tao Jiang and D.T. Lee, editors, *Computing and Combinatorics*, volume 1276 of *Lecture Notes in Computer Science*, pages 146–155, Berlin, 1997. Springer. Third Annual Conference, COCOON’97, Shanghai, China, August 1997.
- 19 Vasco Brattka. Computable invariance. *Theoretical Computer Science*, 210:3–20, 1999.

- 20 Vasco Brattka. Effective Borel measurability and reducibility of functions. *Mathematical Logic Quarterly*, 51(1):19–44, 2005.
- 21 Vasco Brattka. Computability and analysis, a historical approach. In Arnold Beckmann, Laurent Bienvenu, and Nataša Jonoska, editors, *Pursuit of the Universal*, volume 9709 of *Lecture Notes in Computer Science*, pages 45–57, Switzerland, 2016. Springer. 12th Conference on Computability in Europe, CiE 2016, Paris, France, June 27 - July 1, 2016.
- 22 Vasco Brattka. The discontinuity problem. arXiv 2012.02143, 2020.
- 23 Vasco Brattka. Stashing-parallelization pentagons. *Logical Methods in Computer Science*, 17(4):20:1–20:29, 2021.
- 24 Vasco Brattka. Weihrauch complexity and the Hagen school of computable analysis. In Benedikt Löwe and Deniz Sarikaya, editors, *60 Jahre DVMLG*, volume 48 of *Tributes*, pages 13–44. College Publications, London, 2022.
- 25 Vasco Brattka. The discontinuity problem. *Journal of Symbolic Logic*, 88(3):1191–1212, 2023.
- 26 Vasco Brattka. On the complexity of computing Gödel numbers. arXiv 2302.04213, 2023.
- 27 Vasco Brattka. On the complexity of learning programs. In Gianluca Della Vedova, Besik Dundua, Steffen Lempp, and Florin Manea, editors, *Unity of Logic and Computation*, volume 13967 of *Lecture Notes in Computer Science*, pages 166–177, Cham, 2023. Springer. 19th Conference on Computability in Europe.
- 28 Vasco Brattka. Loops, inverse limits and non-determinism. arXiv arXiv:2501.17734, 2025.
- 29 Vasco Brattka, Andrea Cettolo, Guido Gherardi, Alberto Marcone, and Matthias Schröder. Addendum to: “The Bolzano-Weierstrass theorem is the jump of weak König’s lemma”. *Annals of Pure and Applied Logic*, 168(8):1605–1608, 2017.
- 30 Vasco Brattka, Matthew de Brecht, and Arno Pauly. Closed choice and a uniform low basis theorem. *Annals of Pure and Applied Logic*, 163:986–1008, 2012.
- 31 Vasco Brattka, Damir Dzhafarov, Alberto Marcone, and Arno Pauly, editors. *Special issue: Dagstuhl Seminar on Measuring the Complexity of Computational Content 2018*, volume 9 of *Computability - The Journal of the Association CiE*. IOS Press, 2020.
- 32 Vasco Brattka, Damir D. Dzhafarov, Alberto Marcone, and Arno Pauly, editors. *Measuring the Complexity of Computational Content: From Combinatorial Problems to Analysis (Dagstuhl Seminar 18361)*, volume 8 of *Dagstuhl Reports*, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- 33 Vasco Brattka and Guido Gherardi. Borel complexity of topological operations on computable metric spaces. *Journal of Logic and Computation*, 19(1):45–76, 2009.
- 34 Vasco Brattka and Guido Gherardi. Effective choice and boundedness principles in computable analysis. In Andrej Bauer, Peter Hertling, and Ker-I Ko, editors, *CCA 2009, Proceedings of the Sixth International Conference on Computability and Complexity in Analysis*, pages 95–106, Schloss Dagstuhl, Germany, 2009. Leibniz-Zentrum für Informatik.
- 35 Vasco Brattka and Guido Gherardi. Weihrauch degrees, omniscience principles and weak computability. In Andrej Bauer, Peter Hertling, and Ker-I Ko, editors, *CCA 2009, Proceedings of the Sixth International Conference on Computability and Complexity in Analysis*, pages 83–94, Schloss Dagstuhl, Germany, 2009. Leibniz-Zentrum für Informatik.
- 36 Vasco Brattka and Guido Gherardi. Effective choice and boundedness principles in computable analysis. *The Bulletin of Symbolic Logic*, 17(1):73–117, 2011.
- 37 Vasco Brattka and Guido Gherardi. Weihrauch degrees, omniscience principles and weak computability. *The Journal of Symbolic Logic*, 76(1):143–176, 2011.
- 38 Vasco Brattka and Guido Gherardi. Weihrauch goes Brouwerian. arXiv 1809.00380, 2018.
- 39 Vasco Brattka and Guido Gherardi. Completion of choice. arXiv 1910.13186, 2019.
- 40 Vasco Brattka and Guido Gherardi. Weihrauch goes Brouwerian. *The Journal of Symbolic Logic*, 85(4):1614–1653, 2020.

- 41 Vasco Brattka and Guido Gherardi. Completion of choice. *Annals of Pure and Applied Logic*, 172(3):102914, 2021.
- 42 Vasco Brattka, Guido Gherardi, and Rupert Hölzl. Las Vegas computability and algorithmic randomness. In Ernst W. Mayr and Nicolas Ollinger, editors, *32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015)*, volume 30 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 130–142, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- 43 Vasco Brattka, Guido Gherardi, and Rupert Hölzl. Probabilistic computability and choice. *Information and Computation*, 242:249–286, 2015.
- 44 Vasco Brattka, Guido Gherardi, Rupert Hölzl, and Arno Pauly. The Vitali covering theorem in the Weihrauch lattice. In Adam Day, Michael Fellows, Noam Greenberg, Bakhadyr Khoussainov, Alexander Melnikov, and Frances Rosamond, editors, *Computability and Complexity: Essays Dedicated to Rodney G. Downey on the Occasion of His 60th Birthday*, volume 10010 of *Lecture Notes in Computer Science*, pages 188–200. Springer, Cham, 2017.
- 45 Vasco Brattka, Guido Gherardi, and Alberto Marcone. The Bolzano-Weierstrass theorem is the jump of weak König’s lemma. *Annals of Pure and Applied Logic*, 163:623–655, 2012.
- 46 Vasco Brattka, Guido Gherardi, and Arno Pauly. Weihrauch complexity in computable analysis. arXiv 1707.03202, 2017.
- 47 Vasco Brattka, Guido Gherardi, and Arno Pauly. Weihrauch complexity in computable analysis. In Vasco Brattka and Peter Hertling, editors, *Handbook of Computability and Complexity in Analysis*, Theory and Applications of Computability, pages 367–417. Springer, Cham, 2021.
- 48 Vasco Brattka, Noam Greenberg, Iskander Kalimullin, and Mariya Soskova, editors. *Special issue: Oberwolfach Workshop on Computability Theory 2021*, volume 11 of *Computability - The Journal of the Association CiE*. IOS Press, 2022.
- 49 Vasco Brattka, Matthew Hendtlass, and Alexander P. Kreuzer. On the uniform computational content of computability theory. *Theory of Computing Systems*, 61(4):1376–1426, 2017.
- 50 Vasco Brattka, Matthew Hendtlass, and Alexander P. Kreuzer. On the uniform computational content of the Baire category theorem. *Notre Dame Journal of Formal Logic*, 59(4):605–636, 2018.
- 51 Vasco Brattka and Peter Hertling, editors. *Handbook of Computability and Complexity in Analysis*, Theory and Applications of Computability, Cham, 2021. Springer.
- 52 Vasco Brattka, Rupert Hölzl, and Rutger Kuyper. Monte Carlo computability. In Heribert Vollmer and Brigitte Vallée, editors, *34th Symposium on Theoretical Aspects of Computer Science (STACS 2017)*, volume 66 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:14, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- 53 Vasco Brattka, Akitoshi Kawamura, Alberto Marcone, and Arno Pauly, editors. *Measuring the Complexity of Computational Content: Weihrauch Reducibility and Reverse Analysis (Dagstuhl Seminar 15392)*, volume 5 of *Dagstuhl Reports*, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- 54 Vasco Brattka, Stéphane Le Roux, Joseph S. Miller, and Arno Pauly. The Brouwer fixed point theorem revisited. In Arnold Beckmann, Laurent Bienvenu, and Nataša Jonoska, editors, *Pursuit of the Universal*, volume 9709 of *Lecture Notes in Computer Science*, pages 58–67, Switzerland, 2016. Springer. 12th Conference on Computability in Europe, CiE 2016, Paris, France, June 27 - July 1, 2016.
- 55 Vasco Brattka, Stéphane Le Roux, Joseph S. Miller, and Arno Pauly. Connected choice and the Brouwer fixed point theorem. *Journal of Mathematical Logic*, 19(1):1–46, 2019.

- 56 Vasco Brattka, Stéphane Le Roux, and Arno Pauly. On the computational content of the Brouwer Fixed Point Theorem. In S. Barry Cooper, Anuj Dawar, and Benedikt Löwe, editors, *How the World Computes*, volume 7318 of *Lecture Notes in Computer Science*, pages 57–67, Berlin, 2012. Springer. Turing Centenary Conference and 8th Conference on Computability in Europe, CiE 2012, Cambridge, UK, June 2012.
- 57 Vasco Brattka and Arno Pauly. Computation with advice. In Xizhong Zheng and Ning Zhong, editors, *CCA 2010, Proceedings of the Seventh International Conference on Computability and Complexity in Analysis*, volume 24 of *Electronic Proceedings in Theoretical Computer Science*, pages 41–55, 2010.
- 58 Vasco Brattka and Arno Pauly. On the algebraic structure of Weihrauch degrees. *Logical Methods in Computer Science*, 14(4:4):1–36, 2018.
- 59 Vasco Brattka and Tahina Rakotoniana. On the uniform computational content of Ramsey’s theorem. *Journal of Symbolic Logic*, 82(4):1278–1316, 2017.
- 60 Vasco Brattka and Emmanuel Rauzy. Effective second countability in computable analysis. In Arnold Beckmann, Isabel Oitavem, and Florin Manea, editors, *Crossroads of Computability and Logic: Insights, Inspirations, and Innovations*, volume 15764 of *Lecture Notes in Computer Science*, pages 19–33, Cham, 2025. Springer. 21st Conference on Computability in Europe.
- 61 Vasco Brattka and Hendrik Smischlaew. Computability of initial value problems. arXiv arXiv:2501.00451, 2024.
- 62 Vasco Brattka and Hendrik Smischlaew. Computability of initial value problems. In Arnold Beckmann, Isabel Oitavem, and Florin Manea, editors, *Crossroads of Computability and Logic: Insights, Inspirations, and Innovations*, volume 15764 of *Lecture Notes in Computer Science*, page to appear, Cham, 2025. Springer. 21st Conference on Computability in Europe.
- 63 Matthew de Brecht, Arno Pauly, and Matthias Schröder. Overt choice. *Computability*, 9(3-4):169–191, 2020.
- 64 Merlin Carl. Generalized effective reducibility. arXiv 1601.01899, 2016.
- 65 Merlin Carl. Generalized effective reducibility. In Arnold Beckmann, Laurent Bienvenu, and Nataša Jonoska, editors, *Pursuit of the Universal*, volume 9709 of *Lecture Notes in Computer Science*, pages 225–233, Switzerland, 2016. Springer. 12th Conference on Computability in Europe, CiE 2016, Paris, France, June 27 - July 1, 2016.
- 66 Merlin Carl. *Ordinal Computability, An Introduction to Infinitary Machines*, volume 9. de Gruyter, Berlin, 2019.
- 67 Merlin Carl. Effectivity and reducibility with ordinal Turing machines. *Computability*, 10(4):289–304, 2021.
- 68 Raphaël Carroy. *Functions of the first Baire class*. PhD thesis, University of Lausanne and University Paris 7, 2013.
- 69 Raphaël Carroy. A quasi-order on continuous functions. *Journal of Symbolic Logic*, 78(2):663–648, 2013.
- 70 Raphaël Carroy and Yann Pequignot. A well-quasi-order for continuous functions. arXiv 2410.13150v1, 2024.
- 71 Peter A. Cholak, Damir D. Dzhafarov, Denis R. Hirschfeldt, and Ludovic Patey. Some results concerning the SRT_2^2 vs. COH problem. *Computability*, 9(3-4):193–217, 2020.
- 72 Vittorio Cipriani. *Many Problems, Different Frameworks, Classification of Problems in Computable Analysis and Algorithmic Learning Theory*. Ph.D. thesis, Università degli Studi di Udine, 2023.
- 73 Vittorio Cipriani, Alberto Marcone, and Manlio Valenti. The Weihrauch lattice at the level of $\Pi_1^1\text{-CA}_0$: the Cantor-Bendixson theorem. arXiv 2210.15556, 2022.

- 74 Vittorio Cipriani, Alberto Marcone, and Manlio Valenti. The Weihrauch lattice at the level of Π_1^1 - CA_0 : the Cantor-Bendixson theorem. *Journal of Symbolic Logic*, pages 1–39, 2025.
- 75 Vittorio Cipriani and Arno Pauly. The complexity of finding supergraphs. In Gianluca Della Vedova, Besik Dundua, Steffen Lempp, and Florin Manea, editors, *Unity of Logic and Computation*, volume 13967 of *Lecture Notes in Computer Science*, pages 178–189, Cham, 2023. Springer. 19th Conference on Computability in Europe.
- 76 Caleb Davis, Denis R. Hirschfeldt, Jeffrey Hirst, Jake Pardo, Arno Pauly, and Keita Yokoyama. Combinatorial principles equivalent to weak induction. *Computability*, 9(3-4):219–229, 2020.
- 77 Caleb Davis, Denis R. Hirschfeldt, Jeffrey L. Hirst, Jake Pardo, Arno Pauly, and Keita Yokoyama. Combinatorial principles equivalent to weak induction. arXiv 1812.09943, 2018.
- 78 Adam R. Day, Rod Downey, and Linda Westrick. Three topological reducibilities for discontinuous functions. *Transactions of the American Mathematical Society. Series B*, 9:859–895, 2022.
- 79 Matthew de Brecht. Levels of discontinuity, limit-computability, and jump operators. In Vasco Brattka, Hannes Diener, and Dieter Spreen, editors, *Logic, Computation, Hierarchies*, Ontos Mathematical Logic, pages 93–122. Walter de Gruyter, Boston, 2014.
- 80 François G. Dorais, Damir D. Dzhafarov, Jeffrey L. Hirst, Joseph R. Mileti, and Paul Shafer. On uniform relationships between combinatorial problems. *Transactions of the American Mathematical Society*, 368(2):1321–1359, 2016.
- 81 Rod Downey, Alexander Melnikov, and Keng Meng Ng. Foundations of online structure theory II: The operator approach. arXiv 2007.07401, 2020.
- 82 Damir Dzhafarov, Stephen Flood, Reed Solomon, and Linda Brown Westrick. Effectiveness for the dual Ramsey theorem. arXiv 1710.00070, 2017.
- 83 Damir Dzhafarov, Reed Solomon, and Manlio Valenti. The tree pigeonhole principle in the Weihrauch degrees. arXiv 2312.10535, 2023.
- 84 Damir Dzhafarov, Reed Solomon, and Manlio Valenti. The tree pigeonhole principle in the Weihrauch degrees. *Journal of Symbolic Logic*, 2025.
- 85 Damir Dzhafarov, Reed Solomon, and Keita Yokoyama. On the first-order parts of problems in the Weihrauch degrees. *Computability*, 13(3-4):363–375, 2024.
- 86 Damir D. Dzhafarov. Cohesive avoidance and strong reductions. *Proceedings of the American Mathematical Society*, 143(2):869–876, 2015.
- 87 Damir D. Dzhafarov. Strong reductions between combinatorial principles. *Journal of Symbolic Logic*, 81(4):1405–1431, 2016.
- 88 Damir D. Dzhafarov. Joins in the strong Weihrauch degrees. *Mathematical Research Letters*, 26(3):749–767, 2019.
- 89 Damir D. Dzhafarov, Jun Le Goh, Denis R. Hirschfeldt, Ludovic Patey, and Arno Pauly. Ramsey’s theorem and products in the Weihrauch degrees. arXiv 1804.10968, 2018.
- 90 Damir D. Dzhafarov, Jun Le Goh, Denis R. Hirschfeldt, Ludovic Patey, and Arno Pauly. Ramsey’s theorem and products in the Weihrauch degrees. *Computability*, 9(2):85–110, 2020.
- 91 Damir D. Dzhafarov, Denis R. Hirschfeldt, and Sarah C. Reitzes. Reduction games, provability, and compactness. arXiv 2008.00907, 2020.
- 92 Damir D. Dzhafarov, Denis R. Hirschfeldt, and Sarah C. Reitzes. Reduction games, provability, and compactness. *Journal of Mathematical Logic*, 22(3):2250009, 2022.
- 93 Damir D. Dzhafarov and Carl Mummert. *Reverse Mathematics. Theory and Applications of Computability*. Springer, 2022.
- 94 Damir D. Dzhafarov and Ludovic Patey. COH, SRT22, and multiple functionals. *Computability*, 10(2):111–121, 2021.

- 95 Damir D. Dzhabarov, Ludovic Patey, Reed Solomon, and Linda Brown Westrick. Ramsey's theorem for singletons and strong computable reducibility. *Proceedings of the American Mathematical Society*, 145, 2017.
- 96 Damir D. Dzhabarov, Reed Solomon, and Keita Yokoyama. On the first-order parts of problems in the Weihrauch degrees. arXiv 2301.12733, 2023.
- 97 Marta Fiori-Carones and Alberto Marcone. To reorient is easier than to orient: an on-line algorithm for reorientation of graphs. *Computability*, 10(3):215–233, 2021.
- 98 Marta Fiori-Carones, Paul Shafer, and Giovanni Soldà. An inside/outside Ramsey theorem and recursion theory. arXiv 2006.16969, 2020.
- 99 Marta Fiori-Carones, Paul Shafer, and Giovanni Soldà. An inside/outside Ramsey theorem and recursion theory. *Transactions of the American Mathematical Society*, 375(3):1977–2024, 2022.
- 100 Stephen Flood, Matthew Jura, Oscar Levin, and Tyler Markkanen. The computational strength of matchings in countable graphs. *Annals of Pure and Applied Logic*, 173(8):Paper No. 103133, 26, 2022.
- 101 Makoto Fujiwara. Parallelizations in Weihrauch reducibility and constructive reverse mathematics. In Marcella Anselmo, Gianluca Della Vedova, Florin Manea, and Arno Pauly, editors, *Beyond the Horizon of Computability*, pages 38–49, Cham, 2020. Springer International Publishing.
- 102 Makoto Fujiwara. Weihrauch and constructive reducibility between existence statements. *Computability*, 10(1), 2021.
- 103 Makoto Fujiwara, Kojiro Higuchi, and Takayuki Kihara. On the strength of marriage theorems and uniformity. *Mathematical Logic Quarterly*, 60(3):136–153, 2014.
- 104 Lorenzo Galeotti and Hugo Nobrega. Towards computable analysis on the generalized real line. In Jarkko Kari, Florin Manea, and Ion Petre, editors, *Unveiling Dynamics and Complexity*, volume 10307 of *Lecture Notes in Computer Science*, pages 246–257, Cham, 2017. Springer. 13th Conference on Computability in Europe, CiE 2017, Turku, Finland, June 12–16, 2017.
- 105 Guido Gherardi. An analysis of the lemmas of Urysohn and Urysohn-Tietze according to effective Borel measurability. In A. Beckmann, U. Berger, B. Löwe, and J.V. Tucker, editors, *Logical Approaches to Computational Barriers*, volume 3988 of *Lecture Notes in Computer Science*, pages 199–208, Berlin, 2006. Springer. Second Conference on Computability in Europe, CiE 2006, Swansea, UK, June 30–July 5, 2006.
- 106 Guido Gherardi. Effective Borel degrees of some topological functions. *Mathematical Logic Quarterly*, 52(6):625–642, 2006.
- 107 Guido Gherardi. *Some Results in Computable Analysis and Effective Borel Measurability*. PhD thesis, University of Siena, Department of Mathematics and Computer Science, Siena, 2006.
- 108 Guido Gherardi and Alberto Marcone. How incomputable is the separable Hahn-Banach theorem? In Vasco Brattka, Ruth Dillhage, Tanja Grubba, and Angela Klutsch, editors, *CCA 2008, Fifth International Conference on Computability and Complexity in Analysis*, volume 221 of *Electronic Notes in Theoretical Computer Science*, pages 85–102. Elsevier, 2008. CCA 2008, Fifth International Conference, Hagen, Germany, August 21–24, 2008.
- 109 Guido Gherardi and Alberto Marcone. How incomputable is the separable Hahn-Banach theorem? *Notre Dame Journal of Formal Logic*, 50(4):393–425, 2009.
- 110 Guido Gherardi, Alberto Marcone, and Arno Pauly. Projection operators in the Weihrauch lattice. arXiv 1805.12026, 2018.
- 111 Guido Gherardi, Alberto Marcone, and Arno Pauly. Projection operators in the Weihrauch lattice. *Computability*, 8(3, 4):281–304, 2019.
- 112 Kenneth Gill. Indivisibility and uniform computational strength. arXiv 2312.03919, 2023.

- 113 Kenneth Gill. Indivisibility and uniform computational strength. *Logical Methods in Computer Science*, 21(2):22:1–22:23, June 2025.
- 114 Jun Le Goh. *Measuring the Relative Complexity of Mathematical Constructions and Theorems*. Ph.D. thesis, Cornell University, August 2019.
- 115 Jun Le Goh. Some computability-theoretic reductions between principles around ATR_0 . arXiv 1905.06868, 2019.
- 116 Jun Le Goh. Compositions of multivalued functions. *Computability*, 9(3-4):231–247, 2020.
- 117 Jun Le Goh. Embeddings between well-orderings: Computability-theoretic reductions. *Annals of Pure and Applied Logic*, 171(6):102789, 2020.
- 118 Jun Le Goh, Arno Pauly, and Manlio Valenti. Finding descending sequences through ill-founded linear orders. arXiv 2010.03840, 2020.
- 119 Jun Le Goh, Arno Pauly, and Manlio Valenti. Finding descending sequences through ill-founded linear orders. *The Journal of Symbolic Logic*, 86(2):817–854, 2021.
- 120 Jun Le Goh, Arno Pauly, and Manlio Valenti. The weakness of finding descending sequences in ill-founded linear orders. arXiv 2401.11807, 2024.
- 121 Jun Le Goh, Arno Pauly, and Manlio Valenti. The weakness of finding descending sequences in ill-founded linear orders. In Ludovic Levy Patey, Elaine Pimentel, Lorenzo Galeotti, and Florin Manea, editors, *Twenty Years of Theoretical and Practical Synergies*, pages 339–350, Cham, 2024. Springer Nature Switzerland.
- 122 Noam Greenberg, Rutger Kuyper, and Dan Turetsky. Cardinal invariants, non-lowness classes, and Weihrauch reducibility. *Computability*, 8(3, 4):305–346, 2019.
- 123 Noam Greenberg, Joseph S. Miller, and An Nies. Highness properties close to PA completeness. *Israel Journal of Mathematics*, 244:419–465, 2021.
- 124 Kirill Gura, Jeffery L. Hirst, and Carl Mummert. On the existence of a connected component of a graph. *Computability*, 4(2):103–117, 2015.
- 125 Peter Hertling. Stetige Reduzierbarkeit auf Σ^ω von Funktionen mit zweielementigem Bild und von zweistetigen Funktionen mit diskretem Bild. Informatik Berichte 153, FernUniversität Hagen, Hagen, December 1993.
- 126 Peter Hertling. A topological complexity hierarchy of functions with finite range. Technical Report 223, Centre de recerca matemàtica, Institut d’estudis catalans, Barcelona, Barcelona, October 1993. Workshop on Continuous Algorithms and Complexity, Barcelona, October, 1993.
- 127 Peter Hertling. Topologische Komplexitätsgrade von Funktionen mit endlichem Bild. Informatik Berichte 152, FernUniversität Hagen, Hagen, December 1993.
- 128 Peter Hertling. *Unstetigkeitsgrade von Funktionen in der effektiven Analysis*. PhD thesis, Fachbereich Informatik, FernUniversität Hagen, 1996. Dissertation.
- 129 Peter Hertling. Forests describing Wadge degrees and topological Weihrauch degrees of certain classes of functions and relations. *Computability*, 9(3-4):249–307, 2020.
- 130 Peter Hertling and Victor Selivanov. Complexity issues for preorders on finite labeled forests. In Benedikt Löwe, Dag Normann, Ivan Soskov, and Alexandra Soskova, editors, *Models of computation in context*, volume 6735 of *Lecture Notes in Computer Science*, pages 112–121, Heidelberg, 2011. Springer. 7th Conference on Computability in Europe, CiE 2011, Sofia, Bulgaria, June 27–July 2, 2011.
- 131 Peter Hertling and Victor Selivanov. Complexity issues for preorders on finite labeled forests. In Vasco Brattka, Hannes Diener, and Dieter Spreen, editors, *Logic, Computation, Hierarchies*, *Ontos Mathematical Logic*, pages 165–190. Walter de Gruyter, Boston, 2014.
- 132 Peter Hertling and Klaus Weihrauch. Levels of degeneracy and exact lower complexity bounds for geometric algorithms. In *Proceedings of the Sixth Canadian Conference on Computational Geometry*, pages 237–242, 1994. Saskatoon, Saskatchewan, August 2–6, 1994.

- 133 Peter Hertling and Klaus Weihrauch. On the topological classification of degeneracies. *Informatik Berichte* 154, FernUniversität Hagen, Hagen, February 1994.
- 134 Kojiro Higuchi. *Degree Structures of Mass Problems and Choice Functions*. PhD thesis, Mathematical Institute, Tohoku University, Sendai, Japan, January 2012.
- 135 Kojiro Higuchi and Takayuki Kihara. Inside the Muchnik degrees I: Discontinuity, learnability and constructivism. *Annals of Pure and Applied Logic*, 165(5):1058–1114, 2014.
- 136 Kojiro Higuchi and Takayuki Kihara. Inside the Muchnik degrees II: The degree structures induced by the arithmetical hierarchy of countably continuous functions. *Annals of Pure and Applied Logic*, 165(6):1201–1241, 2014.
- 137 Kojiro Higuchi and Arno Pauly. The degree structure of Weihrauch reducibility. *Log. Methods Comput. Sci.*, 9(2):2:02, 17, 2013.
- 138 Denis R. Hirschfeldt. *Slicing the Truth: On the Computable and Reverse Mathematics of Combinatorial Principles*, volume 28 of *Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*. World Scientific, Singapore, 2015.
- 139 Denis R. Hirschfeldt. Some questions in computable mathematics. In Adam Day, Michael Fellows, Noam Greenberg, Bakhadyr Khoussainov, Alexander Melnikov, and Frances Rosamond, editors, *Computability and Complexity: Essays Dedicated to Rodney G. Downey on the Occasion of His 60th Birthday*, volume 10010 of *Lecture Notes in Computer Science*, pages 22–55. Springer, Cham, 2017.
- 140 Denis R. Hirschfeldt and Carl G. Jockusch. On notions of computability-theoretic reduction between Π_2^1 principles. *Journal of Mathematical Logic*, 16(1):1650002, 59, 2016.
- 141 Jeffry L. Hirst. Leaf management. arXiv 1812.09762, 2018.
- 142 Jeffry L. Hirst. Leaf management. *Computability*, 9(3-4):309–314, 2020.
- 143 Jeffry L. Hirst and Carl Mummert. Reverse mathematics of matroids. In Adam Day, Michael Fellows, Noam Greenberg, Bakhadyr Khoussainov, Alexander Melnikov, and Frances Rosamond, editors, *Computability and Complexity: Essays Dedicated to Rodney G. Downey on the Occasion of His 60th Birthday*, volume 10010 of *Lecture Notes in Computer Science*, pages 143–159. Springer, Cham, 2017.
- 144 Jeffry L. Hirst and Carl Mummert. Using Ramsey’s theorem once. *Archive for Mathematical Logic*, 58(7-8):857–866, 2019.
- 145 Jeffry L. Hirst and Carl Mummert. Banach’s theorem in higher-order reverse mathematics. *Computability*, 12(3):203–225, 2023.
- 146 Rupert Hölzl and Paul Shafer. Universality, optimality, and randomness deficiency. *Annals of Pure and Applied Logic*, 166(10):1049–1069, 2015.
- 147 Mathieu Hoyrup. Genericity of weakly computable objects. *Theory of Computing Systems*, 60(3):396–420, 2017.
- 148 Mathieu Hoyrup. Notes on overt choice. *Computability*, 12(4):351–369, 2023.
- 149 Mathieu Hoyrup. *Topological Aspects of Representations in Computable Analysis*. PhD thesis, Laboratoire Lorrain de Recherche en Informatique et ses Applications, Nancy, France, 2023. Habilitation Thesis.
- 150 Mathieu Hoyrup, Cristóbal Rojas, and Klaus Weihrauch. Computability of the Radon-Nikodym derivative. In Benedikt Löwe, Dag Normann, Ivan Soskov, and Alexandra Soskova, editors, *Models of Computation in Context*, volume 6735 of *Lecture Notes in Computer Science*, pages 132–141, Heidelberg, 2011. Springer.
- 151 Mathieu Hoyrup, Cristóbal Rojas, and Klaus Weihrauch. Computability of the Radon-Nikodym derivative. *Computability*, 1(1):3–13, 2012.
- 152 Noah A. Hughes. *Applications of Computability Theory to Infinitary Combinatorics*. Ph.D. thesis, University of Connecticut, 2021.

- 153 Akitoshi Kawamura. Lipschitz continuous ordinary differential equations are polynomial-space complete. In *24th Annual IEEE Conference on Computational Complexity*, pages 149–160. IEEE Computer Soc., Los Alamitos, CA, 2009.
- 154 Akitoshi Kawamura. Lipschitz continuous ordinary differential equations are polynomial-space complete. *Computational Complexity*, 19(2):305–332, 2010.
- 155 Akitoshi Kawamura and Stephen Cook. Complexity theory for operators in analysis. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, pages 495–502, New York, 2010. ACM.
- 156 Akitoshi Kawamura and Hiroyuki Ota. Small complexity classes for computable analysis. In *Mathematical foundations of computer science 2014. Part II*, volume 8635 of *Lecture Notes in Comput. Sci.*, pages 432–444. Springer, Heidelberg, 2014.
- 157 Akitoshi Kawamura, Hiroyuki Ota, Carsten Rösnick, and Martin Ziegler. Computational complexity of smooth differential equations. In *Mathematical foundations of computer science 2012*, volume 7464 of *Lecture Notes in Comput. Sci.*, pages 578–589. Springer, Heidelberg, 2012.
- 158 Akitoshi Kawamura, Hiroyuki Ota, Carsten Rösnick, and Martin Ziegler. Computational complexity of smooth differential equations. *Logical Methods in Computer Science*, 10:1:6,15, 2014.
- 159 Akitoshi Kawamura and Arno Pauly. Function spaces for second-order polynomial time. In *Language, life, limits*, volume 8493 of *Lecture Notes in Comput. Sci.*, pages 245–254. Springer, Cham, 2014.
- 160 Takayuki Kihara. Borel-piecewise continuous reducibility for uniformization problems. *Logical Methods in Computer Science*, 12(4), October 2016.
- 161 Takayuki Kihara. Degrees of incomputability, realizability and constructive reverse mathematics. arXiv 2002.10712, 2020.
- 162 Takayuki Kihara. Lawvere-Tierney topologies for computability theorists. arXiv 2106.03061, 2021.
- 163 Takayuki Kihara. Topological reducibilities for discontinuous functions and their structures. *Israel Journal of Mathematics*, 2022.
- 164 Takayuki Kihara. Lawvere-Tierney topologies for computability theorists. *Transactions of the American Mathematical Society, Series B*, 10(2):48–85, 2023.
- 165 Takayuki Kihara, Alberto Marcone, and Arno Pauly. Searching for an analogue of ATR in the Weihrauch lattice. arXiv 1812.01549, 2018.
- 166 Takayuki Kihara, Alberto Marcone, and Arno Pauly. Searching for an analogue of ATR_0 in the Weihrauch lattice. *Journal of Symbolic Logic*, 85(3):1006–1043, 2020.
- 167 Takayuki Kihara and Arno Pauly. Dividing by zero - how bad is it, really? In Piotr Faliszewski, Anca Muscholl, and Rolf Niedermeier, editors, *41st International Symposium on Mathematical Foundations of Computer Science (MFCS 2016)*, volume 58 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 58:1–58:14, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- 168 Takayuki Kihara and Arno Pauly. Dividing by zero—how bad is it, really? In *41st International Symposium on Mathematical Foundations of Computer Science*, volume 58 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 58, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.
- 169 Takayuki Kihara and Arno Pauly. Finite choice, convex choice and sorting. In T.V. Gopal and Junzo Watada, editors, *Theory and applications of models of computation*, volume 11436 of *Lecture Notes in Computer Science*, pages 378–393. Springer, Cham, 2019. 15th Annual Conference, TAMC 2019, Kitakyushu, Japan, April 13–16, 2019.
- 170 Ulrich Kohlenbach. On the reverse mathematics and Weihrauch complexity of moduli of regularity and uniqueness. *Computability*, 8(3, 4):377–387, 2019.

- 171 Alexander P. Kreuzer. On the strength of weak compactness. *Computability*, 1(2):171–179, 2012.
- 172 Alexander P. Kreuzer. Bounded variation and the strength of Helly’s selection theorem. *Logical Methods in Computer Science*, 10(4:16):1–23, 2014.
- 173 Alexander P. Kreuzer. From Bolzano-Weierstraß to Arzelà-Ascoli. *Mathematical Logic Quarterly*, 60(3):177–183, 2014.
- 174 Oleg V. Kudinov, Victor L. Selivanov, and Anton V. Zhukov. Undecidability in Weihrauch degrees. In Fernando Ferreira, Benedikt Löwe, Elvira Mayordomo, and Luís Mendes Gomes, editors, *Programs, Proofs, Processes*, volume 6158 of *Lecture Notes in Computer Science*, pages 256–265, Berlin, 2010. Springer. 6th Conference on Computability in Europe, CiE 2010, Ponta Delgada, Azores, Portugal, June/July 2010.
- 175 Rutger Kuyper. On Weihrauch reducibility and intuitionistic reverse mathematics. *Journal of Symbolic Logic*, 82(4):1438–1458, 2017.
- 176 Stéphane Le Roux and Arno Pauly. Closed choice for finite and for convex sets. In Paola Bonizzoni, Vasco Brattka, and Benedikt Löwe, editors, *The Nature of Computation. Logic, Algorithms, Applications*, volume 7921 of *Lecture Notes in Computer Science*, pages 294–305, Berlin, 2013. Springer. 9th Conference on Computability in Europe, CiE 2013, Milan, Italy, July 1-5, 2013.
- 177 Stéphane Le Roux and Arno Pauly. Finite choice, convex choice and finding roots. *Logical Methods in Computer Science*, 11(4):4:6, 31, 2015.
- 178 Stéphane Le Roux and Arno Pauly. Weihrauch degrees of finding equilibria in sequential games (extended abstract). In Arnold Beckmann, Victor Mitrană, and Mariya Soskova, editors, *Evolving Computability*, volume 9136 of *Lecture Notes in Computer Science*, pages 246–257, Cham, 2015. Springer. 11th Conference on Computability in Europe, CiE 2015, Bucharest, Romania, June 29–July 3, 2015.
- 179 Steffen Lempp, Alberto Marcone, and Manlio Valenti. Chains and antichains in the Weihrauch lattice. arXiv 2411.07792, 2024.
- 180 Steffen Lempp, Joseph S. Miller, Arno Pauly, Mariya I. Soskova, and Manlio Valenti. Minimal covers in the Weihrauch degrees. arXiv 2311.12676, 2023.
- 181 Steffen Lempp, Joseph S. Miller, Arno Pauly, Mariya I. Soskova, and Manlio Valenti. Minimal covers in the Weihrauch degrees. *Proceedings of the American Mathematical Society*, 152(11):4893–4901, 2024.
- 182 Ang Li. Countable ordered groups and Weihrauch reducibility. arXiv 2409.19229, 2024.
- 183 Patrick Lutz. *Results on Martin’s Conjecture*. PhD thesis, University of California, Berkeley, 2021.
- 184 Patrick Lutz. The Solecki dichotomy and the Possner-Robinson theorem are almost equivalent. arXiv 2301.07259, 2023.
- 185 Patrick Lutz and Benjamin Siskind. Part 1 of Martin’s conjecture for order-preserving and measure-preserving functions. arXiv 2305.19646, 2023.
- 186 Patrick Lutz and Benjamin Siskind. Part 1 of Martin’s conjecture for order-preserving and measure-preserving functions. *Journal of the American Mathematical Society*, 2024.
- 187 Alberto Marcone and Gian Marco Osso. The Galvin-Prikry theorem in the Weihrauch lattice. arXiv 2410.06928, 2024.
- 188 Alberto Marcone and Manlio Valenti. The open and clopen Ramsey theorems in the Weihrauch lattice. arXiv 2003.04245, 2020.
- 189 Alberto Marcone and Manlio Valenti. Effective aspects of Hausdorff and Fourier dimension. arXiv 2108.06941, 2021.
- 190 Alberto Marcone and Manlio Valenti. The open and clopen Ramsey theorems in the Weihrauch lattice. *The Journal of Symbolic Logic*, 86(1):316–351, 2021.

- 191 Alberto Marcone and Manlio Valenti. Effective aspects of Hausdorff and Fourier dimension. *Computability*, 11(3-4):299–333, 2022.
- 192 Samuele Maschio and Davide Trotta. A topos for extended Weihrauch degrees. arXiv 2505.08697, 2025.
- 193 Benoit Monin and Ludovic Patey. Π_1^0 -encodability and omniscient reductions. *Notre Dame Journal of Formal Logic*, 60(1):1–12, 2019.
- 194 Daniel Mourad. There is no composition in the computable reducibility degrees. arXiv 2405.15281, 2024.
- 195 Uwe Mylatz. *Vergleich unstetiger Funktionen: “Principle of Omniscience” und Vollständigkeit in der C-Hierarchie*. PhD thesis, Faculty for Mathematics and Computer Science, University Hagen, Hagen, Germany, 2006. Ph.D. thesis.
- 196 Eike Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11:4:20,44, 2015.
- 197 Eike Neumann and Arno Pauly. A topological view on algebraic computation models. *Journal of Complexity*, 44(Supplement C):1–22, 2018.
- 198 David Nichols. Strong reductions between relatives of the stable Ramsey’s theorem. arXiv 1711.06532, 2017.
- 199 Hugo Nobrega. *Games for functions - Baire classes, Weihrauch degrees, Transfinite Computations, and Ranks*. PhD thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, 2018.
- 200 Hugo Nobrega and Arno Pauly. Game characterizations and lower cones in the Weihrauch degrees. In Jarkko Kari, Florin Manea, and Ion Petre, editors, *Unveiling Dynamics and Complexity*, volume 10307 of *Lecture Notes in Computer Science*, pages 327–337, Cham, 2017. Springer. 13th Conference on Computability in Europe, CiE 2017, Turku, Finland, June 12-16, 2017.
- 201 Hugo Nobrega and Arno Pauly. Game characterizations and lower cones in the Weihrauch degrees. *Logical Methods in Computer Science*, 15(3):Paper No. 11, 29, 2019.
- 202 Ludovic Patey. *The reverse mathematics of Ramsey-type theorems*. PhD thesis, Université Paris Diderot, Paris, France, 2016.
- 203 Ludovic Patey. The weakness of being cohesive, thin or free in reverse mathematics. *Israel Journal of Mathematics*, 216:905–955, 2016.
- 204 Arno Pauly. How discontinuous is computing Nash equilibria? (Extended abstract). In Andrej Bauer, Peter Hertling, and Ker-I Ko, editors, *6th International Conference on Computability and Complexity in Analysis (CCA’09)*, volume 11 of *OpenAccess Series in Informatics (OASICS)*, Dagstuhl, Germany, 2009. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- 205 Arno Pauly. How incomputable is finding Nash equilibria? *Journal of Universal Computer Science*, 16(18):2686–2710, 2010.
- 206 Arno Pauly. On the (semi)lattices induced by continuous reducibilities. *Mathematical Logic Quarterly*, 56(5):488–502, 2010.
- 207 Arno Pauly. *Computable Metamathematics and its Application to Game Theory*. PhD thesis, University of Cambridge, Computer Laboratory, Clare College, Cambridge, 2011. Ph.D. thesis.
- 208 Arno Pauly. Computability on the space of countable ordinals. arXiv 1501.00386, 2015.
- 209 Arno Pauly. Many-one reductions and the category of multivalued functions. *Mathematical Structures in Computer Science*, 27(3):376–404, 2017.
- 210 Arno Pauly. An update on Weihrauch complexity, and some open questions. arXiv 2008.11168, 2020.
- 211 Arno Pauly and Matthew de Brecht. Towards synthetic descriptive set theory: An instantiation with represented spaces. arXiv 1307.1850, 2013.

- 212 Arno Pauly and Matthew de Brecht. Non-deterministic computation and the Jayne-Rogers theorem. In Benedikt Löwe and Glynn Winskel, editors, *Proceedings 8th International Workshop on Developments in Computational Models, DCM 2012, Cambridge, United Kingdom, 17 June 2012.*, volume 143 of *Electronic Proceedings in Theoretical Computer Science*, pages 87–96, 2014.
- 213 Arno Pauly and Matthew de Brecht. Descriptive set theory in the category of represented spaces. In *30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 438–449, 2015.
- 214 Arno Pauly, Willem Fouché, and George Davie. Weihrauch-completeness for layerwise computability. *Logical Methods in Computer Science*, 14(2), May 2018.
- 215 Arno Pauly and Willem L. Fouché. How constructive is constructing measures? *Journal of Logic & Analysis*, 9(c3):1–44, 2017.
- 216 Arno Pauly, Cécilia Pradic, and Giovanni Soldà. On the Weihrauch degree of the additive Ramsey theorem. *Computability*, 13(3–4):459–483, 2024.
- 217 Arno Pauly and Giovanni Soldà. Sequential discontinuity and first-order problems. arXiv 2401.12641, 2024.
- 218 Arno Pauly and Giovanni Soldà. Sequential discontinuity and first-order problems. In Ludovic Levy Patey, Elaine Pimentel, Lorenzo Galeotti, and Florin Manea, editors, *Twenty Years of Theoretical and Practical Synergies*, pages 351–365, Cham, 2024. Springer Nature Switzerland.
- 219 Arno Pauly and Florian Steinberg. Representations of analytic functions and Weihrauch degrees. In *Computer science—theory and applications*, volume 9691 of *Lecture Notes in Computer Science*, pages 367–381, Cham, 2016. Springer.
- 220 Arno Pauly and Florian Steinberg. Comparing representations for function spaces in computable analysis. *Theory of Computing Systems*, 62:557–582, 2018.
- 221 Arno Pauly and Hideki Tsuiki. Computable dyadic subbases and \mathbb{T}^ω -representations of compact sets. arXiv 1604.0258, 2016.
- 222 Michelle Porter, Adam Day, and Rodney Downey. Notes on computable analysis. *Theory of Computing Systems*, 60(1):53–111, 2017.
- 223 Cécilia Pradic and Ian Price. Weihrauch problems as containers. arXiv 2501.17250, 2025.
- 224 Pierre Pradic and Giovanni Soldà. On the Weihrauch degree of the additive Ramsey theorem over the rationals. In *Revolutions and revelations in computability*, volume 13359 of *Lecture Notes in Computer Science*, pages 259–271. Springer, Cham, 2022.
- 225 Tahina Rakotonaiaina. *On the Computational Strength of Ramsey’s Theorem*. PhD thesis, Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, South Africa, 2015. Ph.D. thesis.
- 226 Victor Selivanov. Total representations. *Logical Methods in Computer Science*, 9:2:5, 30, 2013.
- 227 Victor Selivanov. Q-Wadge degrees as free structures. *Computability*, 9(3–4):327–341, 2020.
- 228 Giovanni Soldà and Manlio Valenti. Algebraic properties of the first-order part of a problem. arXiv 2203.16298, 2022.
- 229 Giovanni Soldà and Manlio Valenti. Algebraic properties of the first-order part of a problem. *Annals of Pure and Applied Logic*, 174(7):103270, 2023.
- 230 Reed Solomon. Computable reductions and reverse mathematics. In Arnold Beckmann, Laurent Bienvenu, and Nataša Jonoska, editors, *Pursuit of the Universal*, volume 9709 of *Lecture Notes in Computer Science*, pages 182–191, Switzerland, 2016. Springer. 12th Conference on Computability in Europe, CiE 2016, Paris, France, June 27 - July 1, 2016.
- 231 Yudai Suzuki and Keita Yokoyama. Searching problems above arithmetical transfinite recursion. *Annals of Pure and Applied Logic*, 175(10):Paper No. 103488, 31, 2024.

- 232 Nazanin R. Tavana and Klaus Weihrauch. Turing machines on represented sets, a model of computation for analysis. *Logical Methods in Computer Science*, 7(2):2:19, 21, 2011.
- 233 Holger Thies. *Uniform computational complexity of ordinary differential equations with applications to dynamical systems and exact real arithmetic*. PhD thesis, Graduate School of Arts and Sciences, University of Tokyo, Tokyo, Japan, 2018.
- 234 Davide Trotta, Manlio Valenti, and Valeria de Paiva. Categorifying computable reducibilities. arXiv 2208.08656, 2022.
- 235 Patrick Uftring. The characterization of Weihrauch reducibility in systems containing $E-PA^\omega + QF-AC^{0,0}$. arXiv 2003.13331, 2020.
- 236 Patrick Uftring. The characterization of Weihrauch reducibility in systems containing $E-PA^\omega + QF-AC^{0,0}$. *The Journal of Symbolic Logic*, 86(1):224–261, 2021.
- 237 Patrick Uftring. Weihrauch degrees without roots. arXiv 2102.11832, 2023.
- 238 Manlio Valenti. *A journey through computability, topology and analysis*. Ph.D. thesis, Università degli Studi di Udine, 2021.
- 239 Klaus Weihrauch. The degrees of discontinuity of some translators between representations of the real numbers. Technical Report TR-92-050, International Computer Science Institute, Berkeley, July 1992.
- 240 Klaus Weihrauch. The degrees of discontinuity of some translators between representations of the real numbers. Informatik Berichte 129, FernUniversität Hagen, Hagen, July 1992.
- 241 Klaus Weihrauch. The TTE-interpretation of three hierarchies of omniscience principles. Informatik Berichte 130, FernUniversität Hagen, Hagen, September 1992.
- 242 Klaus Weihrauch. *Computable Analysis*. Springer, Berlin, 2000.
- 243 Klaus Weihrauch. Computable planar curves intersect in a computable point. *Computability*, 8(3, 4):399–415, 2019.
- 244 Linda Westrick. A note on the diamond operator. arXiv 2001.09372, 2020.
- 245 Linda Westrick. A note on the diamond operator. *Computability*, 10(2):107–110, 2021.

Participants

- Djamel – Eddine Amir
University Paris – Saclay – Orsay, FR
- Andrej Bauer
University of Ljubljana, SI
- Laurent Bienvenu
CNRS & Université de Bordeaux, Talence, FR
- Vasco Brattka
Universität der Bundeswehr – München, DE
- Merlin Carl
Europa – Universität – Flensburg, DE
- Raphaël Carroy
University of Torino, IT
- Vittorio Cipriani
TU Wien, AT
- Matthew de Brecht
Kyoto University, JP
- Damir D. Dzhafarov
University of Connecticut – Storrs, US
- Johanna N. Y. Franklin
Hofstra University – Hempstead, US
- Anton Freund
Universität Würzburg, DE
- Makoto Fujiwara
Tokyo University of Science, JP
- Giorgio Genovesi
University of Leeds, GB
- Kenneth Gill
La Salle University – Philadelphia, US
- Jun Le Goh
National University of Singapore, SG
- Peter Hertling
Universität der Bundeswehr – München, DE
- Jeffrey L. Hirst
Appalachian State University – Boone, US
- Rupert Hölzl
Universität der Bundeswehr – München, DE
- Akitoshi Kawamura
Kyoto University, JP
- Takayuki Kihara
Nagoya University, JP
- Ulrich Kohlenbach
TU Darmstadt, DE
- Davide Manca
Universität Würzburg, DE
- Alberto Marcone
University of Udine, IT
- Joseph S. Miller
University of Wisconsin – Madison, US
- Daniel Mourad
Nanjing University, CN
- Carl Mummert
Marshall University – Huntington, US
- Keng Meng Ng
Nanyang TU – Singapore, SG
- Arno Pauly
Swansea University, GB
- Cécilia Pradic
Swansea University, GB
- Emmanuel Rauzy
Universität der Bundeswehr – München, DE
- Matthias Schröder
TU Darmstadt, DE
- Paul Shafer
University of Leeds, GB
- Giovanni Soldà
Ghent University, BE
- Mariya I. Soskova
University of Wisconsin – Madison, US
- Ivan Titov
University of Bordeaux, FR
- Patrick Uftring
Universität Würzburg, DE
- Manlio Valenti
Swansea University, GB
- Java Darleen Villano
University of Connecticut – Storrs, US
- Andrea Volpi
University of Udine, IT
- Keita Yokoyama
Tohoku University, JP



Approximation Algorithms for Stochastic Optimization

Lisa Hellerstein^{*1}, Viswanath Nagarajan^{*2}, and Kevin Schewior^{*3}

1 NYU – New York, US. lisa.hellerstein@nyu.edu

2 University of Michigan – Ann Arbor, US. viswa@umich.edu

3 Universität Köln, DE. kschewior@gmail.com

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 25132 “Approximation Algorithms for Stochastic Optimization”. In this seminar, we gathered researchers from different areas interested in combinatorial optimization problems in which there is some stochasticity in the input. The focus was on approximation algorithms for computing adaptive or non-adaptive strategies to interact with this stochastic uncertainty as well as structural measures such as the adaptivity gap.

Seminar March 23–28, 2025 – <https://www.dagstuhl.de/25132>

2012 ACM Subject Classification Mathematics of computing → Approximation algorithms; Mathematics of computing → Probability and statistics

Keywords and phrases adaptivity, approximation algorithms, combinatorial optimization, stochastic optimization


Digital Object Identifier 10.4230/DagRep.15.3.159

1 Executive Summary

Kevin Schewior (Universität Köln, DE)

Lisa Hellerstein (NYU – New York, US)

Viswanath Nagarajan (University of Michigan – Ann Arbor, US)

License  Creative Commons BY 4.0 International license
© Kevin Schewior, Lisa Hellerstein, and Viswanath Nagarajan

Combinatorial optimization is a classic field, whose results are applied in numerous domains, including logistics, telecommunication, production scheduling, and health care. Many of the problems arising in this field are computationally hard (often NP-hard) to solve exactly. Therefore, approximation algorithms, i.e., efficient algorithms with provable performance guarantees, have been extensively investigated.

An aspect that is very relevant in practice, but which is not well understood, is uncertainty in the input. Stochastic models for uncertainty, where there is some probabilistic information about the uncertain parameters, are arguably the most common approach for algorithms under uncertainty. The main question that this seminar addresses is: Can we approximate (or even compute exactly) the best *strategy* to interact with stochastic uncertainty?

Such a strategy may adapt when uncertainty gets resolved. An additional, structural, question is the question for the *adaptivity gap*, i.e., to what degree adaptivity helps in the objective function. Existing and envisioned tools comprise (but are not limited to) probabilistic approaches (concentration inequalities, martingales, etc.), linear or convex optimization, rounding techniques, and dynamic programming.

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Approximation Algorithms for Stochastic Optimization, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 159–176

Editors: Lisa Hellerstein, Viswanath Nagarajan, and Kevin Schewior



Dagstuhl Reports

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

In this Dagstuhl Seminar, we gathered several researchers from different communities (approximation algorithms, algorithmic game theory, operations research, online algorithms, learning theory) that have been interested in these or related questions from different angles. The goal was to develop new techniques, to identify new models and directions, and to solve some of the many open problems related to the above question.

The specific problems considered in this seminar were stochastic function-evaluation problems, stochastic probing and selection problems, stochastic scheduling, online stochastic problems, and related problems. We also set out to investigate aspects such as sample-complexity bounds, approximate solutions with low regret, and relaxing the common independence assumption.

Organization of the seminar

We gathered 26 participants from different communities (approximation algorithms, algorithmic game theory, operations research, online algorithms, learning theory). To develop a common background and language, we scheduled four overview talks throughout the week:

- Sahil Singla. Towards Tackling Adaptivity and Correlations in Stochastic Optimization (one hour)
- Tonguç Ünlüyurt. A Review of the Sequential Testing Problem and its Extensions (30 minutes)
- Nicole Megow. Query Minimization for Stochastic Selection Problems (one hour)
- Thomas Kesselheim. Learning for Stochastic Optimization: Samples, Bandits, Contexts (one hour)

In addition, there were 14 talks lasting 30 minutes each, covering many different topics relevant to the workshop. There were two open-problem sessions, and we kept the late afternoons free for discussions. On the last day, we held a round-up session to discuss outcomes of the workshop. On one evening, we scheduled a joint session with the parallel seminar on Weihrauch complexity, to spark cross-disciplinary discussion.

Outcomes

The workshop was very well received. In the survey, the participants praised the research theme, the composition of the workshop, the inspiring atmosphere, and the talks. The only negative aspect mentioned several times was that the collaboration time could have been scheduled in the early (rather than late) afternoon and could have been more structured. It was mentioned once that there could have been more participants from industry.

It seems that we have identified an area that researchers from different communities are interested in and that greatly benefits from a workshop like this. Judging from the discussions started at the workshop, we expect that the workshop will have a longer-term impact in terms of results, research proposals, community building, follow-up workshops, etc. Specific directions that were discussed several times at the workshop but still seem underexplored were correlated random variables and sample-complexity bounds.

2 Table of Contents

Executive Summary

Kevin Schewior, Lisa Hellerstein, and Viswanath Nagarajan 159

Overview of Talks

Job selection and scheduling on unreliable machines <i>Alessandro Agnetis</i>	163
Subsampling Suffices for Adaptive Data Analysis <i>Guy Blanc</i>	163
Semi-Bandit Learning for Monotone Stochastic Optimization <i>Rohan Ghuge</i>	164
Online and Stochastic Matching <i>Nathaniel Grammel</i>	164
Unifying Pathwise and Expanding Search <i>Svenja M. Griesbach</i>	165
Learning from a Sample in Online Algorithms <i>Anupam Gupta</i>	165
Learning for Stochastic Optimization: Samples, Bandits, Contexts <i>Thomas Kesselheim</i>	166
Search games with predictions <i>Thomas Lidbetter</i>	166
Decomposing Probability Marginals Beyond Affine Requirements <i>Jannik Matuschke</i>	167
Query Minimization for Stochastic Set Selection Problems <i>Nicole Megow</i>	167
Testing whether a Partition Matroid has an Active Basis <i>Benedikt Plank</i>	168
Towards Tackling Adaptivity and Correlations in Stochastic Optimization <i>Sahil Singla</i>	168
Approximation Algorithms for Correlated Knapsack Orienteering <i>Chaitanya Swamy</i>	169
Stochastic Scheduling of Bernoulli Jobs through Policy Stratification <i>Marc Uetz</i>	169
Provably Accurate Shapley Value Estimation via Leverage Score Sampling <i>Teal Witter</i>	170
Approximating Optimal Binary Search Trees under Uncertainty <i>Sorrachai Yingchareonthawornchai</i>	170
Bayesian Probing on Graphs <i>Rudy Zhou</i>	171
A review of the sequential testing problem and its extensions <i>Tonguç Ünüyurt</i>	171

Open problems

Sequencing replicated jobs on parallel machines to maximize the probability of a full kit <i>Alessandro Agnetis</i>	172
Hardness of Correlated Prophet Inequality <i>Andrés Cristi</i>	172
Searching on a path with predictions <i>Thomas Lidbetter</i>	173
Stochastic Combinatorial Optimization under a Budget Constraint in Expectation <i>Kevin Schewior</i>	173
The deliberate idleness problem <i>Marc Uetz</i>	173
Stochastic Bin Packing with Overflow <i>Rudy Zhou</i>	174
Participants	176

3 Overview of Talks

3.1 Job selection and scheduling on unreliable machines

Alessandro Agnetis (University of Siena, IT)

License © Creative Commons BY 4.0 International license
© Alessandro Agnetis

Joint work of Alessandro Agnetis, Leus Roel, Emmeline Perneel, Ilara Salvadori, Kevin Schewior

We address the following problem, denoted as the Unreliable Job Selection and Sequencing Problem (UJSSP). Given a set J of jobs, a subset $S \subseteq J$ must be selected for processing on a single machine that is subject to failure. Each job j incurs a cost c_j if selected and yields a reward R_j upon successful completion. A job j is completed successfully only if the machine does not fail before or during its execution, with job-specific failure probabilities p_j . The objective is to determine an optimal subset and sequence of jobs to maximize the expected net profit. We review some known results for the case with $c_j = 0$ (i.e., all jobs are selected), for both the single-machine and the parallel-machine cases [1, 2]. We establish the computational complexity of UJSSP, proving its NP-hardness when job costs are not identical. The relationship of UJSSP with other submodular selection problems is discussed [3, 4], showing that the special cases in which all jobs have the same cost ($c_j = c$ for all j) or, respectively, the same failure probability ($p_j = p$ for all j) can be solved in polynomial time, while the case in which all jobs have the same reward remains open.

References

- 1 Agnetis, A., Detti, P., Pranzo, M. and Sodhi, M.S., Sequencing unreliable jobs on parallel machines, *Journal of Scheduling*, 12(1), 45–54, 2009.
- 2 Agnetis, A., Lidbetter, T., List scheduling is 0.8531-approximate for scheduling unreliable jobs on m parallel machines, *Operations Research Letters*, 48, 405–409, 2020.
- 3 W. Stadje. Selecting jobs for scheduling on a machine subject to failure, *Discrete Applied Mathematics*, 63(3), 257–265, 1995.
- 4 Olszewski, W. and Vohra, R., Simultaneous selection, *Discrete Applied Mathematics*, 200, 161–169, 2016.

3.2 Subsampling Suffices for Adaptive Data Analysis

Guy Blanc (Stanford University, US)

License © Creative Commons BY 4.0 International license
© Guy Blanc

Main reference Guy Blanc: “Subsampling Suffices for Adaptive Data Analysis”, CoRR, Vol. abs/2302.08661, 2023.
URL <https://doi.org/10.48550/ARXIV.2302.08661>

Ensuring that analyses performed on a dataset are representative of the entire population is one of the central problems in statistics. Most classical techniques assume that the dataset is independent of the analyst’s query and break down in the common setting where a dataset is reused for multiple, adaptively chosen, queries. This problem of *adaptive data analysis* was formalized in the seminal works of Dwork et al. (STOC, 2015) and Hardt and Ullman (FOCS, 2014).

We identify a remarkably simple set of assumptions under which the queries will continue to be representative even when chosen adaptively: The only requirements are that each query takes as input a random subsample and outputs few bits. This result shows that the

noise inherent in subsampling is sufficient to guarantee that query responses generalize. The simplicity of this subsampling-based framework allows it to model a variety of real-world scenarios not covered by prior work.

In addition to its simplicity, we demonstrate the utility of this framework by designing mechanisms for two foundational tasks, statistical queries and median finding. In particular, our mechanism for answering the broadly applicable class of statistical queries is both extremely simple and state of the art in many parameter regimes.

3.3 Semi-Bandit Learning for Monotone Stochastic Optimization

Rohan Ghuge (Georgia Institute of Technology – Atlanta, US)

License © Creative Commons BY 4.0 International license
© Rohan Ghuge

Joint work of Arpit Agarwal, Rohan Ghuge, Viswanath Nagarajan
Main reference Arpit Agarwal, Rohan Ghuge, Viswanath Nagarajan: “Semi-Bandit Learning for Monotone Stochastic Optimization”, in Proc. of the 65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024, pp. 1260–1274, IEEE, 2024.
URL <https://doi.org/10.1109/FOCS61266.2024.00083>

Stochastic optimization is a widely used approach for optimization under uncertainty, where uncertain input parameters are modeled by random variables. Exact or approximation algorithms have been obtained for several fundamental problems in this area. However, a significant limitation of this approach is that it requires full knowledge of the underlying probability distributions. Can we still get good algorithms if these distributions are unknown, and the algorithm needs to learn them through repeated interactions? In this talk, I will discuss a generic online learning algorithm that obtains optimal regret bounds relative to the best algorithms (under known distributions) for a large class of “monotone” stochastic problems. This class includes fundamental problems like single-resource revenue management, Pandora’s box, and stochastic knapsack. Notably, our online algorithm works in a semi-bandit setting, where in each period, the algorithm only observes samples from the random variables that were actually probed.

3.4 Online and Stochastic Matching

Nathaniel Grammel (University of Maryland – College Park, US)

License © Creative Commons BY 4.0 International license
© Nathaniel Grammel

Joint work of Nathaniel Grammel, Brian Brubach, Will Ma, Aravind Srinivasan
Main reference Nathaniel Grammel, Brian Brubach, Will Ma, Aravind Srinivasan: “Follow Your Star: New Frameworks for Online Stochastic Matching with Known and Unknown Patience”, in Proc. of the The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event, Proceedings of Machine Learning Research, Vol. 130, pp. 2872–2880, PMLR, 2021.
URL <http://proceedings.mlr.press/v130/grammel21a.html>

We consider variants of the classical graph matching problem that exhibit various forms of uncertainty. Online Bipartite Matching considers the problem of finding a maximum weight matching in a bipartite graph when the vertices of one side arrive one by one and, at each arrival, the algorithm must make an immediate and irrevocable decision about whether to match the vertex to one of its neighbors. Variants of the problem consider stochastic arrival models (e.g., arriving vertices are sampled IID from a known distribution).

Stochastic Matching (or Matching with Stochastic Edges) considers the standard graph matching problem in settings where the existence of each edge is (initially) unknown to the algorithm and only discovered after querying the edge. Variants of the problem consider the case where each vertex has a “patience”, indicating the maximum number of its incident edges that may be queried. Typically, this patience value is deterministic and known to the algorithm. Online Stochastic Matching combines both of these problems, so that vertices on one side of a bipartite graph arrive one by one, and in each step the algorithm may make multiple attempts to match the vertex until one succeeds, or until the patience is exhausted. We discuss results for some of these variants. A final variant considers the case of unknown (stochastic) patience, i.e. where the patience is drawn from some known distribution and is only learned by the algorithm once it becomes exhausted.

3.5 Unifying Pathwise and Expanding Search

Svenja M. Griesbach (University of Chile – Santiago de Chile, CL)

License © Creative Commons BY 4.0 International license

© Svenja M. Griesbach

Joint work of Svenja M. Griesbach, Felix Hommelsheim, Max Klimm, Kevin Schewior

We establish a framework unifying the pathwise search problem and the expanding search problem. We consider a graph $G = (V, E)$ with non-negative vertex weights and a designated start vertex s . Furthermore, each edge e is equipped with a non-negative cost and a discount factor $\alpha_e \in [0, 1]$ such that for the second and further traversals of this edge, its cost is multiplied by α_e . For a path that starts in s , the latency of a vertex is the total cost of that path until the vertex is visited for the first time. The goal is to find a path that starts in s and visits all vertices with positive weight such that the weighted sum of the latencies of all vertices is minimized. If $\alpha_e = 0$ for all $e \in E$, the problem corresponds to the expanding search problem, and if $\alpha_e = 1$ for all $e \in E$, it corresponds to the pathwise search problem. We give a polynomial time algorithm that yields a constant approximation factor for all choices of $\alpha \in [0, 1]^{|E|}$. For $\alpha = 0$ and $\alpha = 1$, our factor attains the same ratio as the so far best factors for expanding and pathwise search, respectively.

3.6 Learning from a Sample in Online Algorithms

Anupam Gupta (New York University, US)

License © Creative Commons BY 4.0 International license

© Anupam Gupta

Joint work of C. J. Argue, Alan M. Frieze, Anupam Gupta, Christopher Seiler


Main reference C. J. Argue, Alan M. Frieze, Anupam Gupta, Christopher Seiler: “Learning from a Sample in Online Algorithms”, in Proc. of the Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 – December 9, 2022, 2022.

URL http://papers.nips.cc/paper_files/paper/2022/hash/5a093120ff4776b4f0dc452e3e3b6652-Abstract-Conference.html

While analyzing algorithms in the worst-case has long served us well, recent years have seen an exciting surge in analyzing algorithms in models that go beyond the worst case. We consider the classical problem of load-balancing, where jobs arrive online and must be assigned to collections of machines to minimize the maximum load. Can we get results better than the adversarial setting if we have a small sample of the upcoming data? The results in this talk are based on work with C.J. Argue, Alan Frieze, and Chris Seiler.

3.7 Learning for Stochastic Optimization: Samples, Bandits, Contexts

Thomas Kesselheim (Universität Bonn, DE)

License  Creative Commons BY 4.0 International license
© Thomas Kesselheim

Commonly, in stochastic optimization, one assumes to know the probability distribution(s) the input is coming from. This is often motivated by the availability of historic data. In this talk, we survey different recent results formalizing this aspect and attempting to understand how one can learn the distribution well enough. For example, we might get a number of samples from the distribution. Can we still show that the optimal policy on the empirical distribution has a good performance? And how should we behave in repeated settings, where we get one sample from the distribution per day?

3.8 Search games with predictions

Thomas Lidbetter (Rutgers University – Newark, US)

License  Creative Commons BY 4.0 International license
© Thomas Lidbetter

Joint work of Spyros Angelopoulos, Thomas Lidbetter, Konstantinos Panagiotou
Main reference Spyros Angelopoulos, Thomas Lidbetter, Konstantinos Panagiotou: “Search Games with Predictions”, CoRR, Vol. abs/2401.01149, 2024.
URL <https://doi.org/10.48550/ARXIV.2401.01149>

We introduce the study of search games between a mobile Searcher and an immobile Hider in a new setting in which the Searcher has some potentially erroneous information, i.e., a prediction on the Hider’s position. The objective is to establish tight tradeoffs between the consistency of a search strategy (i.e., its worst case expected payoff assuming the prediction is correct) and its robustness (i.e., the worst case expected payoff with no assumptions on the quality of the prediction). Our study is the first to address the full power of mixed (randomized) strategies; previous work focused only on deterministic strategies, or relied on stochastic assumptions that do not guarantee worst-case robustness in adversarial situations. We give Pareto-optimal strategies for three fundamental problems, namely searching in discrete locations, searching with stochastic overlook, and searching in the infinite line. As part of our contribution, we provide a novel framework for proving optimal tradeoffs in search games which is applicable, more broadly, to any two-person zero-sum games in learning-augmented settings.

3.9 Decomposing Probability Marginals Beyond Affine Requirements

Jannik Matuschke (KU Leuven, BE)

License © Creative Commons BY 4.0 International license
© Jannik Matuschke

Main reference Jannik Matuschke: “Decomposing Probability Marginals Beyond Affine Requirements”, in Proc. of the Integer Programming and Combinatorial Optimization – 25th International Conference, IPCO 2024, Wrocław, Poland, July 3-5, 2024, Proceedings, Lecture Notes in Computer Science, Vol. 14679, pp. 309–322, Springer, 2024.

URL https://doi.org/10.1007/978-3-031-59835-7_23

Consider the triplet (E, \mathcal{P}, π) , where E is a finite ground set, $\mathcal{P} \subseteq 2^E$ is a collection of subsets of E and $\pi : \mathcal{P} \rightarrow [0, 1]$ is a requirement function. Given a vector of marginals $\rho \in [0, 1]^E$, our goal is to find a distribution for a random subset $S \subseteq E$ such that $\Pr[e \in S] = \rho_e$ for all $e \in E$ and $\Pr[P \cap S \neq \emptyset] \geq \pi_P$ for all $P \in \mathcal{P}$, or to determine that no such distribution exists.

Generalizing results of Dahan, Amin, and Jaillet, we devise a generic decomposition algorithm that solves the above problem when provided with a suitable sequence of admissible support candidates (ASCs). We show how to construct such ASCs for numerous settings, including supermodular requirements, Hoffman-Schwartz-type lattice polyhedra, and abstract networks where π fulfils a conservation law. The resulting algorithm can be carried out efficiently when \mathcal{P} and π can be accessed via appropriate oracles. For any system allowing the construction of ASCs, our results imply a simple polyhedral description of the set of marginal vectors for which the decomposition problem is feasible. Finally, we characterize balanced hypergraphs as the systems (E, \mathcal{P}) that allow the perfect decomposition of any marginal vector $\rho \in [0, 1]^E$, i.e., where we can always find a distribution reaching the highest attainable probability $\Pr[P \cap S \neq \emptyset] = \min \sum_{e \in P} \rho_e, 1$ for all $P \in \mathcal{P}$.

3.10 Query Minimization for Stochastic Set Selection Problems

Nicole Megow (Universität Bremen, DE)

License © Creative Commons BY 4.0 International license
© Nicole Megow

Main reference Evripidis Bampis, Christoph Dürr, Thomas Erlebach, Murilo Santos de Lima, Nicole Megow, Jens Schlöter: “Orienting (Hyper)graphs Under Explorable Stochastic Uncertainty”, in Proc. of the 29th Annual European Symposium on Algorithms, ESA 2021, September 6-8, 2021, Lisbon, Portugal (Virtual Conference), LIPIcs, Vol. 204, pp. 10:1–10:18, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.

URL <https://doi.org/10.4230/LIPICS.ESA.2021.10>

Main reference Nicole Megow, Jens Schlöter: “Set Selection Under Explorable Stochastic Uncertainty via Covering Techniques”, in Proc. of the Integer Programming and Combinatorial Optimization – 24th International Conference, IPCO 2023, Madison, WI, USA, June 21-23, 2023, Proceedings, Lecture Notes in Computer Science, Vol. 13904, pp. 319–333, Springer, 2023.

URL https://doi.org/10.1007/978-3-031-32726-1_23

I will give an overview of basic set selection problems in a model where querying uncertain data incurs a cost. Given subsets of uncertain values, we study two tasks: identifying the minimum element in each set and selecting the subset of minimum total value while querying as few values as possible. These problems fall under the umbrella of explorable uncertainty. In the adversarial setting, strong lower bounds on query complexity extend to a wide range of classical problems such as knapsack, matchings, and linear programming. We then introduce a stochastic variant, where each weight is drawn independently from a known distribution, and present algorithms that, in expectation, beat these adversarial bounds. Our approach builds on a careful analysis of the underlying offline problems, exploiting connections to vertex covers and LP formulations. Finally, I will outline further research directions involving parallelization, robustification, and other extensions.

3.11 Testing whether a Partition Matroid has an Active Basis

Benedikt Plank (Berlin, DE)


License  Creative Commons BY 4.0 International license
© Benedikt Plank

Joint work of Lisa Hellerstein, Benedikt Plank, Kevin Schewior

We consider the following Stochastic Boolean Function Evaluation problem, which is closely related to several problems from the literature. A matroid \mathcal{M} (in compact representation) on ground set E is given, and each element $i \in E$ is active independently with known probability $p_i \in (0, 1)$. The elements can be queried, upon which it is revealed whether the respective element is active or not. The goal is to find an adaptive querying strategy for determining whether there is a basis of \mathcal{M} in which all elements are active, with the objective of minimizing the expected number of queries. When \mathcal{M} is a uniform matroid, this is the problem of evaluating a k -of- n function, first studied in the 1970s. This problem is well-understood, and has an optimal adaptive strategy that can be computed in polynomial time. Interestingly, already when \mathcal{M} is a partition matroid, we show that the standard approaches fail to give even a constant-factor approximation. Our main result is a randomized polynomial-time constant-factor approximation algorithm for this problem. Our algorithm adaptively interleaves solutions to several instances of a novel type of stochastic querying problem, with a constraint on the expected cost. We believe that this problem is of independent interest and that several of our techniques have the potential for more general applications.

3.12 Towards Tackling Adaptivity and Correlations in Stochastic Optimization

Sahil Singla (Georgia Institute of Technology – Atlanta, US)

License  Creative Commons BY 4.0 International license
© Sahil Singla

In this survey talk, we will consider discrete optimization problems where the inputs include probability distributions, and the goal is to maximize the expected reward. Two key benchmarks for these problems are the hindsight (offline) optimum and the optimal (online) policy. A central challenge, regardless of the chosen benchmark, is that optimal algorithms often adapt to the realizations of random variables. This adaptivity can lead to decision trees of exponential size, making the problem computationally intractable. We will explore techniques that focus on non-adaptive algorithms, which offer simpler and more efficient solutions, with only a small loss in performance. We will also discuss advances in stochastic discrete optimization models that incorporate correlations, while maintaining tractability in algorithm design.

3.13 Approximation Algorithms for Correlated Knapsack Orienteering

Chaitanya Swamy (University of Waterloo, CA)

License © Creative Commons BY 4.0 International license
© Chaitanya Swamy

Joint work of Chaitanya Swamy, Davis Alemán Espinosa

Main reference David Alemán Espinosa, Chaitanya Swamy: “Approximation Algorithms for Correlated Knapsack Orienteering”, in Proc. of the Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2024, August 28-30, 2024, London School of Economics, London, UK, LIPIcs, Vol. 317, pp. 29:1–29:24, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.

URL <https://doi.org/10.4230/LIPICS.APPROX/RANDOM.2024.29>

We consider the *correlated knapsack orienteering* (CSKO) problem: we are given a travel budget B , processing-time budget W , finite metric space (V, d) with root $\rho \in V$, where each vertex is associated with a job with possibly correlated random size and random reward that become known only when the job completes. Random variables are independent across different vertices. The goal is to compute a ρ -rooted path of length at most B , in a possibly adaptive fashion, that maximizes the reward collected from jobs that are processed by time W . To our knowledge, CSKO has not been considered before, though prior work has considered the uncorrelated problem, *stochastic knapsack orienteering*, and *correlated orienteering*, which features only one budget constraint on the *sum* of travel-time and processing-times.

We show that the *adaptivity gap* of CSKO is not a constant, and is at least $\Omega(\max\{\sqrt{\log B}, \sqrt{\log \log W}\})$. Complementing this, we devise *non-adaptive* algorithms that obtain: (a) $O(\log \log W)$ -approximation in quasi-polytime; and (b) $O(\log W)$ -approximation in polytime. We obtain similar guarantees for CSKO with cancellations, wherein a job can be cancelled before its completion time, foregoing its reward. We also consider the special case of CSKO, wherein job sizes are weighted Bernoulli distributions, and more generally where the distributions are supported on at most two points (2-CSKO). Although weighted Bernoulli distributions suffice to yield an $\Omega(\sqrt{\log \log B})$ adaptivity-gap lower bound for (uncorrelated) *stochastic orienteering*, we show that they are easy instances for CSKO. We develop non-adaptive algorithms that achieve $O(1)$ -approximation in polytime for weighted Bernoulli distributions, and in $(n + \log B)^{O(\log W)}$ -time for the more general case of 2-CSKO.

3.14 Stochastic Scheduling of Bernoulli Jobs through Policy Stratification

Marc Uetz (University of Twente – Enschede, NL)

License © Creative Commons BY 4.0 International license
© Marc Uetz

Joint work of Antonios Antoniadis, Ruben Hoeksma, Kevin Schewior, Marc Uetz

Main reference Antonios Antoniadis, Ruben Hoeksma, Kevin Schewior, Marc Uetz: “Stochastic scheduling with Bernoulli-type jobs through policy stratification”, CoRR, Vol. abs/2505.03349, 2025.

URL <https://doi.org/10.48550/ARXIV.2505.03349>

This talk addresses the problem of computing a scheduling policy that minimizes the total expected completion time of a set of N jobs with stochastic processing times on m parallel identical machines. When all processing times follow Bernoulli-type distributions, Gupta et al. (SODA '23) exhibited approximation algorithms with an approximation guarantee $\tilde{O}(\sqrt{m})$, where m is the number of machines and $\tilde{O}(\cdot)$ suppresses polylogarithmic factors in N , improving upon an earlier $O(m)$ approximation by Eberle et al. (OR Letters '19) for a special case. The present paper shows that, quite unexpectedly, the problem with

Bernoulli-type jobs admits a PTAS whenever the number of different job-size parameters is bounded by a constant. The result is based on a series of transformations of an optimal scheduling policy to a “stratified” policy that makes scheduling decisions at specific points in time only, while losing only a negligible factor in expected cost. An optimal stratified policy is computed using dynamic programming. Two technical issues are solved, namely (i) to ensure that, with at most a slight delay, the stratified policy has an information advantage over the optimal policy, allowing it to simulate its decisions, and (ii) to ensure that the delays do not accumulate, thus solving the trade-off between the complexity of the scheduling policy and its expected cost. Our results also imply a quasi-polynomial $\tilde{O}(\log N)$ -approximation for the case with an arbitrary number of job sizes.

3.15 Provably Accurate Shapley Value Estimation via Leverage Score Sampling

Teal Witter (New York University, US)

License © Creative Commons BY 4.0 International license
© Teal Witter

Joint work of Christopher Musco, R. Teal Witter

Main reference Christopher Musco, R. Teal Witter: “Provably Accurate Shapley Value Estimation via Leverage Score Sampling”, in Proc. of the The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025, OpenReview.net, 2025.

URL <https://openreview.net/forum?id=wg3rBImn3O>

Originally introduced in game theory, Shapley values have emerged as a central tool in explainable machine learning, where they are used to attribute model predictions to specific input features. However, computing Shapley values exactly is expensive: for a general model with n features, $O(2^n)$ model evaluations are necessary. To address this issue, approximation algorithms are widely used. One of the most popular is the Kernel SHAP algorithm, which is model agnostic and remarkably effective in practice. However, to the best of our knowledge, Kernel SHAP has no strong non-asymptotic complexity guarantees. We address this issue by introducing Leverage SHAP, a light-weight modification of Kernel SHAP that provides provably accurate Shapley value estimates with just $O(n \log n)$ model evaluations. Our approach takes advantage of a connection between Shapley value estimation and agnostic active learning by employing leverage score sampling, a powerful regression tool. Beyond theoretical guarantees, we show that Leverage SHAP consistently outperforms even the highly optimized implementation of Kernel SHAP available in the ubiquitous SHAP library [Lundberg & Lee, 2017].

3.16 Approximating Optimal Binary Search Trees under Uncertainty

Sorrachai Yingchareonthawornchai (ETH Zürich, CH)

License © Creative Commons BY 4.0 International license
© Sorrachai Yingchareonthawornchai

Joint work of Parinya Chalermsook, Zahra Hadizadeh, Wanchote Jiamjitrak, Sorrachai Yingchareonthawornchai

Constructing an optimal binary search tree (BST) has long been a foundational problem in data structures. Given a fixed, known probability distribution over keys, Knuth’s seminal result (1971) provides an efficient method for computing the optimal static BST. Later, Mehlhorn (1975) showed that a near-optimal BST can be approximated in linear time.

We initiate the study of the robust optimization variant of the classical BST problem by considering settings where the underlying distribution is uncertain. Instead of a single known distribution, we are given k different distributions $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_k$. The goal is to construct a single BST T that performs well across all of them – minimizing the worst-case expected access cost.

In this talk, I will give a simple 1.5-approximation algorithm for $k = 2$, which can be generalized to a $0.804k$ -approximation for arbitrary k . This improvement is achieved by carefully combining two distinct 2-approximation algorithms, leveraging their strengths to refine the approximation ratio. I will also show a hardness result under the Unique Games Conjecture: when k is large, computing an optimal BST is NP-hard, even when each distribution has support of size at most two.

3.17 Bayesian Probing on Graphs

Rudy Zhou (Carnegie Mellon University – Pittsburgh, US)

License © Creative Commons BY 4.0 International license

© Rudy Zhou

Joint work of Anupam Gupta, Benjamin Moseley, Rudy Zhou

We introduce a stochastic probing problem with correlated items, which we call Bayesian probing, where the correlations are modeled by an underlying graph. In our model, each vertex has a known probability of being active or inactive. Each item is an edge in the graph, and its distribution is the product of its two endpoints. The goal is to adaptively probe items/edges subject to a knapsack constraint to maximize the expected total reward obtained from all probed edges. This problem is a special case of stochastic knapsack with correlations across items and Bayesian active search problems considered in machine learning.

The distinguishing feature of Bayesian probing is that the probing an edge reveals the outcome of both of its endpoints, which induces a Bayesian update on the expected reward of all other incident edges. We design approximation algorithms computing policies that are either fully non-adaptive, or they make a single Bayesian update after using half of the knapsack budget. Our approximation ratios depend on natural graph parameters of the underlying correlation graph.

3.18 A review of the sequential testing problem and its extensions

Tonguç Ünlüyurt (Sabanci University – Istanbul, TR)

License © Creative Commons BY 4.0 International license


© Tonguç Ünlüyurt

In the sequential testing problem, the goal is to evaluate a Boolean (or discrete) function with the minimum expected cost where the values of the variables can be learned by paying a cost. The variables take values independent of each other with known probabilities. The problem has been studied in different domains for various applications. In this talk, we concentrate on works that have been published in the last 20 years and provide a general review of the results that have been obtained for different special cases and extensions of the problem. We also provide insights to explore potential research areas for future research.

4 Open problems

4.1 Sequencing replicated jobs on parallel machines to maximize the probability of a full kit

Alessandro Agnetis (University of Siena, IT)

License  Creative Commons BY 4.0 International license
© Alessandro Agnetis

Consider n job types, m machines and m copies of each job type. When processed by a machine, a job of type j is successfully carried out with probability π_j , while with probability $(1 - \pi_j)$ it fails. In the latter case, the machine halts and cannot process all subsequently scheduled jobs. The problem is to decide how to sequence the n job copies on each of the m machines, in order to maximize the probability of having at least one “full kit” of jobs, i.e., at least one copy of each job successfully carried out.

For $m = 2$, the optimal solution is simply found by arbitrarily sequencing the n job copies on the first machine, and by sequencing them in the reverse order on the second machine. Another easily solvable case is when $n = 2$. In this case the optimal solution can be found in $O(\log m)$. These results can be found in [1].


As far as I know, the complexity of the problem is open for $m \geq 3$. Even the case $m = n$, or the case with identical probabilities do not seem obvious...

References

- 1 Agnetis, A., Benini, M., Detti, P., Hermans, B., Pranzo, M., Replication and sequencing of unreliable jobs on parallel machines, *Computers and Operations Research*, 139, 105634, 2022, doi:10.1016/j.cor.2021.105634.

4.2 Hardness of Correlated Prophet Inequality

Andrés Cristi (EPFL – Lausanne, CH)

License  Creative Commons BY 4.0 International license
© Andrés Cristi

In the prophet inequality we observe a sequence of n random variables one by one and we want to decide when to stop. The reward of a online stopping policy is the variable it stops with. When the random variables are independent and the distributions are known, the optimal stopping policy can be computed in polynomial time via backward induction. A large body of literature is dedicated to bound the ratio between the expected reward of the optimal online policy and the expected offline optimum, under many variants of the problem. When the variables are correlated, no constant approximation with respect to the offline optimum is possible. Moreover, when the variables are correlated, the computation of the optimal online policy becomes challenging, because a naive backward induction algorithm would need to “remember” at each time all previous realizations. Here I propose two models to study the computational complexity of correlated optimal stopping, one easy and one apparently hard:

- (1) We are given a set of m scenarios. Each scenario is a deterministic sequence of n values. A realization of the n random variables is drawn as a uniformly chosen scenario. We can compute the optimal online algorithm via backward induction, by noticing that at any given time, the observed values must agree with at least one of the scenarios. Therefore, we can encode past observations by just choosing one scenario that is compatible with them. Thus we can compute the optimal online policy in time $\text{poly}(n, m)$.

(2) We are given a set of m scenarios, but now a scenario is a sequence of n distributions, all supported in a set of k different values. A realization of the n random variables is drawn by first choosing uniformly one scenario, and then drawing each variable independently according to the n distributions of the chosen scenario. Can we compute the optimal online policy in time $\text{poly}(m, n, k)$? If not, can we approximate it?

4.3 Searching on a path with predictions


Thomas Lidbetter (*Rutgers University – Newark, US*)

License  Creative Commons BY 4.0 International license
© Thomas Lidbetter

A target is located at one of the two ends of a discrete path according to a known probability distribution. A Searcher begins at a root node O at time 0. In each time step, she moves to an adjacent node. Upon reaching a node, she receives a signal that points either left or right. (We assume there is no signal at O at time 0.) With probability $p \geq 1/2$, the signal points towards the target, otherwise it points in the opposite direction. We wish to find a policy to minimize the expected time to reach the target. Attached is a solution for a path of length 3. Is there a closed form policy in general?

4.4 Stochastic Combinatorial Optimization under a Budget Constraint in Expectation


Kevin Schewior (*Universität Köln, DE*)

License  Creative Commons BY 4.0 International license
© Kevin Schewior
Joint work of Lisa Hellerstein, Benedikt Plank, Kevin Schewior

Motivated by Benedikt Plank's talk, I propose to relax stochastic combinatorial optimization problems in which there is a hard budget constraint by considering them under a budget constraint *in expectation*. One can then ask (i) how to efficiently compute or approximate the best strategy and (ii) by what factor the objective-function value changes in the worst case. Such problems may be interesting by itself, but Benedikt Plank's talk has shown another motivation. For the k -of- n SFBE problem considered in that talk the answer to (ii) was non-constant. What else can we say, e.g., for the ProbeMax problem?

4.5 The deliberate idleness problem

Marc Uetz (*University of Twente – Enschede, NL*)

License  Creative Commons BY 4.0 International license
© Marc Uetz

The *deliberate idleness problem* is a problem in stochastic machine scheduling. In stochastic machine scheduling, we are concerned with the question how to optimally schedule n jobs with stochastic processing requirements on m machines. More specifically, the processing times follows distributions $p_j \sim X_j$, $j = 1 \dots, n$. The jobs are nonpreemptive, all available

at time 0, and have to be scheduled on m parallel, identical machines. Each machine can only do one job at a time, and a job can go on any of the m machines. Moreover, each job has a weight w_j , and we want to find a scheduling policy that minimizes the expected value of the weighted sum of completion times, $E[\sum_j w_j C_j]$. An instance consists of the input of jobs (w_j, X_j) , $j = 1, \dots, n$, and the encoding of the number of machines m .

The solution to such a problem is not a schedule, but a scheduling *policy* which tells us, at any point in time t (typically when a machine falls idle, but possibly also at other points in time), which job(s) to schedule next. This decision may depend on the input of the problem, and the state of the system at time t . The latter is given by time t , the set of jobs already completed, the set of jobs currently running together with their conditional distribution of remaining processing time, and the set of jobs not yet started.

Question: Assume $m \geq 3$ machines, and assume that all jobs follow an exponential distribution, $p_j \sim \exp(\lambda_j)$, that is, the processing times are memoryless. Does there always exist an optimal policy that avoids deliberate idleness? That is, as long as there are unprocessed jobs, it would never leave a machine deliberately idle. Some background information follows.

- For arbitrary $p_j \sim X_j$ there are examples showing that deliberate idleness can be necessary, even on $m = 2$ machines. See Uetz: When Greediness Fails: Examples from Stochastic Scheduling, OR Lett. 31, 2003, 413-419. (Franziska Eberle might have an example even when all $w_j = 1$).
- For $m = 2$ machines and $p_j \sim \exp(\lambda_j)$ an optimal policy always exists that avoids deliberate idleness. This is not totally trivial, but not too difficult either, using an exchange argument (and induction).
- WSEPT (greedily schedule jobs in order of ratios w_j/Ep_j) has a performance guarantee of $2 - 1/m$, whenever the distribution have coefficient of variation ≤ 1 , including exponential distributions. See Möhring, Schulz, Uetz: Approximation in stochastic scheduling: the power of LP-based priority policies, J. ACM 46 (1999), 924-942. For a more recent improvement of this upper bound to $4/3$, see Jäger, Skutella: Generalizing the Kawaguchi-Kyan bound to stochastic parallel machine scheduling, 35th STACS, LIPIcs no. 43, vol. 96, 2018
- When all $w_j = 1$, the problem is solved optimally by the SEPT rule, greedily schedule jobs with shortest expected processing time first. See Bruno, Downey, Frederickson: Sequencing Tasks with Exponential Service Times to Minimize the Expected Flow Time or Makespan, J. ACM 28 (1981), 100-113.

4.6 Stochastic Bin Packing with Overflow

Rudy Zhou (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 4.0 International license
© Rudy Zhou

Joint work of Sebastian Perez-Salazar, Mohit Singh, Alejandro Toriello, Rudy Zhou

Main reference Sebastian Perez-Salazar, Mohit Singh, Alejandro Toriello: “Adaptive Bin Packing with Overflow”, Math. Oper. Res., Vol. 47(4), pp. 3317–3356, 2022.

URL <https://doi.org/10.1287/MOOR.2021.1239>

In the stochastic bin packing problem, we must adaptively pack items with random sizes into unit-sized bins. Each item size is independent with known distribution but unknown realized value. Because of this, we may place an item into a bin such that its realized size

overflows the bin capacity. Overflowing a bin incurs an additive penalty. Thus, our objective is to pack all items to minimize the expected number of bins opened plus the penalties for any overflowed bins.

This problem was introduced by Sebastian Perez-Salazar, Mohit Singh, and Alejandro Toriello (<https://arxiv.org/abs/2007.11532>). Among other things, they prove that in the online setting (items arrive one-by-one and must be packed irrevocably upon arrival), one can achieve a $O(1)$ -approximation if all items are drawn i.i.d. from a known distribution and a $O(\log C)$ approximation if all items are exponentially distributed, where C is a function of the parameters of the exponentials.

It would be interesting to give approximation algorithms for more general distributions or consider the case where we do not know the item size distributions.

Participants

- Alessandro Agnetis
University of Siena, IT
- Guy Blanc
Stanford University, US
- Andrés Cristi
EPFL – Lausanne, CH
- Franziska Eberle
TU Berlin, DE
- Rohan Ghuge
Georgia Institute of Technology – Atlanta, US
- Nathaniel Grammel
University of Maryland – College Park, US
- Svenja M. Griesbach
University of Chile – Santiago de Chile, CL
- Anupam Gupta
New York University, US
- Lisa Hellerstein
NYU – New York, US
- Thomas Kesselheim
Universität Bonn, DE
- Rebecca Lehming
Universität Bonn, DE
- Thomas Lidbetter
Rutgers University – Newark, US
- Naifeng Liu
Universität Mannheim, DE
- Jannik Matuschke
KU Leuven, BE
- Nicole Megow
Universität Bremen, DE
- Viswanath Nagarajan
University of Michigan – Ann Arbor, US
- Benedikt Plank
Berlin, DE
- Kevin Schewior
Universität Köln, DE
- Daniel Schmand
Universität Bremen, DE
- Sahil Singla
Georgia Institute of Technology – Atlanta, US
- Chaitanya Swamy
University of Waterloo, CA
- Tonguç Ünliyurt
Sabanci University – Istanbul, TR
- Marc Uetz
University of Twente – Enschede, NL
- Teal Witter
New York University, US
- Sorrachai Yingchareonthawornchai
ETH Zürich, CH
- Rudy Zhou
Carnegie Mellon University – Pittsburgh, US



Categories for Automata and Language Theory

Achim Blumensath^{*1}, Mikołaj Bojańczyk^{*2}, Bartek Klin^{*3}, and Daniela Petrişan^{*4}

1 Masaryk University – Brno, CZ. blumens@fi.muni.cz

2 University of Warsaw, PL. bojan@mimuw.edu.pl

3 University of Oxford, GB. bartek.klin@cs.ox.ac.uk

4 Université Paris Cité, FR. Petrisan@irif.fr

Abstract

Categorical methods have a long history in automata and language theory, but a coherent theory has started to emerge only in recent years. The abstract viewpoint of category theory can provide a unifying perspective on various forms of automata; it can make it easier to bootstrap a theory in a new setting; and it provides conceptual clarity regarding which aspects and properties are fundamental and which are only coincidental.

Due to being in its early stages, the field is currently still divided into several different communities with little connections between them. One of the central aims of the Dagstuhl Seminar “Categories for Automata and Language Theory” (25141) was to enhance communication between automata theory and category theory communities. To this end, the seminar brought together researchers from both areas and included introductory tutorials designed to provide common ground and help participants better understand each other’s approach and terminology.

The following key topics were explored during the seminar:

- Categorical unification of language theory, either via the theory of monads, or via the category of MSO-transductions and their composition;
- Coalgebraic methods and their applications to automata theory, to infinite trace semantics and connection to bisimulation-invariant fragments of logics;
- Functorial automata and generic algorithms therein;
- Fibrational and monoidal perspectives on language theory;
- Higher-order automata and profinite lambda-calculus.

Seminar March 30 – April 4, 2025 – <https://www.dagstuhl.de/25141>

2012 ACM Subject Classification Theory of computation → Algebraic language theory; Theory of computation → Automata extensions

Keywords and phrases categorical automata theory, automata theory, category theory, monads

Digital Object Identifier 10.4230/DagRep.15.3.177

1 Executive Summary

Achim Blumensath (Masaryk University – Brno, CZ)

Mikołaj Bojańczyk (University of Warsaw, PL)

Bartek Klin (University of Oxford, GB)

Daniela Petrişan (Université Paris Cité, FR)

License  Creative Commons BY 4.0 International license

© Achim Blumensath, Mikołaj Bojańczyk, Bartek Klin, and Daniela Petrişan

Categorical methods have a long history in automata and language theory (see, e.g., [1, 2, 3, 4, 5] for early examples), but only in recent years a coherent theory has started to emerge.

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Categories for Automata and Language Theory, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 177–200

Editors: Achim Blumensath, Mikołaj Bojańczyk, Bartek Klin, and Daniela Petrişan



Dagstuhl Reports

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

As an indication of the increasing popularity, note that papers on categories and automata have regularly appeared in the last three years, both at ICALP(B) and at LICS. These are the top conferences in the field of Track B of Theoretical Computer Science, but a similar pattern holds across other conferences in this field. There are several reasons why these abstract approaches have become popular in automata theory.

- They provide a unifying perspective on various forms of automata. For example, minimisation and learning algorithms for deterministic automata, weighted automata and sequential transducers can be seen as instances of a generic algorithm given at an abstract category-theoretic level [6].
- They make it easier to bootstrap a theory in a new setting. For instance, one of the main motivations of the monadic approach to recognisability [7] was to extend the existing algebraic theories to infinite trees [8].
- They provide conceptual clarity regarding which aspects and properties are fundamental and which are only coincidental. For example, the semiring based formalism to formal languages treats addition and multiplication symmetrically, while a more general approach [9] reveals that multiplication is specific to the shapes of the objects in question while addition is universal and related to the power-set operation.

The field is still in its early stages, and as a result it is still divided into several different communities with little connections between them. The purpose of this Dagstuhl Seminar was to connect researchers active in these communities; to make them aware of the work of other groups; to initiate collaborations; and to discuss recent developments and possible ways to go forward.

Organization of the seminar

The organisers designed the schedule to strike a balance between survey talks, focused presentations, and free time for informal discussions. Participants were encouraged to share their work and highlight recent advances in their respective fields. Given the diverse backgrounds of the attendees, the organisers invited several participants to deliver extended tutorial-style talks on specific topics to help bridge disciplinary gaps. Five such talks were presented, as listed below.

1. Achim Blumensath: *Monads and Recognisability*. This tutorial aims to provide an accessible introduction and a broad overview of a recently developed framework for algebraic language theory, which is based on monads and Eilenberg–Moore algebras.
2. Mikołaj Bojańczyk: *The Composition Method*. This talk explores the connection between algebra and logic through a categorical lens intended to support generalisations beyond word languages, but from a viewpoint dual to the one adopted in the first tutorial: logic serves as the primary notion, and algebraic structures are derived from it. The central topic of interest is the category of MSO-transductions.
3. Paweł Sobocinski: *String Diagrams*. The talk introduced string diagrams – a mathematical notation rooted in (monoidal) category theory – and its applications through computer science. In particular, it discussed monoidal automata and their languages of string diagrams.
4. Yde Venema: *Coalgebra*. The talk offered a gentle introduction to universal coalgebra as a broad categorical framework for modeling state-based evolving systems. It discussed coinduction, behavioral equivalence and bisimilarity.
5. Noam Zeilberger: *A tutorial on (generalized) fibrations for logic, automata and language theory*. The talk provided an introduction to certain fibrational concepts from category theory and their relevance to addressing “lifting problems” in logic, automata, and language theory.

Apart from the tutorials, 30 other participants delivered short presentations on recent work related to the topics listed above. Two additional sessions were set aside to allow time for informal discussions and interactions.

Conclusions

The organizers considered the seminar to be a success. Most participants reported gaining new insights from other areas and many expressed interest in applying these ideas to advance their own research. Among the participants who filled in the survey, more than half evaluated the scientific quality of the seminar as outstanding. We have striven to integrate junior researchers and many of them gave talks. This came at the expense of having less time dedicated to personal interactions.

References

- 1 P. L. F. Rudolf E. Kalman and M. A. Arbib, *Topics in Mathematical System Theory*, McGraw-Hill, 1969.
- 2 S. Eilenberg, *Automata, Languages, and Machines: volume A*, Pure and applied mathematics, Academic Press, 1974.
- 3 M. A. Arbib and E. G. Manes, *A categorist's view of automata and systems*, in Category Theory Applied to Computation and Control, Proceedings of the First International Symposium, San Francisco, CA, USA, February 25–26, 1974, Proceedings, E. G. Manes, ed., vol. 25 of Lecture Notes in Computer Science, Springer, 1974, pp. 51–64.
- 4 H. Ehrig, K. Kiermeier, H. Kreowski, and W. Kühnel, *Universal theory of automata. A categorical approach*, Teubner Studienbücher, Teubner, 1974.
- 5 B. Tilson, *Categories as algebra: An essential ingredient in the theory of monoids*, Journal of Pure and Applied Algebra, 48 (1987), pp. 83–198.
- 6 T. Colcombet, D. Petrişan, and R. Stabile, *Learning automata and transducers: A categorical approach*, in 29th EACSL Annual Conference on Computer Science Logic, CSL 2021, January 25–28, 2021, Ljubljana, Slovenia (Virtual Conference), C. Baier and J. Goubault-Larrecq, eds., vol. 183 of LIPIcs, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, pp. 15:1–15:17.
- 7 M. Bojańczyk, *Recognisable languages over monads*. Unpublished note, arXiv:1502.04898v1.
- 8 *Regular Tree Algebras*, Logical Methods in Computer Science, 16 (2020), pp. 16:1–16:25.
- 9 *The Power-Set Operation for Tree Algebras*, Logical Methods in Computer Science, (to appear).

2 Table of Contents

Executive Summary

Achim Blumensath, Miłojaj Bojańczyk, Bartek Klin, and Daniela Petriřan 177

Overview of Talks


Tutorial: Monads and Recognisability <i>Achim Blumensath</i>	182
Tutorial: The Composition Method <i>Miłojaj Bojańczyk</i>	182
Tutorial: String Diagrams <i>Pawel Sobocinski</i>	183
Tutorial: Coalgebra <i>Yde Venema</i>	183
Tutorial: A tutorial on (generalized) fibrations for logic, automata and language theory <i>Noam Zeilberger</i>	183
Learning automata weighted over number rings: (concretely and) categorically <i>Quentin Aristote</i>	184
On a Monadic Semantics for Circuit Description Languages <i>Ugo Dal Lago</i>	184
Trace semantics of effectful Mealy machines <i>Elena Di Lavore</i>	185
Context-free languages of string diagrams <i>Matthew David Earnshaw</i>	185
Elements of Higher-Dimensional Automata Theory <i>Uli Fahrenberg</i>	185
Systems of fixpoint equations categorically <i>Zeinab Galal</i>	186
Thin Coalgebraic Behaviours Are Inductive <i>Helle Hvid Hansen</i>	186
Algebraic Language Theory with Effects <i>Henning Urbat</i>	187
Automata in W-toposes, and general Myhill-Nerode theorem <i>Victor Iwaniack</i>	187
Graph Automata and Automaton Functors <i>Barbara König</i>	188
The relative family construction, the Brzozowski derivatives, and the Myhill-Nerode theorem <i>Paul-Andre Mellies and Noam Zeilberger</i>	188
Algebraic Recognition of Probabilistic Languages <i>Stefan Milius</i>	189

Higher-order regular languages and profinite lambda-terms	
<i>Vincent Moreau</i>	190
Relative Membership of Regular Languages	
<i>Rémi Morvan</i>	190
On the expressivity of linear recursion schemes	
<i>Andrzej Murawski</i>	191
Algebraic Recognition of Regular Functions	
<i>Lê Thành Dung Nguyễn</i>	191
How are we secretly using category theory while proving hardness of satisfying constraints	
<i>Jakub Opršal</i>	192
Functor automata - minimization and learning	
<i>Daniela Petrişan</i>	192
Equational theories of algebraic operators on Weihrauch problems	
<i>Cécilia Pradic</i>	193
Conformance Testing for Automata in a Category	
<i>Jurriaan Rot</i>	193
Measure-Theoretic Closure Operator on the Local Varieties of Regular Languages	
<i>Ryoma Sin'ya</i>	194
Name Allocation in Nominal Automata	
<i>Lutz Schröder</i>	194
Towards a Theory of Homomorphism Indistinguishability	
<i>Tim Seppelt</i>	195
Arboreal Covers over Relational Structures	
<i>Nihil Shah</i>	196
Model Completeness, MSOL and Temporal Logics	
<i>Silvio Ghilardi</i>	196
Sound and Complete Axiomatizations of (Infinite) Traces for Probabilistic Transition Systems	
<i>Ana Sokolova</i>	197
Tree algebras and bisimulation-invariant MSO on finite graphs	
<i>Thomas Colcombet</i>	197
Profinite words and Stone duality for regular languages	
<i>Sam van Gool</i>	198
Monadic second-order logic modulo bisimilarity, coalgebraically	
<i>Yde Venema</i>	198
Presheaf automata	
<i>Krzysztof Ziemiański</i>	199
Participants	200

3 Overview of Talks

3.1 Tutorial: Monads and Recognisability

Achim Blumensath (*Masaryk University – Brno, CZ*)

License  Creative Commons BY 4.0 International license
© Achim Blumensath


This tutorial intends to give an introduction to and an overview of the recently developed framework for algebraic language theory based on monads and Eilenberg-Moore algebras [3, 4, 5, 1, 2]. This framework was developed to support languages of various types, in particular those of words and trees (both finite and infinite ones). The main results concern the existence of syntactic algebras, an Eilenberg Variety Theorem, and a Reiterman Theorem.

References

- 1 A. Blumensath, *Algebraic Language Theory for Eilenberg-Moore Algebras*, Logical Methods in Computer Science, 17 (2021), pp. 6:1–6:60.
- 2 A. Blumensath, *Abstract Algebraic Language Theory*, book in preparation, <https://www.fi.muni.cz/~blumens/ALT.pdf>.
- 3 M. Bojańczyk, *Recognisable languages over monads*, unpublished note, arXiv:1502.04898v1.
- 4 M. Bojańczyk, *Languages Recognised by Finite Semigroups and their generalisations to objects such as Trees and Graphs with an emphasis on definability in Monadic Second-Order Logic*, lecture notes, arXiv:2008.11635, 2020.
- 5 H. Urbat, J. Adámek, L.-T. Chen, S. Milius, *Eilenberg Theorems for Free*, in 42nd International Symposium on Mathematical Foundations of Computer Science, MFCS 2017, August 21–25, 2017 – Aalborg, Denmark, vol. 83, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017, pp. 43:1–43:15.

3.2 Tutorial: The Composition Method

Mikołaj Bojańczyk (*University of Warsaw, PL*)

License  Creative Commons BY 4.0 International license
© Mikołaj Bojańczyk

In this talk I would like to discuss the connection between algebra (e.g. semigroups) and logic (e.g. MSO) for defining languages, in a categorical perspective, which is meant to be helpful for generalisations beyond words. There are in fact two perspectives: one, using monads, focuses on algebras, and logic becomes a derived notion. This talk is about a dual perspective, where logic is the basic notion, and algebras are derived. The main topic of interest is MSO-transductions.

3.3 Tutorial: String Diagrams

Pawel Sobocinski (Tallinn University of Technology, EE)

License © Creative Commons BY 4.0 International license
© Pawel Sobocinski

This tutorial introduced string diagrams, a mathematical notation that originates in (monoidal) category theory, through a computer science lens. In particular, the focus was on string diagrams *as syntax*, generalising classical syntax. Because string diagrams do not have an implicit assumption on the classicality of the underlying data, they are particularly useful for resource sensitive applications. After some discussion of the notion of monoidal theory, generalising the classical notion of algebraic theory, the tutorial finished with the notion of monoidal automaton (joint work with Matt Earnshaw) that accepts languages of string diagrams.

3.4 Tutorial: Coalgebra

Yde Venema (University of Amsterdam, NL)

License © Creative Commons BY 4.0 International license
© Yde Venema

In this tutorial I gave an introduction to universal coalgebra as a general categorical framework for state-based evolving systems. After presenting some motivating examples I formally introduced coalgebras for a set functor T , as well as their morphisms. I then discussed final coalgebras and the concept of coinduction, as a principle for giving both definitions and proofs. I finished with discussing the notions of behavioral equivalence and bisimilarity, and their relationship.

3.5 Tutorial: A tutorial on (generalized) fibrations for logic, automata and language theory

Noam Zeilberger (Ecole Polytechnique - Palaiseau, FR)

License © Creative Commons BY 4.0 International license
© Noam Zeilberger
Joint work of Paul-André Melliès, Noam Zeilberger
Main reference Paul-André Melliès, Noam Zeilberger: “The categorical contours of the Chomsky-Schützenberger representation theorem”, CoRR, Vol. abs/2405.14703, 2024.
URL <https://doi.org/10.48550/ARXIV.2405.14703>

The talk gave an introduction to some fibrational concepts from category theory and the relevance to studying “lifting problems” in logic, automata, and language theory. After presenting some general ideas on representing deductive systems as “bundles of categories”, the bulk of the talk focused on finite-state automata, starting from the classical idea of representing the transition graph of an NFA by a graph homomorphism $\phi : G \rightarrow B_\Sigma$ into the bouquet graph with set of loops Σ , and considering the corresponding functor $p = F\phi : FG \rightarrow FB_\Sigma$ between free categories. We discussed how to characterize determinism and codeterminism: ϕ represents the transition graph of a complete DFA (respectively coDFA) just in case p is a discrete opfibration (respectively discrete fibration). Next, emphasizing the importance of the general case, we showed how to characterize functors representing arbitrary

(potentially ambiguous) nondeterministic finite-state automata: a functor $p : D \rightarrow FB_\Sigma$ represents an NFA just in case p satisfies the unique lifting of factorizations (ULF) and finitary fiber properties. Finally, we explained how to use this characterization to define NFA over arbitrary categories, recognizing regular languages of arrows.

The tutorial was based on joint work with Paul-André Melliès.

References

- 1 Paul-André Melliès and Noam Zeilberger. *Functors are type refinement systems*. POPL 2015
- 2 Paul-André Melliès and Noam Zeilberger. *The categorical contours of the Chomsky-Schützenberger representation theorem*. LMCS (to appear), 2024, arXiv:2405.14703

3.6 Learning automata weighted over number rings: (concretely and) categorically

Quentin Aristote (IRIF - Paris, FR)

License  Creative Commons BY 4.0 International license
© Quentin Aristote

Joint work of Quentin Aristote, Sam van Gool, Daniela Petrişan, Mahsa Shirmohammadi


We study automata weighted over number rings, that is, rings of integers in an algebraic number field.

We show that number rings are what we call “almost strong Fatou”: if an n -state automaton weighted in a number field recognizes an integer-valued series, then it admits an equivalent $n+1$ -state automaton weighted in the corresponding ring of integers.

We then explain how this fits in a bigger categorical picture: given a well-behaved functor F , we give a generic procedure for retrieving the minimal C -automaton from any D -automaton. This gives in particular a generic reduction of the problem of actively learning C -automata to the problem of actively learning D -automata, which instantiates in particular to a reduction from actively learning automata weighted in number rings to automata weighted in number fields.

3.7 On a Monadic Semantics for Circuit Description Languages

Ugo Dal Lago (University of Bologna, IT)

License  Creative Commons BY 4.0 International license
© Ugo Dal Lago

Joint work of Andrea Colledan, Ugo Dal Lago, Ken Sakayori

A monad-based denotational model is introduced, which is shown to be adequate for the Proto-Quipper family of calculi, themselves being idealized versions of the Quipper programming language. The use of a monadic approach allows us to keep the value to which a term reduces distinct from the circuit that the term itself produces as a side-effect. In turn, this enables the denotational interpretation of rich type systems in which even the size of the produced circuit is controlled, at the same time justifying some of the design novelties present in such calculi.

3.8 Trace semantics of effectful Mealy machines

Elena Di Lavore (University of Pisa, IT)

License © Creative Commons BY 4.0 International license
© Elena Di Lavore

Joint work of Elena Di Lavore, Filippo Bonchi, Mario Román

Main reference Filippo Bonchi, Elena Di Lavore, Mario Román: “Effectful Mealy Machines: Bisimulation and Trace”, CoRR, Vol. abs/2410.10627, 2024.

URL <https://doi.org/10.48550/ARXIV.2410.10627>

The talk introduces effectful Mealy machines and gives them semantics both in terms of bisimilarity and traces. Bisimilarity is characterised syntactically, via free uniform feedback. Traces are a coinductive construction. Effectful Mealy machines, their bisimilarity and trace capture existing flavours of Mealy machines, bisimilarity and trace.

3.9 Context-free languages of string diagrams

Matthew David Earnshaw (Tallinn University of Technology, EE)

License © Creative Commons BY 4.0 International license
© Matthew David Earnshaw

Joint work of Matthew David Earnshaw, Mario Román

Main reference Matt Earnshaw, Mario Román: “Context-Free Languages of String Diagrams”, CoRR, Vol. abs/2404.10653, 2024.

URL <https://doi.org/10.48550/ARXIV.2404.10653>

We introduce context-free languages of morphisms in monoidal categories, extending recent work on the categorification of context-free languages, and regular languages of string diagrams. Context-free languages of string diagrams include classical context-free languages of words, trees, and hypergraphs, when instantiated over appropriate monoidal categories. We prove a representation theorem for context-free languages of string diagrams: every such language arises as the image under a monoidal functor of a regular language of string diagrams.

3.10 Elements of Higher-Dimensional Automata Theory

Uli Fahrenberg (EPITA - Cesson-Sévigné, FR)

License © Creative Commons BY 4.0 International license
© Uli Fahrenberg

Joint work of Uli Fahrenberg, Amazigh Amrane, Hugo Bazille, Emily Clement, Krzysztof Ziemiański

We introduce higher-dimensional automata (HDAs) and their languages, which consist of interval pomsets with interfaces (ipomsets). We then show a decomposition property which allows to develop an isomorphism between the category of ipomsets and a category generated by special discrete ipomsets (“starters” and “terminators”) under the relation which allows to compose subsequent starters and subsequent terminators. This in turn allows us to introduce an operational semantics for HDAs as so-called ST-automata: finite automata over the graph alphabet of starters and terminators.

3.11 Systems of fixpoint equations categorically

Zeinab Galal (University of Bologna, IT)

License  Creative Commons BY 4.0 International license
© Zeinab Galal

Fixpoints play an important role in both denotational semantics where they are used to represent recursively defined programs and data types as well as in operational semantics where many behavioral equivalences are obtained as fixpoints of some relation transformers.

In the categorical theory of fixpoint operators, we usually consider one fixpoint operator at a time and little attention is given to the study of mixed fixpoint operators where we take a different fixpoint operator for each variable. Systems combining least and greatest fixpoints over lattices are an important example as they are the basis of many static analysis and model checking methods.

I will present in this talk an axiomatization of mixed fixpoint operators first in the 1-categorical setting and then briefly mention how to extend to 2-categories in order to capture the examples of initial algebras and coalgebras of accessible functors, analytic and polynomial functors.

3.12 Thin Coalgebraic Behaviours Are Inductive

Helle Hvid Hansen (University of Groningen, NL)

License  Creative Commons BY 4.0 International license
© Helle Hvid Hansen

Joint work of Anton Chervnev, Corina Cirstea, Helle Hvid Hansen, Clemens Kupke
Main reference Anton Chervnev, Corina Cirstea, Helle Hvid Hansen, Clemens Kupke: “Thin Coalgebraic Behaviours Are Inductive”, CoRR, Vol. abs/2504.07013, 2025.
URL <https://doi.org/10.48550/ARXIV.2504.07013>

F-coalgebra automata provide a unifying, categorical setting for studying automata-theoretic verification of a variety of system types. For certain applications in quantitative model checking [1], it is crucial that the property to be checked is given by an unambiguous automaton, i.e., there is at most one accepting run on each input. This leads to the question of when unambiguous F-coalgebra automata exist. This question is also of fundamental interest, beyond verification.

For infinite words, the situation is easy, since parity word automata can be determined. For trees, it is known that deterministic automata are less expressive than nondeterministic ones, but one can recover unambiguous acceptance when restricting to thin trees, i.e., trees with only countably many infinite branches [2, 3].

Inspired by the results on thin trees, we aim to develop a similar theory for thin F-coalgebras. This talk presents the first part of this program. We show that for analytic functors F , we can define thin F-coalgebras as those coalgebras with only countably many infinite paths from each state. Our main result is an inductive characterisation of thinness via an initial algebra. To this end, we develop a syntax for thin behaviours and give a sound and complete axiomatisation of when two terms represent the same thin behaviour. Finally, for the special case of polynomial functors (the type of ranked ordered trees), we retrieve from our syntax the notion of Cantor-Bendixson rank of a thin tree.

References

- 1 Cîrstea, C. and Kupke, C., “Measure-theoretic semantics for quantitative parity automata,” in *31st EACSL Annual Conference on Computer Science Logic (CSL 2023)*, B. Klin and E. Pimentel, Eds., ser. Leibniz International Proceedings in Informatics (LIPIcs), vol. 252, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 14:1–14:20.
- 2 Idziaszek, T., Skrzypczak, M., and Bojańczyk, M., “Regular languages of thin trees”, *Theory Comput. Syst.*, vol. 58, no. 4, pp. 614–663, 2016.
- 3 Skrzypczak, M. . “Recognition by thin algebras,” in *Descriptive Set Theoretic Methods in Automata Theory: Decidability and Topological Complexity*. Springer Berlin Heidelberg, 2016, pp. 121–135.

3.13 Algebraic Language Theory with Effects

Henning Urbat (*Universität Erlangen-Nürnberg, DE*)

License © Creative Commons BY 4.0 International license
© Henning Urbat

Joint work of Fabian Lenke, Henning Urbat, Stefan Milius, Thorsten Wißmann
Main reference Fabian Lenke, Stefan Milius, Henning Urbat, Thorsten Wißmann: “Algebraic Language Theory with Effects”, in Proc. of the 52nd International Colloquium on Automata, Languages, and Programming, ICALP 2025, July 8-11, 2025, Aarhus, Denmark, LIPIcs, Vol. 334, pp. 165:1–165:20, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.
URL <https://doi.org/10.4230/LIPICS.ICALP.2025.165>

Regular languages – the languages accepted by deterministic finite automata – are known to be precisely the languages recognized by finite monoids. This characterization is the origin of algebraic language theory. We generalize the correspondence between automata and monoids to automata with generic computational effects given by a monad, providing the foundations of an effectful algebraic language theory. We show that, under suitable conditions on the monad, a language is accepted by an effectful finite automaton precisely when it is recognizable by (1) an effectful monoid morphism into an effect-free finite monoid, and (2) a monoid morphism into a monad-monoid bialgebra whose carrier is a finitely generated algebra for the monad, the former mode of recognition being conceptually completely new. As applications we obtain novel algebraic characterizations of probabilistic finite automata, nondeterministic probabilistic finite automata, and for weighted finite automata over unrestricted semirings, generalizing previous work on weighted algebraic recognition over commutative rings.

3.14 Automata in W -toposes, and general Myhill-Nerode theorem

Victor Iwaniack (*University of Côte d’Azur - Nice, FR*)

License © Creative Commons BY 4.0 International license
© Victor Iwaniack

Main reference Fabian Lenke, Stefan Milius, Henning Urbat, Thorsten Wißmann: “Algebraic Language Theory with Effects”, in Proc. of the 52nd International Colloquium on Automata, Languages, and Programming, ICALP 2025, July 8-11, 2025, Aarhus, Denmark, LIPIcs, Vol. 334, pp. 165:1–165:20, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.
URL <https://doi.org/10.4230/LIPICS.ICALP.2025.165>

We enrich the functorial viewpoint on automata of Colcombet and Petrişan to work with automata in any topos. The whole minimisation framework adapts to enrichment with notably enriched Kan extensions and enriched orthogonal factorisation systems. We use this general minimisation to deduce a general Myhill-Nerode theorem. This theorem depends on

a notion of “finiteness”. By instantiating this framework with the topos of nominal sets and the orbit-finiteness, we recover the Myhill-Nerode theorem of Bojańczyk, Klin and Lasota. But the theorem applies to other finiteness conditions, such as Kuratowski-finiteness or “stalkwise-finiteness”. Moreover, the whole enriched framework can work with other monoidal categories than toposes.

3.15 Graph Automata and Automaton Functors

Barbara König (Universität Duisburg-Essen, DE)

License © Creative Commons BY 4.0 International license

© Barbara König

Joint work of Sander Bruggink, Christoph Blume, Barbara König

Main reference H. J. Sander Bruggink, Barbara König: “Recognizable languages of arrows and cospans”, *Math. Struct. Comput. Sci.*, Vol. 28(8), pp. 1290–1332, 2018.

URL <https://doi.org/10.1017/S096012951800018X>

We generalize Courcelle’s recognizable graph languages and results on monadic second-order logic to more general structures.

We give a category-theoretical characterization of recognizability. A recognizable subset of arrows in a category is defined via a functor into the category of relations on finite sets. This can be seen as a straightforward generalization of finite automata and we show how to obtain graph automata - accepting recognizable graph languages – by applying the theory to the category of cospans of graphs.

We also introduce a simple logic that allows to quantify over the subobjects of a categorical object and we show that, for the category of graphs, this logic is equally expressive as monadic second-order graph logic (MSOGL). Furthermore, we explain that in the more general setting of hereditary pushout categories, a class of categories closely related to adhesive categories, we can recover Courcelle’s result that every MSOGL-expressible property is recognizable.

The talk concludes by reviewing a practical implementation of graph automata with applications to the verification of graph transformation systems.

3.16 The relative family construction, the Brzozowski derivatives, and the Myhill-Nerode theorem

Paul-Andre Mellies (Université Paris Cité, FR), Noam Zeilberger (Ecole Polytechnique - Palaiseau, FR)

License © Creative Commons BY 4.0 International license

© Paul-Andre Mellies and Noam Zeilberger

Main reference Paul-André Mellies, Noam Zeilberger: “The categorical contours of the Chomsky-Schützenberger representation theorem”, *CoRR*, Vol. abs/2405.14703, 2024.

URL <https://doi.org/10.48550/ARXIV.2405.14703>

In this talk, I explained how to establish a “run-aware” version of the Myhill-Nerode theorem based on a fibrational / categorical account of Brzozowski derivatives. A non-deterministic finite state automaton is defined as a finitary functor $p : E \rightarrow B$ satisfying the unique lifting of factorisation (ULF) property, where E is a category of states and runs, and B is a category of sorts and words.

A relative family construction is introduced, which provides a “categorified” powerset construction for any functor $p : E \rightarrow B$. One key observation is that the resulting functor $\text{Fam}(E, p) \rightarrow B$ is a fibration if and only if p is a ULF functor.

The observation is applied to the finitary and ULF functor $p^{\text{op}} \times p : E^{\text{op}} \times E \rightarrow B^{\text{op}} \times B$ in order to obtain a fibration $\text{Fam}(E^{\text{op}} \times E, p^{\text{op}} \times p) \rightarrow B^{\text{op}} \times B$ together with a functor $F^\dagger : \text{Fam}(E^{\text{op}} \times E, p^{\text{op}} \times p) \rightarrow \text{Lang}^\rightarrow$ over $B^{\text{op}} \times B$ where the fibration $\text{Lang}^\rightarrow \rightarrow B^{\text{op}} \times B$ is the pullback of the codomain fibration $\text{Set}^\rightarrow \rightarrow \text{Set}$ along the hom functor $\text{hom}_B : B^{\text{op}} \times B \rightarrow \text{Set}$. The functor F^\dagger itself is obtained from the functor $\text{hom}_p : E^{\text{op}} \times E \rightarrow \text{Set}^\rightarrow$ which transports every pair of states (R, S) to the function $\text{Hom}_p : \text{Hom}_E(R, S) \rightarrow \text{Hom}_B(A = pR, B = pS)$ associated to the functor p .

The functor F^\dagger is shown to transport cartesian maps to cartesian maps in $\text{Lang}^\rightarrow \rightarrow B^{\text{op}} \times B$ which appear as a “run-aware” form of Brzozowski derivatives in Lang^\rightarrow . From this follows a Myhill-Nerode theorem for “run-aware” languages, stating that a language L is regular if and only if its class of derivatives is finitely generated (using sums or disjoint unions) at each fiber over an object of the category $B^{\text{op}} \times B$.

3.17 Algebraic Recognition of Probabilistic Languages

Stefan Milius (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 4.0 International license
© Stefan Milius

Joint work of Fabian Lenke, Stefan Milius, Henning Urbat, Thorsten Wißmann


Main reference Fabian Lenke, Stefan Milius, Henning Urbat, Thorsten Wißmann: “Algebraic Language Theory with Effects”, in Proc. of the 52nd International Colloquium on Automata, Languages, and Programming, ICALP 2025, July 8-11, 2025, Aarhus, Denmark, LIPIcs, Vol. 334, pp. 165:1–165:20, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.

URL <https://doi.org/10.4230/LIPICS.ICALP.2025.165>

Regular languages – the languages accepted by deterministic finite automata – are known to be precisely the languages recognized by finite monoids. This characterization is the origin of algebraic language theory. We generalize the correspondence between automata and monoids to probabilistic automata, providing the foundations of a probabilistic algebraic language theory. We show that a language is computable by a probabilistic finite automaton precisely when it is recognizable by (1) a probabilistic monoid morphism into an (ordinary) finite monoid, and (2) an (ordinary) monoid morphism into an fg-carried convex monoid: a finitely generated convex set equipped with a monoid operation that distributes over the convex structure. The former mode of recognition is conceptually completely new. Moreover, every probabilistic language has a syntactic monoid, that is, a minimal algebraic recognizer. However, the syntactic monoid is not fg-carried in general. As an open problem we leave the question whether the syntactic monoid is finitely presentable as a convex monoid.

3.18 Higher-order regular languages and profinite lambda-terms

Vincent Moreau (*Université Paris Cité, FR*)

License  Creative Commons BY 4.0 International license
© Vincent Moreau

Joint work of Vincent Moreau, Sam van Gool, Paul-André Melliès


Main reference Sam van Gool, Paul-André Melliès, Vincent Moreau: “Profinite lambda-terms and parametricity”, in Proc. of the 39th Conference on the Mathematical Foundations of Programming Semantics, MFPS XXXIX, Indiana University, Bloomington, IN, USA, June 21-23, 2023, EPTICS, Vol. 3, EpiSciences, 2023.

URL <https://doi.org/10.46298/ENTICS.12280>

A fundamental observation at the heart of the topological approach to language theory is the fact that the topological space of profinite words is the Stone dual of the Boolean algebra of regular languages $\text{Reg}(\Sigma)$ over the alphabet Σ . Using ideas coming from the seminal work of Salvati on languages of λ -terms, who introduced the family of Boolean algebras $\text{Reg}(A)$ for any type A , we introduce the space of profinite λ -terms of type A as the Stone dual of $\text{Reg}(A)$, which generalizes to the higher-order setting the notion of profinite word. Types and profinite λ -terms assemble into a Stone-enriched cartesian closed category ProLam , which is the free such category \mathcal{C} that recognizes at most regular languages in the sense of Salvati. This demonstrates the compositional aspects of profinite λ -terms, which we think of as the terms of an extension of the λ -calculus with profinite operators.

3.19 Relative Membership of Regular Languages

Rémi Morvan (*University of Bordeaux, FR*)

License  Creative Commons BY 4.0 International license
© Rémi Morvan

Main reference Rémi Morvan: “The Algebras for Automatic Relations”, in Proc. of the 33rd EACSL Annual Conference on Computer Science Logic, CSL 2025, February 10-14, 2025, Amsterdam, Netherlands, LIPIcs, Vol. 326, pp. 21:1–21:21, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.

URL <https://doi.org/10.4230/LIPICS.CSL.2025.21>

Given a regular language Ω and a class of regular languages V , we want to understand when the class of languages that can be written as the intersection of Ω with a language from V has decidable membership. We provide a sufficient condition on Ω such that whenever V has decidable membership (and has some mild closure properties, i.e. is a pseudovariety), then the relativization of V with respect to Ω also has decidable membership.

3.20 On the expressivity of linear recursion schemes

Andrzej Murawski (*University of Oxford, GB*)

License © Creative Commons BY 4.0 International license
© Andrzej Murawski

Joint work of Pierre Clairambault, Andrzej S. Murawski

Main reference Pierre Clairambault, Andrzej S. Murawski: “On the Expressivity of Linear Recursion Schemes”, in Proc. of the 44th International Symposium on Mathematical Foundations of Computer Science, MFCS 2019, August 26-30, 2019, Aachen, Germany, LIPIcs, Vol. 138, pp. 50:1–50:14, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

URL <https://doi.org/10.4230/LIPICS.MFCS.2019.50>

In the last decade or so, higher-order recursion schemes (HORS) have emerged as a promising technique for model-checking higher-order programs. I will discuss several results concerning the case when HORS are typed using linear logic (intuitionistic multiplicative additive linear logic, to be precise). It turns out that such schemes have an automata-theoretic counterpart, namely restricted tree-stack automata, which come from linguistics, where they were introduced to study the so-called multiple context-free languages. This leads to a new perspective on linear HORS and new decidability results. This is joint work with Pierre Clairambault (deterministic case, MFCS’19), Guanyan Li and Luke Ong (probabilistic case, LICS’22).

3.21 Algebraic Recognition of Regular Functions

Lê Thành Dung Nguyễn (*CNRS & Aix-Marseille Univ., FR*)

License © Creative Commons BY 4.0 International license
© Lê Thành Dung Nguyễn

Joint work of Mikołaj Bojańczyk, Lê Thành Dung Nguyễn

Main reference Mikołaj Bojańczyk, Lê Thành Dung Nguyễn: “Algebraic Recognition of Regular Functions”, in Proc. of the 50th International Colloquium on Automata, Languages, and Programming, ICALP 2023, July 10-14, 2023, Paderborn, Germany, LIPIcs, Vol. 261, pp. 117:1–117:19, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.

URL <https://doi.org/10.4230/LIPICS.ICALP.2023.117>

A string-to-string function is regular (i.e. is an MSO-transduction) if and only if it factors as

$$\Sigma^* \xrightarrow{\text{some semigroup homomorphism}} F(\Gamma^*) \xrightarrow{\text{out}_{\Gamma^*}} \Gamma^*$$

where F is a finiteness-preserving endofunctor on semigroups, and out_{Γ^*} is a component of a natural transformation

$$\begin{array}{ccc} & \xrightarrow{[\text{forgetful functor}] \circ F} & \\ \text{Semigroups} & \Downarrow & \text{Sets.} \\ & \xrightarrow{\text{forgetful functor}} & \end{array}$$

3.22 How are we secretly using category theory while proving hardness of satisfying constraints

Jakub Opršal (University of Birmingham, GB)

License © Creative Commons BY 4.0 International license

© Jakub Opršal

Joint work of Maximilian Hadek, Tomas Jakl, Jakub Opršal

Main reference Maximilian Hadek, Tomas Jakl, Jakub Opršal: “A categorical perspective on constraint satisfaction: The wonderland of adjunctions”, CoRR, Vol. abs/2503.10353, 2025.

URL <https://doi.org/10.48550/ARXIV.2503.10353>

In the talk, I have provided a new categorical perspective on the *algebraic approach to the constraint satisfaction problem (CSP)*. The approach has been a prevalent method of the study of the complexity of these problems since the early 2000s, and many breakthrough achievements can be either directly or indirectly attributed to it. A prime result is the Bulatov–Zhuk Dichotomy Theorem, which states that every finite-template CSP is either in P or NP-complete.

I have explained the *gadget reductions* used by the algebraic approach as a case of a well-known categorical construction (left Kan extension along Yoneda), and I have stated and proved the fundamental theorem of the algebraic approach in the categorical language. The theorem provides a condition for a CSP to be NP-complete that covers the hardness part of the dichotomy.

3.23 Functor automata - minimization and learning

Daniela Petrişan (Université Paris Cité, FR)

License © Creative Commons BY 4.0 International license

© Daniela Petrişan

Joint work of Thomas Colcombet, Daniela Petrişan

Main reference Thomas Colcombet, Daniela Petrişan: “Automata Minimization: a Functorial Approach”, Log. Methods Comput. Sci., Vol. 16(1), 2020.

URL [https://doi.org/10.23638/LMCS-16\(1:32\)2020](https://doi.org/10.23638/LMCS-16(1:32)2020)

In this talk I present a categorical approach to automata based on the categorical notion of functor. The basic idea is to see an automaton as a machine taking some input specified by an “input” category and producing some effect – such as non-determinism, words over an output alphabet or probability values – encoded by an “output” category. Usually, the output category is a Kleisli category for a monad specifying a given effect. We discuss how adjunctions between categories can be lifted to adjunctions between categories of automata, encompassing examples such as determinization or completion. We then present sufficient conditions on the output category such that minimization and learning algorithms exist.

References

- 1 Thomas Colcombet and Daniela Petrişan. *Automata Minimization: a Functorial Approach*. Log. Methods Comput. Sci., vol 16 (1), 2020
- 2 Thomas Colcombet, Daniela Petrişan and Riccardo Stabile. *Learning Automata and Transducers: A Categorical Approach*. CSL 2021

3.24 Equational theories of algebraic operators on Weihrauch problems

Cécilia Pradic (Swansea University, GB)

License © Creative Commons BY 4.0 International license
© Cécilia Pradic

Joint work of Cécilia Pradic, Eike Neumann, Arno Pauly, Ian Price

Main reference Cécilia Pradic: “The equational theory of the Weihrauch lattice with (iterated) composition”, CoRR, Vol. abs/2408.14999, 2024.

URL <https://doi.org/10.48550/ARXIV.2408.14999>

Weihrauch reducibility is a notion that has gained a lot of traction in computable analysis in the last decade. It may be regarded as a framework to compare the uncomputational strength of problems, much like reverse mathematics.

Weihrauch problems include natural mathematical problems such as WKL and RT, but the corresponding degrees also enjoy a rich algebraic structure induced by algebraic operation on problems. A great number of these operations correspond to the structure of the category problems and reductions, which is equivalent to (a full subcategory of) the category of containers/polynomials over represented spaces and computable maps, a well-known nice category which is bicartesian closed and monoidal closed among other things.

After introducing these notions, I will discuss the equational theory of the Weihrauch lattice equipped with the composition product and finite iterations thereof. Terms in this theory can be translated to alternating automata, and reductions regarded as a somewhat weird kind of simulation. This leads to decidability and a complete axiomatization.

Very little to the development is specific to Weihrauch problems and could potentially be adapted to handle containers over a range of extensive locally cartesian closed categories with dependent W-types.

3.25 Conformance Testing for Automata in a Category

Jurriaan Rot (Radboud University Nijmegen, NL)

License © Creative Commons BY 4.0 International license
© Jurriaan Rot

Conformance testing is often used to implement the equivalence query in active automata learning. In this talk, I will highlight this application, and go on to discuss the basic notions of n -completeness and the classical W -method. I will then discuss recent joint work with Bálint Kocsis ([1], to appear at FoSSaCS 2025) on generalising part of this theory to the abstract setting of automata in a category.

References

- 1 Bálint Kocsis and Jurriaan Rot, *Complete Test Suites for Automata in Monoidal Closed Categories*, FoSSaCS 2025

3.26 Measure-Theoretic Closure Operator on the Local Varieties of Regular Languages

Ryoma Sin'ya (Akita University, JP)

License  Creative Commons BY 4.0 International license

© Ryoma Sin'ya

Joint work of Takao Yuyama, Yoshiki Nakamura, Yutaro Yamaguchi, Kazuhiro Inaba


A language L is said to be regular measurable if there exists an infinite sequence of regular languages that “converges” to L . This notion was introduced by the speaker in 2021 [1] and used for classifying non-regular languages by using regular languages. In this talk, we describe why this notion was introduced, and give a brief overview of decidability results relating to the measurability on subclasses (local subvarieties) of regular languages, eg., star-free, generalised definite, and group languages.

References

- 1 Ryoma Sin'ya. *Asymptotic Approximation by Regular Languages*, In Proceedings of the 47th International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2021), 2021.

3.27 Name Allocation in Nominal Automata

Lutz Schröder (Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 4.0 International license

© Lutz Schröder

Joint work of Florian Frank, Daniel Hausmann, Dexter Kozen Stefan Milius, Simon Prucker, Henning Urbat, Florian Frank, Daniel Hausmann, Dexter Kozen Stefan Milius, Simon Prucker, Lutz Schröder, Henning Urbat, Thorsten Wißmann

Main reference Lutz Schröder, Dexter Kozen, Stefan Milius, Thorsten Wißmann: “Nominal Automata with Name Binding”, in Proc. of the Foundations of Software Science and Computation Structures - 20th International Conference, FOSSACS 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings, Lecture Notes in Computer Science, Vol. 10203, pp. 124–142, 2017.

URL https://doi.org/10.1007/978-3-662-54458-7_8

Formal languages over infinite alphabets serve the modelling and specification of structures and processes with data; here, infinite alphabets represent infinite data types equipped with very restricted interfaces, typically just checks for (in)equality. Such formal languages are thus typically referred to as data languages. There is a large variety of automata models for data languages; one of the most established models is the classical register automaton model, in which letters encountered in the input can be stored in a fixed number of registers for later comparison with other letters. Register automata are generally equivalent to automata models over nominal sets; for instance, nondeterministic register automata with nondeterministic update are equivalent to nondeterministic orbit-finite automata. In such equivalences, states of a nominal automaton correspond to configurations of a register automaton, consisting of a control state and the current register content.

In register-based models, expressiveness typically varies with the power of control; e.g., nondeterministic register automata are strictly more expressive than deterministic ones. While membership is typically decidable, inclusion of register automata tends to be either undecidable or of prohibitively high complexity unless stringent restrictions are imposed on either the power of control, e.g. requiring unambiguity, or on the number of registers. We introduce nominal automata models with explicit name allocation, which strike a compromise

between expressive power and complexity. In such models, freshness of letters is modelled via alpha-equivalence on words with explicit allocation. Roughly speaking, this means that distinctness of a newly read letter with respect to a stored letter can only be enforced if the stored letter is expected to be seen again. This still allows for typical modes of expression requiring, for instance, explicit closure of named sessions; on the other hand, it allows for inclusion checking in elementary time under unrestricted nondeterminism and without bounding the number of registers. Indeed, the complexity is typically no worse than the complexity of the corresponding finite-alphabet model when the number of registers is fixed as a parameter. This has originally been established for non-deterministic finite-word automata with name allocation; similar results have subsequently been obtained for infinite-word automata, tree automata, and, very recently, for alternating finite-word automata.

3.28 Towards a Theory of Homomorphism Indistinguishability

Tim Seppelt (IT University of Copenhagen, DK)

License © Creative Commons BY 4.0 International license

© Tim Seppelt

Main reference Tim Seppelt: “Homomorphism Indistinguishability”, Dissertation, RWTH Aachen University, Aachen, 2024.

URL <https://doi.org/10.18154/RWTH-2024-11629>

In 1967, Lovász showed that two graphs G and H are isomorphic if, and only if, they admit the same number of homomorphisms from every graph F . Subsequently, it emerged that many natural graph isomorphism relaxations from fields as diverse as finite model theory, category theory, optimization, quantum information, and algebraic graph theory can be characterized as homomorphism indistinguishability relations over natural restricted graph classes.

In this talk, we set out to abstract from these examples and develop a theory of homomorphism indistinguishability. We propose to answer, given a graph class \mathcal{F} , the following questions:

1. What is the distinguishing power of the homomorphism indistinguishability relation $\equiv_{\mathcal{F}}$ of \mathcal{F} ?
2. What is the complexity of deciding homomorphism indistinguishability over \mathcal{F} ?

We discuss progress on both questions covering Roberson’s conjecture on the homomorphism distinguishing closure, a recent result of myself on the complexity of homomorphism indistinguishability over recognizable \mathcal{F} of bounded treewidth, and the role of minor-closed graph classes in the emerging theory of homomorphism indistinguishability.

3.29 Arboreal Covers over Relational Structures

Nihil Shah (University of Cambridge, GB)

License  Creative Commons BY 4.0 International license
© Nihil Shah

Joint work of Nihil Shah, Samson Abramsky, Luca Reggio, Dan Marsden, Yoav Montacute, Anuj Dawar, Tomáš Jakl, Adam O’Conghaile, Amin Karamlou

Main reference Nihil A Shah: “Arboreal covers over relational structures”, University of Oxford, 2024.


URL <https://ora.ox.ac.uk/objects/uuid:87f53217-655e-47b0-8991-14bc0ced4443/files/dwd375x22s>

Model-comparison games are an important tool in both finite and unrestricted model theory to prove when two structures satisfy the same sentences in a logic. Game comonads encode particular model comparison games as comonads on the categories of structures. This started with the work of Abramsky, Dawar, and Wang in 2017 on encoding pebble games for finite variable logics. Since this initial work, several examples of game comonads have been engineered to capture a wide-range of logics. Each game comonad provides a categorical characterisation of equivalence in a logic and its variants. The categorical constructions common to these comonads have proved to be a nice tool for organizing tacit connections between syntactic resources in logic, hierarchical approximations to constraint satisfaction and isomorphism, and well-known combinatorial parameters such as treewidth and tree-depth.

Determining the commonalities shared amongst game comonads that enable them to have these features lead to the axiomatic formulation of arboreal category and arboreal cover by Abramsky and Reggio 2021. Arboreal categories axiomatise the notion of a category with “tree-shaped” objects and provide a native setting for dynamic notions like simulation, bisimulation, and resource-indexing. Arboreal covers are comonadic adjunctions to any category that allow application of these dynamic notions to the static objects of the target category. Game comonads all arise as arboreal covers over categories of relational structures.

3.30 Model Completeness, MSOL and Temporal Logics

Silvio Ghilardi (University of Milan, IT)

License  Creative Commons BY 4.0 International license
© Silvio Ghilardi

Joint work of Silvio Ghilardi, Luca Carai, Sam van Gool

We shall connect classical algebraic model theory with monadic second order logic (MSOL) on infinite structures (natural numbers, infinite binary trees, infinite trees of arbitrary degrees, etc.)

The idea comes from the well-known connection between MSOL and automata: this connection allows a conversion between MSOL formulae and suitable automata (and back). Applying such conversion, every MSOL formula turns out to be equivalent to a formula which is “almost existential”. The “almost” proviso (coming from acceptance conditions) can be removed if the language is enriched with some temporal operators in LTL/CTL style.

First-order theories whose formulae are equivalent to existential formulae are precisely model complete theories. MSOL interpreted on a structure can be viewed as first-order logic (FOL) interpreted on the corresponding power set Boolean first-order structures. From these facts we may formulate the slogan “MSOL is the model companion of temporal logic”. The aim of the talk is to supply results giving a precise formal meaning to such slogan.

3.31 Sound and Complete Axiomatizations of (Infinite) Traces for Probabilistic Transition Systems

Ana Sokolova (*Paris Lodron Universität Salzburg, AT*)

License © Creative Commons BY 4.0 International license

© Ana Sokolova

Joint work of Corina Cirstea, Larry Moss, Todd Schmidt, Tori Noquez, Alexandra Silva, Ana Sokolova

Main reference Corina Cirstea, Lawrence S. Moss, Victoria Noquez, Todd Schmid, Alexandra Silva, Ana Sokolova: “A Complete Inference System for Probabilistic Infinite Trace Equivalence”, in Proc. of the 33rd EACSL Annual Conference on Computer Science Logic, CSL 2025, February 10-14, 2025, Amsterdam, Netherlands, LIPIcs, Vol. 326, pp. 30:1–30:23, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.

URL <https://doi.org/10.4230/LIPICS.CSL.2025.30>

This talk is about recent results on axiomatizing infinite trace semantics, also called stream semantics, for generative probabilistic transition systems, also called labelled Markov chains.

The talk is about trace semantics and its equational (sound and complete) axiomatization. Probabilistic transition systems can be represented by suitable probabilistic expressions, with the property that the expression and the transition system (state) have the same (in this case trace) semantics. We then provide equations on these expressions that fully characterize (infinite) trace semantics: two expressions are trace equivalent iff they are provably equivalent with the presented axioms. The axioms for finite traces have been given by Silva and myself in 2011. The soundness and completeness proof is coalgebraic.

In recent work with the group of coauthors mentioned above, we present the first sound and complete axiomatization of infinite trace semantics for generative probabilistic transition systems. Our approach is categorical, and we build on recent results on proper functors over convex sets - in particular on a novel and simpler proof of properness of the involved functor. At the core of our proof is a characterization of infinite traces as the final coalgebra of a functor over convex algebras. Somewhat surprisingly, our axiomatization of infinite trace semantics coincides with that of finite trace semantics, even though the techniques used in the completeness proof are significantly different.

3.32 Tree algebras and bisimulation-invariant MSO on finite graphs

Thomas Colcombet (*IRIF – CNRS – Université Paris Cité, FR*)

License © Creative Commons BY 4.0 International license

© Thomas Colcombet

Joint work of Amina Doumane, Denis Kuperberg, Thomas Colcombet

Main reference Thomas Colcombet, Amina Doumane, Denis Kuperberg: “Tree Algebras and Bisimulation-Invariant MSO on Finite Graphs”, in Proc. of the 52nd International Colloquium on Automata, Languages, and Programming, ICALP 2025, July 8-11, 2025, Aarhus, Denmark, LIPIcs, Vol. 334, pp. 152:1–152:16, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.

URL <https://doi.org/10.4230/LIPICS.ICALP.2025.152>

In this work with Amina Doumane and Denis Kuperberg, we establish that the bisimulation invariant fragment of MSO over finite transition systems is expressively equivalent over finite transition systems to modal μ -calculus, a question that had remained open for several decades. The proof goes by translating the question to an algebraic framework, and showing that the languages of regular trees that are recognised by finitary tree algebras whose sorts zero and one are finite are the regular ones. This corresponds for trees to a weak form of the key translation of Wilke algebras to ω -semigroup over infinite words, and was also a missing piece in the algebraic theory of regular languages of infinite trees for twenty years.

3.33 Profinite words and Stone duality for regular languages

Sam van Gool (*Université Paris Cité, FR*)

License  Creative Commons BY 4.0 International license
© Sam van Gool

Joint work of Mai Gehrke, Sam van Gool, Paul-Anrdre Melliès, Vincent Moreau, Benjamin Steinberg
Main reference Mai Gehrke, Sam van Gool: “Topological duality for distributive lattices: Theory and Applications”. Cambridge Tracts in Theoretical Computer Science 61, Cambridge University Press (2024).
URL <https://doi.org/10.1017/9781009349680>


In this talk, we show how profinite words are types, in the model-theoretic sense, for monadic second-order logic. Starting from the Stone duality approach to regular languages developed in [1], we show how the free profinite monoid naturally arises as the dual of a colimit chain of finite Boolean algebras with operators. In the first-order setting, this leads to the model-theoretic view on proaperiodic monoids developed in [2]. We also mention links with the duality-theoretic notion of preserving joins at primes [3, Ch. 8]. These considerations were generalized to give profinite lambda-terms in [4], also see V. Moreau’s talk in the same seminar.

References

- 1 Gehrke, M., S. Grigorieff and J.-E. Pin, *Duality and equational theory of regular languages*, in: International Colloquium on Automata, Languages and Programming (ICALP), 246–257, 2008.
- 2 van Gool, S. and Steinberg, B. *Pro-aperiodic monoids via saturated models*, Israel Journal of Mathematics 234: 451–498, 2019.
- 3 Mai Gehrke and Sam van Gool, *Topological duality for distributive lattices: Theory and Applications*. Cambridge Tracts in Theoretical Computer Science 61. Cambridge University Press, 2024.
- 4 Sam van Gool, Paul-André Melliès and Vincent Moreau, *Profinite lambda-terms and parametricity*, Mathematical Foundations of Programming Semantics (MFPS), 2023.

3.34 Monadic second-order logic modulo bisimilarity, coalgebraically

Yde Venema (*University of Amsterdam, NL*)

License  Creative Commons BY 4.0 International license
© Yde Venema

Joint work of Sebastian Enqvist, Fatemeh Seifan, Yde Venema
Main reference Sebastian Enqvist, Fatemeh Seifan, Yde Venema: “An expressive completeness theorem for coalgebraic modal mu-calculi”, Log. Methods Comput. Sci., Vol. 13(2), 2017.
URL [https://doi.org/10.23638/LMCS-13\(2:14\)2017](https://doi.org/10.23638/LMCS-13(2:14)2017)

The Janin-Walukiewicz Theorem states that on Kripke models, the modal μ -calculus is the bisimulation-invariant fragment of monadic second-order logic. In the talk I showed how to generalise this result to T-coalgebras for an arbitrary set functor T (satisfying some conditions). In the proof I introduced the notion of coalgebra automata.

3.35 Presheaf automata

Krzysztof Ziemianski (University of Warsaw, PL)

License © Creative Commons BY 4.0 International license
© Krzysztof Ziemianski

Joint work of Krzysztof Ziemianski, Georg Struth

Main reference Georg Struth, Krzysztof Ziemianski: “Presheaf automata”, CoRR, Vol. abs/2409.04612, 2024.

URL <https://doi.org/10.48550/ARXIV.2409.04612>

I introduce presheaf automata as presheaves over categories with two distinguished families of morphisms. Presheaf automata are a generalisation of different variants of higher-dimensional automata and other automata-like formalisms, including Petri nets and pushdown automata. I develop the foundations of a language theory for them and define runs, languages, notions of regularity and rationality of languages, determinism and bisimulations.

Participants

- Samson Abramsky
University College London, GB
- Bahareh Afshari
University of Gothenburg, SE
- Quentin Aristote
IRIF – Paris, FR
- Achim Blumensath
Masaryk University – Brno, CZ
- Mikołaj Bojańczyk
University of Warsaw, PL
- Célia Borlido
University of Coimbra, PT
- Thomas Colcombet
IRIF – CNRS – Université
Paris Cité, FR
- Ugo Dal Lago
University of Bologna, IT
- Elena Di Lavore
University of Pisa, IT
- Matthew David Earnshaw
Tallinn University of
Technology, EE
- Uli Fahrenberg
EPITA – Cesson-Sévigné, FR
- Zeinab Galal
University of Bologna, IT
- Silvio Ghilardi
University of Milan, IT
- Helle Hvid Hansen
University of Groningen, NL
- Victor Iwaniack
University of Côte d’Azur –
Nice, FR
- Tomas Jakl
Czech Technical University –
Prague, CZ
- Bartek Klin
University of Oxford, GB
- Barbara König
Universität Duisburg-Essen, DE
- Aliaume Lopez
University of Warsaw, PL
- Fosco Loregian
Tallinn University of
Technology, EE
- Jérémie Marquès
University of Milan, IT
- Paul-Andre Mellies
Université Paris Cité, FR
- Karla Messing
Universität Duisburg-Essen, DE
- Stefan Milius
Universität Erlangen-
Nürnberg, DE
- Vincent Moreau
Université Paris Cité, FR
- Rémi Morvan
University of Bordeaux, FR
- Andrzej Murawski
University of Oxford, GB
- Lê Thành Dung Nguyen
CNRS & Aix-Marseille Univ., FR
- Jakub Opršal
University of Birmingham, GB
- Daniela Petrişan
Université Paris Cité, FR
- Cécilia Pradic
Swansea University, GB
- Jurriaan Rot
Radboud University
Nijmegen, NL
- Lutz Schröder
Universität Erlangen-
Nürnberg, DE
- Tim Seppelt
IT University of
Copenhagen, DK
- Nihil Shah
University of Cambridge, GB
- Ryoma Sin’ya
Akita University, JP
- Pawel Sobocinski
Tallinn University of
Technology, EE
- Ana Sokolova
Paris Lodron Universität
Salzburg, AT
- Rafal Stefanski
University of Warsaw, PL
- Howard Straubing
Boston College, US
- Henning Urbat
Universität Erlangen-
Nürnberg, DE
- Sam van Gool
Université Paris Cité, FR
- Yde Venema
University of Amsterdam, NL
- Jana Wagemaker
Radboud University
Nijmegen, NL
- Noam Zeilberger
Ecole Polytechnique –
Palaiseau, FR
- Krzysztof Ziemiański
University of Warsaw, PL



Explainability in Focus: Advancing Evaluation through Reusable Experiment Design

Simone Stumpf^{*1}, Stefano Teso^{*2}, and Elizabeth M. Daly^{*3}

1 University of Glasgow, UK. simone.stumpf@glasgow.ac.uk

2 University of Trento, IT. stefano.teso@unitn.it

3 IBM Research, IE. elizabeth.daly@ie.ibm.com

Abstract

This report summarizes the outcomes of Dagstuhl Seminar 25142, which convened leading researchers and practitioners to address the pressing challenges in evaluating explainable artificial intelligence (XAI). The seminar focused on developing reusable experimental designs and robust evaluation frameworks that balance technical rigor with human-centered considerations. Key themes included the need for standardized metrics, the contextual relevance of evaluation criteria, and the integration of human understanding, trust, and reliance into assessment methodologies. Through a series of talks, collaborative discussions, and case studies across domains such as healthcare, hiring, and decision support, the seminar identified critical gaps in current XAI evaluation practices and proposed actionable strategies to bridge them. The report presents a refined taxonomy of evaluation criteria, practical guidance for experimental design, and a roadmap for future interdisciplinary collaboration in responsible and transparent AI development.

Seminar March 30 – April 2, 2025 – <https://www.dagstuhl.de/25142>

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Human-centered computing → Human computer interaction (HCI); Computing methodologies → Machine learning

Keywords and phrases Explainability, Mental Models, interactive machine learning, Experiment Design, Human-centered AI Dagstuhl Seminar

Digital Object Identifier 10.4230/DagRep.15.3.201

1 Executive Summary

Simone Stumpf (University of Glasgow, simone.stumpf@glasgow.ac.uk)

Stefano Teso (University of Trento, Italy, stefano.teso@unitn.it)

Elizabeth M. Daly (IBM Research, Ireland, elizabeth.daly@ie.ibm.com)

License  Creative Commons BY 4.0 International license
© Simone Stumpf, Elizabeth Daly, and Stefano Teso

This summary outlines the key outcomes of Dagstuhl Seminar 25142, which focused on the role of explanations in advancing Responsible and Ethical AI. The discussion emphasized the importance of explainability in AI systems to:

- **Demystify AI systems:** Helping users understand the rationale behind AI-generated outcomes.
- **Promote accountability:** Enabling users to verify that decisions are based on valid, unbiased data.

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Explainability in Focus: Advancing Evaluation through Reusable Experiment Design, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 201–224

Editors: Elizabeth M. Daly, Simone Stumpf, and Stefano Teso



Dagstuhl Reports

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

- **Encourage transparency:** Reinforcing trust and confidence in AI technologies through clear, interpretable outputs.
- **Support debugging and decision-making:** Assisting users in evaluating whether to trust a prediction or recommendation.

This seminar brought together researchers, practitioners, and experts in the field of explainable AI to collaboratively develop reusable resources aimed at standardizing the evaluation of explainability methods. The goal was to ensure that evaluation practices are robust, consistent, and adaptable across diverse contexts and applications.

A major outcome of the seminar was the identification of three key challenges:

1. Balancing technical rigor with human-centric considerations when determining which aspects of explanations should be assessed;
2. Developing consistent and reliable metrics for evaluating the selected criteria; and
3. Ensuring that both criteria and measurements are appropriately tailored to specific use cases where explainability is critical.

To illustrate practical applications of the discussed frameworks, we presented case studies showcasing end-to-end evaluation examples.

2 Table of Contents

Executive Summary

Simone Stumpf, Elizabeth Daly, and Stefano Teso 201

Overview of Talks

Why are so many studies measuring XAI wrong?
Simone Stumpf 204

Replication in explainable AI: a case study in group recommender systems
Nava Tintarev 204

A Review of Taxonomies of XAI Evaluation Methods
Timo Speith 205

Doing multiclass Shapley values properly
Peter Flach 205

Explanations as Constructed Arguments
Peter Clark 206

Introduction 206

Evaluation Criteria for Explanations 207

Self-Reported and Observed Understanding 208

Explanation Fidelity and Stability 209

Trust, Reliance and Performance 211

Relationship between Usage Context and Criteria 213

Metrics for Model Improvement 213

Metrics for Capability Assessment/Auditing 214

Case Study 216

Model Improvement (Medical Domain) 216

Capability Assessment (Hiring) 217

Decision Support (ICU Triage) 219

Conclusion 221

Participants 224

3 Overview of Talks

3.1 Why are so many studies measuring XAI wrong?

Simone Stumpf (University of Glasgow – UK, simone.stumpf@glasgow.ac.uk)

License  Creative Commons BY 4.0 International license
© Simone Stumpf

Reflecting on the original vision of XAI in 2016, three points were important:

1. providing an *explanation* of a “decision” and the “reasoning” behind it;
2. to increase *understanding* or knowledge of the AI;
3. it should be useful to a “*user*” who doesn’t really have AI knowledge to result in *appropriate trust*


There are nowadays lots of “*technical*” ways to measure XAI explanations (e.g. complexity, fidelity, consistency, etc) but these are described in different terms and measured in different ways, making their consistent application problematic.

Most importantly, many XAI studies are *never evaluated with humans*. We lack consistent human-centered XAI measures, possibly both subjective and objective measurements revolving around:

- Understandability and preferences
- Understanding
- satisfaction
- trust and reliance
- other effects of explanations (e.g. actionability, model improvements, etc)

3.2 Replication in explainable AI: a case study in group recommender systems

Nava Tintarev (Maastricht University, NL, n.tintarev@maastrichtuniversity.nl)

License  Creative Commons BY 4.0 International license
© Nava Tintarev

Joint work of Cedric Waterschoot, Raciél Yera Toledo, Francesco Barile, Nava Tintarev

We have few instances of reproduction or replication studies in XAI. I discuss a series of replication studies using group recommender systems as an application area [1]. I highlight several design considerations including the choice of baseline, experimental procedure (within or between subjects; internal vs external evaluator), and task complexity[1]. I conclude with a brief introduction of a recent study evaluating objective (task performance) and subjective (perceived) understanding of explanations in group recommender systems.¹

This work was led by Francesco Barile.

References

- 1 F. Barile, T. Draws, and O. et al. Incl. Evaluating explainable social choice-based aggregation strategies for group recommendation. *User Model User-Adap Inter*, 34:1–58, 2024.

¹ To appear UMAP’25 [2].

- 2 Cedric Waterschoot, Raciél Yera Toledo, Francesco Barile, and Nava Tintarev. With friends like these, who needs explanations? evaluating user understanding of group recommendations. In *UMAP (to appear)*, 2025.

3.3 A Review of Taxonomies of XAI Evaluation Methods

Timo Speith (University of Bayreuth – Bayreuth, Germany, timo.speith@uni-bayreuth.de)

License © Creative Commons BY 4.0 International license
© Timo Speith

The evaluation of explainable AI (XAI) systems remains a fragmented field, with diverse metrics and taxonomies across the literature. In this talk, I present preliminary insights from a systematic literature review of XAI evaluation methods taxonomies. Across 160 publications, I found nearly 250 properties that were proposed to evaluate XAI systems. Taxonomic efforts often center around the type of evaluation used (human-based vs. mathematical), yet newer approaches emphasize process-oriented perspectives. I highlight the challenges posed by terminological inconsistencies—such as synonymous or overlapping terms and conceptual ambiguities—and propose that evaluating explainability should better attend to the objects of measurement (e.g., understanding of explanation vs. understanding of output vs. understanding of model). This talk aims to contribute to the development of reusable experimental designs by advocating for more coherent evaluation frameworks.

3.4 Doing multiclass Shapley values properly

Peter Flach (University of Bristol, UK)

License © Creative Commons BY 4.0 International license
© Peter Flach

Joint work of Paul-Gauthier Noé, Miquel Perelló-Nieto, Jean-François Bonastre, Peter A. Flach


Main reference Paul-Gauthier Noé, Miquel Perelló-Nieto, Jean-François Bonastre, Peter A. Flach: “Explaining a Probabilistic Prediction on the Simplex with Shapley Compositions”, in Proc. of the ECAI 2024 – 27th European Conference on Artificial Intelligence, 19-24 October 2024, Santiago de Compostela, Spain – Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024), Frontiers in Artificial Intelligence and Applications, Vol. 392, pp. 1124–1131, IOS Press, 2024.

URL <https://doi.org/10.3233/FAIA240605>

Originating in game theory, Shapley values are widely used for explaining a machine learning model’s prediction by quantifying the contribution of each feature’s value to the prediction. This requires a scalar prediction as in binary classification, whereas a multiclass probabilistic prediction is a discrete probability distribution, living on a multidimensional simplex. In such a multiclass setting the Shapley values are typically computed separately on each class in a one-vs-rest manner, ignoring the compositional nature of the output distribution. I gave a brief introduction to *Shapley compositions*, a well-founded way to properly explain a multiclass probabilistic prediction, using the Aitchison geometry from compositional data analysis. In particular, the norm of Shapley decompositions can be used to quantify feature compositions over all classes.

3.5 Explanations as Constructed Arguments

Peter Clark (Allen Institute for AI – Seattle, US, peterc@allenai.org)

License  Creative Commons BY 4.0 International license
© Peter Clark

In this talk I'll offer some perspectives about explanations and their role. I use the following definition: *Explanations are constructions to convey a well-founded argument about why a conclusion is valid.* While a human or machine may arrive at a decision via some opaque method, e.g., with an LLM, we may then *explain* those decisions in a symbolic way, showing how the conclusion systematically follows from facts which the model believes or is provided with. Note that the explanation does not necessarily reflect what the model *did*, but rather why the conclusion is valid. The explanation can be viewed as an orthogonal, but equally valid, way of showing why the model's output is rational given its inputs. In the work my group has been doing, we have been using textual entailment as the formalism for building such chains of arguments, in which a LM first generates an explanation then validates that it itself believes (via self-querying) both the facts and inferences in that explanation – hence it is a “faithful” explanation. If the system's conclusion turns out to be incorrect, we thus now have a way of debugging where the error was (a fact, or an inference) in the system's argument, and potentially correcting that error by updating the model [1]. Four evaluation criteria are useful for such explanations: (a) are the basic facts correct? (b) is the reasoning accurate? (c) Can the user comprehend it? and more generality (d) does the explanation also help the user predict answers to other questions, i.e., has the explanation conveyed a broader “mental model” of the machine? [2, 3, 4]. Argument-based explanations like these are particularly useful for model improvement, as users finally have an interpretable view of what the model “knows” and how that knowledge justifies its conclusions.

Providing a documentation for a Dagstuhl Seminar is mandatory. We focus on talk abstracts and show that a talk abstract can be tagged with co-authors appearing in the joint-work-of-field. Furthermore, a talk abstract can state one main reference on which the talk is based.

References

- 1 Bhavana Dalvi, Oyvind Tafjord, and Peter Clark. Towards teachable reasoning systems: Using a dynamic memory of user feedback for continual system improvement. *ArXiv*, abs/2204.13074, 2022.
- 2 Harsh Jhamtani and Peter Clark. Learning to explain: Datasets and models for identifying valid reasoning chains in multihop question-answering. In *EMNLP*, 2020.
- 3 Peter Clark, Bhavana Dalvi, and Oyvind Tafjord. Barda: A belief and reasoning dataset that separates factual accuracy and reasoning ability. *ArXiv*, abs/2312.07527, 2023.
- 4 Bhavana Dalvi, Peter Alexander Jansen, Oyvind Tafjord, Zhengnan Xie, Hannah Smith, Leighanna Pipatanangkura, and Peter Clark. Explaining answers with entailment trees. In *EMNLP*, 2021.

4 Introduction

Explanations have garnered escalating interest within the AI and Machine Learning (ML) communities. Yet, at times, a crucial aspect that tends to be overlooked is the recognition that explanations can be leveraged for different objects and when evaluating the utility of

these methods the objective needs to be taken into account. Explanations can enhance transparency, help users for a cognitive model of a trained ML system, aid in debugging, or assist users in determining whether to place trust in a prediction or recommendation. While many explanatory mechanisms have been proposed in the community, comparing these solutions remains challenging without the adoption of more standardized practices in terms of evaluation. Compounding this issue is the versatile nature of explanations, meaning algorithm designers in reality should tailor their evaluation strategies to specific tasks. To address this, we build upon the taxonomy presented [15] to identify the different objectives and tasks for explainability methods create guidelines for adaptable tasks and experiments for the community.

Several other taxonomies and frameworks to enable practitioners to develop and evaluate explanations about AI system behaviour [15, 20]. Fundamental questions should be considered when evaluating the impact of an explainability method, what are the **goals or objectives** of the XAI method, what **tasks or usage context** with the method be used for and importantly who are the **stakeholders** of the system [14].

5 Evaluation Criteria for Explanations

To frame our discussion, we began by defining and refining our working definition of evaluation criteria for explainable AI (XAI). Our review of the taxonomy provided by [15] highlighted that, while it serves as a useful starting point, it presents several limitations:

- The criteria are heavily focused on computational or technical aspects, with limited attention to human-centered metrics.
- The taxonomy does not comprehensively map the space of possible evaluation methods.
- Several important criteria are either missing or inadequately defined, including:
 - Trust, Calibrated Trust, and Reliance
 - Human-AI Team Performance
 - Situational Awareness
 - Cognitive Load
 - Explanation Satisfaction
 - Fluency in Human-Autonomy Teaming
 - User Satisfaction
 - Efficiency (e.g., number of explanations/interactions required)
 - Accessibility and Modifiability
 - Distinctions between model and explanation evaluation
 - Calibration and Human-AI Alignment

These gaps reflect a broader issue: current evaluation metrics tend to focus solely on the XAI method itself. Moreover, the criteria are often not well-defined or easily interpretable for HCI researchers and practitioners. Given that usage contexts vary, evaluation criteria must be carefully selected and adapted accordingly. It is rarely feasible to optimize for all criteria simultaneously, making it essential to prioritize based on context. However, the lack of clarity around evaluation criteria makes it difficult to reason about or prioritize trade-offs.

5.1 Self-Reported and Observed Understanding

The idea of *ultra-strong machine learning* suggests that ML systems should not only be able to learn hypotheses in symbolic form but also teach humans about what they have learned, enabling stronger human-AI team performance overall [17]. Explainable AI can help to achieve this end by supporting humans-in-the-loop to develop strong, accurate, and aligned mental models about AI system behavior which enable them to flexibly interact with and apply these systems across all necessary operational contexts. In this sense, one of the main objectives of XAI is to build robust human mental models that facilitate user perception, comprehension, and prediction of AI behavior. In order to know whether explanations have achieved this end, we need a way to assess a human user’s comprehension of a system before and after receiving explanations about its functioning. Thus, one key criteria to consider in assessing the overall efficacy of any given explanation is user understanding. We suggest evaluating understanding through a suite of objective and subjective assessments, which we break into two primary categories: self-reported understanding and observed understanding. The value of assessing understanding both subjectively and objectively is that this allows us to compare a user’s perceived understanding versus their true understanding and whether these two are aligned – in other words, whether understanding is well-calibrated. This is critical as over- or under-confidence could lead to over-use or under-use of this system, hampering team performance overall [18].

5.1.1 Self-Reported Understanding

Criterion: Perceived Understanding

Definition: The extent to which users believe they understand the model, its outputs, and the explanations.

Source: Human

Type: Subjective (e.g., self-reported via questionnaire)

In order to assess self-reported understanding, in line with other subjective assessments from the human factors literature [22], we suggest developing a suite of Likert scale-based questions, which probe users about their individual perceptions of how well they comprehend an AI system given any explanations that they have been provided. While Hoffman et al. have proposed explanation satisfaction and trust scales [10], scales that focus on self-reported human understanding have been underexplored to date.

Previous examples of such questions include [3, 24] (1) I understand how the model works to predict whether a defendant will reoffend [whether the primary tree species in an area is spruce/fir; “I understand the admission algorithm”]; (2) I can predict how the model will behave.

In addition to asking for Likert-based responses to questions from the categories above, it would be additionally useful to ask users to self-report their confidence for each item. We note that items may vary in terms of understanding complexity, and show different patterns in performance across a set of participants [25]. Therefore, we also recommend analysing performance on individual or specific questions rather than computing them on aggregate (e.g., sum of accurate responses).

5.1.2 Observed Understanding

Criterion: Actual Understanding

Definition: The accuracy of a user's understanding of the model, its inputs, outputs, and explanations.

Source: Human

Type: Objective (e.g., mental model elicitation)

The fundamental challenge in assessing user understanding in an objective manner is selecting assessments and metrics that faithfully reflect the user's underlying mental model. Mental models can be shallow, covering only a functional understanding of a system (e.g. a driver knows how to operate a car), or deep, achieving a more structural understanding of how a system functions (e.g. a mechanic understands the inner-workings of a car and how to fix broken cars or extend their capabilities) [13]. Importantly, the appropriateness of such mental models depends on a user's context, including their attributes and expertise, tasks, use cases, and goals. Previous work has proposed a mental model soundness score which addresses these factors and incorporates domain-specific comprehension questions [13].

One structured approach to determine a user's goal-informed informational needs within a given context is based on the situation awareness (SA) framework from the human factors literature. Endsley defines SA as the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status into the future [7]. SA requirements for a given context define a person's informational needs which can be met in part through the application of XAI when a human is working with an AI system within their context. Sanneman and Shah apply the SA framework to XAI, and define the following three levels of XAI [20]:

- Level 1: XAI for Perception—explanations of what an AI system did or is doing and the decisions made by the system
- Level 2: XAI for Comprehension—explanations of why an AI system acted in a certain way or made a particular decision and what this means in terms of the system's goals
- Level 3: XAI for Projection—explanations of what an AI system will do next, what it would do in a similar scenario, or what would be required for an alternate outcome

A process such as goal-directed task analysis (GDTA) can be applied to elicit a user's informational needs at these three levels [8], applicable explanations techniques can be selected to meet these needs, and their efficacy at supporting a user's mental model can be assessed by applying a technique such as the situation awareness global assessment technique (SAGAT), a validated SA assessment from the human factors literature [7, 19].

There are various other models. Speith et al. [21], for example, generalize Sannemann and Shah's model by incorporating findings from various disciplines outside of human factors (e.g., cognitive psychology, philosophy). Their model comprises six levels of skills that can be tested in studies and are said to correlate with varying degrees of understanding. Their aim is to make studies more comparable.

5.2 Explanation Fidelity and Stability

After generating an explanation, the first step is to assess its correctness. To this end, we identified two key properties that allow for a technical evaluation of explanations, namely explanation fidelity and stability. According to the state of the art, these criteria are essential

for assessing the reliability and trustworthiness of explanations, although their formalization and practical application are still challenging [2, 9]. In particular, the literature offers many different definitions of fidelity, faithfulness and correctness, each leading to distinct technical implementations, depending on the context, kind of data, ML model and explanation. Given the complexity of this setting, our first priority is to verify the technical validity of the explanation. Our rationale is that, before evaluating the quality of an explanation through user studies, we must first ensure that it is mathematically sound.

To this end, we consider an ML model $b(\cdot)$ that has already been trained on a dataset D_{train} and exhibits good predictive performance, with generalization capabilities and no signs of overfitting. To explain our model $b(\cdot)$, we consider an explanation method $g(\cdot)$ that outputs explanations e . At this stage, we intentionally keep the definitions at a high level to allow for the definition of criteria that are general and not tailored to a specific ML task or explanation type. As such, the explanation in this setting may take various forms, including feature importance, decision rule, or even a chain of thoughts. The same applies to the ML task.

5.2.1 Explanation fidelity

Criterion: Fidelity

Definition: The degree to which the explanation accurately reflects the model’s decision-making process.

Source: Model (may require human-verified ground truth)

Type: Objective

We first address the definition of explanation fidelity (or faithfulness). The objective in this case is to evaluate whether an explanation accurately reflects the internal reasoning of the model. We consider this an objective property, since it concerns the alignment between the explanation and the model’s actual behavior, independently of any human interpretation. The definition of explanation fidelity is not straightforward. In the literature, we can find many terms, including faithfulness and correctness. However, faithfulness may be misleading due to its connotation of belief or trust, which does not align with the technical nature of the concept. In addition, its definition is highly task-dependent and varies significantly depending on the type of explanation considered. For instance, when considering feature importance-based explanations, faithfulness can be evaluated by masking the most important features at prediction time, but this only applies for this specific kind of explanations. Rather than relying on a single notion of fidelity, a useful perspective may come from related concepts such as *comprehensiveness* and *sufficiency* [6]. A comprehensive explanation contains all the critical components that influence the prediction, while a sufficient explanation identifies the minimal set of elements necessary to reach the same outcome. In particular, an explanation makes claims about the factors that cause or influence the model’s prediction. Therefore, the main objective of *explanation fidelity* is to verify whether these factors are truly influential. In practice, if a change in the explanation leads to a change in the prediction, this supports the causal relevance of the explanation’s components. As already mentioned, the technical implementation of this criteria can be challenging given the different ML tasks available, as well as the various kind of explanations. However, the evaluation can be approached by altering the components identified in the explanation, such as masking an important feature in a classifier or removing a rule in an expert system, and observing whether the model’s output changes accordingly.

5.2.2 Explanation stability

Criterion: Stability

Definition: The consistency of explanations for similar inputs or outputs.

Source: Model / Human

Type: Objective / Subjective

Another important aspect of the explanation is its stability. Also in this case we can find different terms, such as consistency or robustness, and many definitions. After our discussion, we claim that stability refers to the consistency of a model's predictions and explanations when provided with similar inputs. A stable explanation method should produce similar outputs and explanations for inputs that are close according to a similarity metric. To set the stage, we can consider two similar records, x and x^1 that give the same output, $b(x) = b(x^1)$. In this case, we expect two similar explanations, e and e^1 . Therefore, the evaluation requires two components: a metric for measuring similarity between input instances x and x^1 , and a metric for comparing the corresponding explanations e and e^1 . This property can be assessed through objective criteria, but we prefer to consider it as an *objective and subjective* criteria. In fact, when defining what constitutes *similar* inputs or explanations, it may also involve human intervention to define or validate similarity measures. The specific approach to measuring stability can vary depending on the task at hand and the type of explanation being considered.

5.3 Trust, Reliance and Performance

It is necessary to clearly distinguish between “trust” and “reliance” in a system. Although these concepts are interconnected, they are fundamentally separate [1]. Reliance is a behavior when adopting the recommendations of the system or delegating (sub)tasks to it, while trust is considered a relational process between a trustor and a trustee in a specific context with a trust goal [1]. Reliance can indicate trust, but does not entail it. While reliance on a system can be observed and measured in different ways, there is no distinct way to measure users' trust in a system.

5.3.1 Performance

Criterion: AI-Human Team Performance

Definition: The overall performance of decisions made by the AI-human team.

Source: Human

Type: Objective (e.g., task performance)

Although the performance of a Human-AI team is often the primary goal in most applications and heavily affected by users' trust and reliance, the measurement of performance is left here vague on purpose, as it is very task specific.

You would use the same metric that you would use to evaluate the model in isolation (or the user in isolation). For instance, in model debugging performance means the quality of the model one obtains, whereas in a classification task it is rather the accuracy the joint performance of human and system.

5.3.2 Self-reported Trust

Criterion: Self-Reported Trust

Definition: How much the user reports trusting the AI system.

Source: Human

Type: Subjective (e.g., questionnaire)

Self-reported trust is the extent to which a user believes they trust in the AI model. There is a multitude of questionnaires to measure self-reported trust (see [12] for an overview). Some of the frequently used questionnaires may not be appropriate in every situation. For example, the questionnaire by Hoffman et al. [10] tends to include concepts different from trust.

Additionally, self-reported trust can be measured by means of betting markets, i.e., how much the participants are willing to bet on the model giving a correct output in situations without having personal stake.

Additional thoughts on self-reported trust in a specific instance:

- possibly only for experts, as laypeople might not distinguish between the system and its individual explanations
- maybe the very same metrics apply (e.g. trust in automation), with some customization

5.3.3 Observed Reliance

Criterion: (Appropriate) Reliance

Definition: The extent to which users appropriately follow correct or

Source: Human

Type: Objective (e.g., behavioral observation)

Methods to measure observed reliance include:

1. Investment games in which the participant initially has several points that can be used to bet on the AI system. The number of points they are willing to spend indicates the reliance on the system. Two interesting resources for this kind of task are: [16] for general for general XAI models and [11] for explainable reinforcement learning.
2. Stakes scenarios that investigate the behavior in different settings, e.g., using the system to recommend a movie vs. to diagnose a life-threatening illness
3. Delegation tasks that examine in which cases the users entirely delegate a (sub)task to the system. These can be augmented with betting.
4. Measurement of “switch rates” [26]: How often and when do users switch to the system’s recommendation in case it deviates from their own initial assumption or outcome? This can be designed as a multi-step approach where the user makes an initial assumption and then is presented the system’s recommendation. In case the user and system results are different, the participants in a final step have to decide whether they keep their own result or adopt the system’s recommendation.
5. Willingness to follow advice: how much do people deviate towards the algorithmic estimate based on their own estimate. This approach is similar to the switch rate task but applied to numeric decisions.

6 Relationship between Usage Context and Criteria

As highlighted by [15] different criteria become more relevant in assessing explanations depending on the target usage context. In the following sections we begin to reason about and prioritise the most relevant criteria and measurements for a subset of the usage contexts.

6.1 Metrics for Model Improvement

Besides helping the user, one important objective of explanation is to help experts and/or system builders **improve the problem-solving system** (“model”, henceforth) itself. Such model improvements can occur:

- During model development, to inspect how the model could be improved and make such improvements
- after model deployment, to verify if the model is behaving as intended or needs further improvement

There are numerous mechanisms that can be employed to modify model behavior, e.g., adding extra training data and retraining; modifying a rule or concept in a rule-based component; changing weights on features; masking out elements known to be irrelevant to a result. The role of explanation is to help the expert/system designer understand why a model produced a wrong answer, and what kinds of interventions might correct the error both for a specific case and future cases. This whole endeavor is not just about producing good explanations to help in this process, but also designing a model architecture in the first place that supports such explanations and allows easy model improvements – the technology of interactive explainable AI (XAI) [23]. Note that model improvement may occur both

6.1.1 Relevant Metrics

How can we measure whether explanations help experts/designers improve models? We consider two types of measurements:

Primary Measures.

We identify several **primary measures** that can be used to directly measure model improvement:

- Performance, e.g., accuracy, performance of the human AI team
- Quality of the explanations (that aid in the end goal of improving performance)). Measures of quality (defined and described in more detail elsewhere in this document) include:
 - Stability
 - Faithfulness
 - Understanding
 - Contextless (knowing the bounds or limitations of the explanations e.g. where it would not hold)
 - Action-ability

Note that simply measuring the change in such metrics before/after a model update is not sufficient to show that explanations help. Rather, the experiment should compare improvement without/with explanations helping the person improving the model.

Secondary Measures.

In addition, there are some **secondary measures** that do not directly measure performance improvement, but are likely correlated with it and desirable to also improve (and at least observe):

- Change in trust
- Faithfulness

Finally we note that other measures, e.g., end-user satisfaction, are less relevant for the specific goal of model improvement (though clearly critical for the end system).

6.1.2 Trust and Model Improvement

There is a delicate relationship between trust and model improvement. If a domain expert is involved, then exposing them to model errors may reduce their trust in the system. Conversely, if they have been involved in improving the model, then this may help increase their trust, and in fact involving domain experts may ultimately help them appropriately calibrate the right level of trust they should have in the system's behavior. This is an important aspect to track and study in model improvement experiments, even though it is not the primary objective.

6.1.3 Sources of Model Misalignment

Training data misalignment Data misalignment between training data vs test (debugging) data Fundamentally missing from the training data (vs complete ground truth of all the possible data) Improving alignment between the AI model and expert user's knowledge on the task Example based correction: by explaining the missing examples, one can point out which part of data is missing.

Why might a system be making mistakes in the first place? It is useful to consider two dimensions of misalignment (between the actual model and the ideal/perfect target model):

Training data limitations.

In a machine learning context, models are only exposed to a sample (namely, the training data) taken over the distribution of problem-solving tasks, and thus the learned model may be somewhat misaligned with the actual ideal model. While we do not have access to that ideal model (ideal ground truth), we approximate this by using an independent, hidden test set (a sample of that ideal ground truth), to measure model performance.

Domain expertise.

There may also be additional knowledge beyond that captured in examples that is relevant to the task, again contributing to a misalignment between the actual model and an ideal oracle model. The model improvement process provides an opportunity for experts to inject that extra knowledge into the system, e.g., by providing additional examples for areas of the problem space that the model is either ignorant of or unsure about.

6.2 Metrics for Capability Assessment/Auditing

An important usage context of explanations is the assessment of a system's capabilities (e.g., fairness, safety, performance), also known as *auditing*. In general, a capability is part of the system; and there are specific criteria that the model has to satisfy to have a certain

capability. Auditing, then, is the way to find out whether the system satisfies specific criteria and, thus, has the corresponding capabilities. As explanations can help to find out whether specific criteria are satisfied, they are a means to auditing.

Since most of these criteria are those that concern the whole system, global explanations are most helpful for auditing. Nevertheless, the usefulness of local explanations should not be overlooked, as they can indicate that something is not correct and can be the base for further deliberations on a capability of interest. If explanations reveal that some criteria for a system capability are not met (e.g., for fairness that no protected attributes unduly influence decision-making), then the system is flawed in this respect and mitigation strategies must be initiated.

In general, it is quite open who can be the auditor. Obvious possibilities are external (accredited) watchdog organizations (such as the TÜV in Germany), but also regulators and interested individuals. The only requirement is that they are as objective as possible with regard to the capability they are auditing. An example of an interested individual who does audits is the developer who aims at a high accuracy of the system.

Audits can take place at various points in time. They can be conducted regularly, for example when expiring certificates need to be renewed or when required by regulations, or only when it is discovered that something has gone wrong. In the latter case in particular, the affected party is usually left out of the loop because they either did not realize that they had been negatively affected (which is often the case with discrimination, for example) or do not have the means to defend themselves against a false assessment. This raises the question of how affected parties can be better involved in the auditing process. Related to this question, but also going beyond it, is the question of which explanations are most useful at what part of the process. Accordingly, the measures for evaluating these explanations are also diverse.

In the case of auditing, metrics such as (human-AI) performance and reliance are not relevant. The most important metrics are:

- (actual) understanding: the auditor should understand the system
- fidelity and stability: the explanations should reliably and truthfully track the system's decision-making processes.
- coverage (i.e., the distribution of explanations): the explanations should cover as many cases as possible.
- self-reported trust: the auditor should believe that they trust the system.

However, auditing using XAI can be difficult in many cases if the important criteria can only be checked when the system is in use or are even outside the (usual) realm of XAI. An example of this is security, where the provenance of the training data is also important. Another example is fairness [4, 5]. Especially when it comes to fairness, a mere consideration of outcome fairness is often not enough [5]. One reason for this is that there is no ground truth in certain areas. In the case of loan applications, for example, there is only partial ground truth data, as it is never known whether someone who has been refused a loan would not have repaid it after all.

Furthermore, there are various types of fairness that can also be important. Informational fairness (which is not part of the model), for example, deals with the question of what information a particular party has received about a process and whether this information is sufficient, faithful, and adequately prepared. Procedural fairness, on the other hand, asks whether the decision-making process itself is designed in such a way that it leads to fair outcomes. Both are types of fairness whose fulfillment cannot be determined by traditional XAI explanations. Accordingly, auditing requires explanations that XAI does not yet provide.

7 Case Study

In this section, we discuss an end-to-end pipeline for designing an XAI experiment. The participants broke up into groups and collaboratively developed concrete evaluation scenarios across different application domains. Each group explored how to operationalise key explainability criteria within their context, identified appropriate stakeholders and metrics, and proposed experimental designs to assess the impact of explanations on human-AI interaction.

7.1 Model Improvement (Medical Domain)

We focus on image-based classification task in the medical domain. Specifically, we discussed classification of tumors or skin lesions. The goal is to classify an image based on the tumor type detected in the image. There are five tumor types considered in this example, with values 1 – 5. The label indicates the progression of the tumor, with 1 being the least and 5 the most dangerous. The challenge of classification problem is in distinguishing between similar classes, specifically discerning between types 2 and 3 of tumors.

7.1.1 Stakeholders

We distinguish between stakeholders and their roles in pre and post deployment scenarios. We identify the main roles from the perspective of model improvement task as following:

1. Pre-deployment:
 - a. Model developers: the goal is to improve the model for deployment
 - b. Domain experts: might be consulted during pre-deployment to add expert knowledge to the model. Domain experts could spot model errors and gaps, identify corrections, provide labels, annotations, and data in general
2. Post-deployment:
 - a. System user: with the goal of performing the end task
 - b. Model auditor:

7.1.2 xAI Pipeline

AI Model. To implement image-based classification, we focus on the following two classification models:

1. CNN – representing a black-box approach to image classification task.
2. Concept bottleneck model – capable of providing more high-level concept explanations.

xAI Methods. To generate explanations for the AI model’s decisions we discussed the following explanation methods:

1. Saliency maps could identify the parts of the images that led the AI to make a specific decision. Could be especially interesting to uncover spurious correlations in data.
2. Concept level explanation: these could be a result on the concept bottleneck models.
3. Example based: to explain a current decision, a previous example where the same decision was reached could be informative.
4. Prototypical: explains decisions by offering “prototypes” of different classes (e.g. the input is classified as y because it looks like the representative of this class).
5. Counterfactual/Contrastive (near miss example): could highlight the nuances between similar classes for critical decisions on the decision boundary.

Data Collection and Preparation. To train the AI model a dataset of labeled images from this medical domain is required. The features and labels in the dataset can be verified by domain experts during pre-deployment. The data should be split into three sections – train, validation and test. A (potentially flawed) model is then trained on the train dataset. The validation dataset is presented to the user or domain expert who can advise on potential corrections. Then the retrained model (taking into account user/domain expert’s feedback) is evaluated on the test split.

Additionally, at this point it might be advisable to identify the type of errors on which the model improvement task focuses. These can be a consequence of known spurious correlations in the data, missing or incorrectly labeled data.

xAI Study Data Collection and Preparation. A dataset of 20 images can then be selected to be used in the xAI user study. As the goal of the study is to investigate model improvement, the model should likely make mistakes on some of the presented input images. However, a large number of mistakes could quickly lead to deterioration of user trust and impact the measurements of other xAI metrics. The advised number of mistakes on a dataset of 20 images is 3.

xAI Study Design. The study should begin by informing the participants of the domain problem, the AI model, explanations (depending on the study condition) and their task in correcting the model outputs. It is advisable to inform the participants that the AI system is well trained, however, it can also make mistakes. Otherwise, users might quickly lose trust in the system after observing errors. However, it is not clear should the reported accuracy equal the real accuracy or be set to a predefined value (e.g. 95%).

To investigate if the expert’s corrections are a consequence of the presented explanations, it is likely that a baseline condition is needed where experts are asked to correct the model behavior without access to explanations. Furthermore each explanation type is going to elicit another study condition.

xAI Metrics. The following metrics were discussed to evaluate the quality of explanations for the model improvement task:

1. Human-AI performance: measures the accuracy of the expert who can receive recommendations on the decision. Does the expert follow the model’s outputs or makes its own decisions? In the ideal scenario, the expert agrees with the model when the model is correct but corrects it when it makes an error.
2. Model performance after corrections: after receiving expert’s corrected labels and retraining the model, does the performance (on some metric like accuracy) improve?

7.2 Capability Assessment (Hiring)

We considered capability assessment – and specifically fairness assessment – of an AI system for hiring. The system is intended to pre-select or screen promising candidates out of a pool, so that successful applicants will later on undergo a hiring interview. This can be most naturally be cast as a binary classification task: “pass” vs “fail”.

7.2.1 Stakeholders

In order to get a sense of what criteria and metrics would be useful for the specific use case, we consider different users that might be interested in assessing fairness of this AI, and specifically:

- The *job applicant*: they are trying to get a job and they presumably receive a pre-screening output from the AI. Naturally, it is in their interest to verify that the AI is indeed fair.
- The *hiring manager* of the company offering the job.
- The “*watchdogs*”, e.g., the ethics department of a company, the union, etc.
- (Optionally) The *job centre personnel*, who are responsible for facilitating job applications.

7.2.2 Detailed Setup

We assume the model is a machine learning classifier trained to discriminate between promising candidates and the rest on historical data. The training examples and the input instance would consist of text records (e.g. CV, education, prior positions) either pre-processed into or paired with tabular data (e.g., personal information). We do not focus on a specific classifier architecture, for two reasons. First, we assume the classifier has been trained – presumably by either the company or the job centre – for good performance out of the many options available. Second, many high-quality explainability techniques are model agnostic and can provide competitive explanations for a variety of classifier architectures. Of course, classification performance should be tracked (for instance, via model accuracy or F_1 score) for consistency with the primary goal of selecting promising candidates. It is a prerequisite that the classifier achieves non-trivial prediction accuracy.

We would expect four possible kinds of information would be of interest for assessing fairness for the four stakeholders we consider:

- Prediction confidence – useful across the board.
- Feature relevance – this is especially useful for hiring managers and watchdogs for understanding whether the model is, e.g., leveraging protected attributes for its decisions.
- Counterfactual explanations – these are especially useful for the applicant (and potentially the job centre facilitator) so as to gain actionable insights about the decision.
- Prototype-based explanation, e.g., distance from “ideal” applicants – these *might* be useful to get a sense what kind of profile(s) the classifier is expecting successful applicants to have, and whether these are in any way undesirable.

7.2.3 Evaluation Metrics

- *Perceived prediction quality and fairness*: this is the basic capability being assessed.
- *Satisfaction with explanation*: whether explanations are perceived as useful for assessing fairness.
- *Actual understanding*: whether explanations have been in fact understood by stakeholders (rather than merely perceived as useful).
- *Self-reported trust*: whether the stakeholder believes they can rely on the AI doing a good job at generating fair predictions.
- “Correctness” of explanations, and specifically:
 - *Fidelity*: lack of fidelity means that it may be impossible to map poor justifications for the predictions (e.g., reliance on protected attributes) to the model’s actual behavior and capabilities.
 - *Stability & Coverage*: feature relevance explanations only provide a local, per-candidate view of the model’s reasoning; this does not necessarily generalize to other instances unless the explanations are somehow stable, i.e., do not vary enormously for similar candidates (and potentially decisions); coverage refers to the fact that in order

to get a sense of the overall capabilities of the classifier as a whole, for all possible instances, it may be necessary to obtain local explanations for a sizeable number of individual cases, for statistical reasons.

All these metrics can be computed mechanically without users studies.

7.2.4 Evaluation

The idea is to use a between-participant experimental design to assess the relative performance of different UI designs. We suggest to focus on a total of 6 designs, one for each combination of explanation type (3 total) and (2 total).

An online study seems to be sufficient for evaluating the AI and its explanations given the chosen metrics and setup. The task could be briefly summarized as “look at UIs and then complete a questionnaire: how well do you think it does its job, given your specific role?”

The question is how to evaluate the UIs from different user perspectives, and specifically how to carry out the recruiting. Ideally, we would need a reasonably large sample of applicants, job centre personnel, hiring manager, and watchdogs. This immediately poses the issue of how to get access to such a sample. Naturally, power analysis can help to identify a sufficient sample size. For certain users – e.g., applicants – one could implement a role playing setup in which Prolific participants are asked to act as applicants and evaluate the AI system under the aforementioned metrics, obtaining feedback via questionnaires. This is more problematic for specialized users like hiring managers, who might be more difficult to simulate or role play properly.

7.2.5 Hic Sunt Leones

Mental model /situational awareness extraction framework
Questionnaire about perceived fairness / trust
Satisfaction with explanations
Open text response: any other info that would have helped you assess?
Demographics

7.3 Decision Support (ICU Triage)

In this section we describe an example evaluation methodology for decision support systems.

To ground the discussion, we chose an example use-case instead of describing a general decision support problem.

7.3.1 Chosen use case for the evaluation

For the use-case, we decided on a medical diagnosis task. Specifically, a situation where a healthcare professional has to make a single decision. An example could be ICU triage – one single patient where the decision has to be made whether the person needs to wait or needs emergency treatment now. So, we can think of a single nurse that is working during a particular shift and one ICU bed has opened up, who of the top 5 patients at risk do they admit.

7.3.2 Things to think about that you might need for your evaluation

For our use-case we mainly need to things. Medical Use Case/Data: We need a medical use case given by patient data. This data normally consists of both self-reported (interview) data and objective measurements (vitals). The modality of this data is often very multi-modal including text, images and tabular time series (EHR). Ideally, we would like to get this data from use case studies for medical training.

Model: A predictive AI Model that outputs urgency rankings for each patient. Based on this, the model gives a ranking where, e.g., Person A is more urgent than Person B.

7.3.3 Chosen sample explanation method

To ground the discussion some more, we imagined two example explanations we want to compare. Because the task is comparative (which of the participants do you admit to the ICU), we chose two contrastive explanations. First, counterfactual explanations that change the input such that the model rates Person B as more urgent than Person A. Second, feature attribution that show how relevant each input was for the AI's decision to assigning more urgency to Person A than Person B. In general, our evaluation example focuses on local explanations.

7.3.4 Participants

Because medical background knowledge is crucial to properly interpret the explanations we would aim to recruit healthcare professionals for our user study. In particular, for our use case, we would try to recruit nurses.

7.3.5 Task Setup

At first, the participants are given a description of the task and of the AI and Explanations methods that they will see during the study. [optional:] They will also get a pre-questionnaire about their demographics, medical expertise and previous experience with AI and XAI.

Afterwards, the participants will be given X decision tasks where they are told that one ICU bed opened up and they have to decide which of 10 patients they admit. First, they have to decide by themselves, without AI support, which patient they admit. After that, they see an AI Recommendation for the urgency ranking of the participants. Depending on the condition, they will also see an explanation here.

After X decisions, they will [optional: move to the understanding task and then] get a pos-questionnaire.

7.3.6 Conditions

We envision a between subject design.

- Baseline Condition: These participants only see the recommendation of the AI.
- Explanation Condition 1: These participants see the AI recommendation together with the first explanation method – in our case counterfactual explanations.
- Explanation Condition 2: These participants see the AI recommendation together with the second explanation method – in our case contrastive feature attribution.

7.3.7 Metrics

We will measure two tothree observed (or objective) metrics:

- **Performance:** Did nurses/doctors select the correct patient? And also interesting: how far is the distance from the 1st rank is the admitted person (relative performance). To incorporate the possibility of the AI making mistakes we will assume a somewhat realistic accuracy of 80%. That means that in 80% of the example decision, the AI will be correct. We do not choose a lower rate, since it might unrealistically bias the participants against the AI. However, in addition to the general performance, we will also reported individual performance for the group of correct and incorrect AI predictions.

- **Appropriate AI Reliance:** To what extent did people deviate towards the AI advice, given that they gave their initial estimate, then received the AI's advice, and based on the latter, decided to comply (or not – or to what extent) with the AI's estimate (“weight on advice”)

Appropriate reliance = AI was recommending patient #1 ranking, and the user was following it (final user estimate = AI estimate = best) Underreliance = AI was recommending patient #1 ranking but the user did not follow it (final user estimate \neq AI estimate & AI estimate = best)

Overreliance = was NOT recommending patient #1 ranking and the user followed it (final user estimate AI estimate & AI estimate \neq best)

Special case = doctor's initial estimate = best = AI estimate \rightarrow “reinforced” appropriate reliance
- **Observed Understanding [Optional, not as crucial for this task]:** After the X decisions, we could add a prediction task as proposed by [10]. In this task, participants will see Y additional examples. Here, they will only see the input and the explanation (or no explanation for the baseline). Based on this they have to predict the urgency ranking of the AI. Depending on how good they are at predicting the AI will be used to judge their understanding of the AI models reasoning.

Additionally, we will measure several self-reported (subjective) measures in the post-questionnaire:

- Self-reported Trust: Trust in Automation (e.g., Perceived Competence, \rightarrow Madsen & Gregor, 2000)
- Perceived Understanding
- Perceived Helpfulness
- Satisfaction with Explanation
- User experience questionnaire? (Maybe)
- Task Load / Perceived Accomplishment

8 Conclusion

This seminar marked an important step in bridging the gap between human-computer interaction (HCI) and artificial intelligence (AI) research communities. By bringing together experts from both fields, we created a multidisciplinary forum that encouraged critical reflection on the goals, assumptions, and evaluation methods of explainable AI (XAI). Discussions focused not only on technical soundness but also on human-centered evaluation and usability in real-world contexts. Through collaborative case studies, taxonomy refinement, and shared methodological frameworks, we initiated a dialogue that will continue to shape the development of robust, transparent, and user-aligned AI systems. This seminar laid the groundwork for ongoing collaboration between HCI and AI researchers, emphasizing the importance of inclusive, reproducible, and context-aware evaluation practices in the evolving landscape of responsible AI.

References

- 1 Michaela Benk, Sophie Kerstan, Florian von Wangenheim, and Andrea Ferrario. Twenty-four years of empirical research on trust in ai: a bibliometric review of trends, overlooked issues, and future directions. *AI & SOCIETY*, October 2024.

- 2 Francesco Bodria et al. Benchmarking and survey of explanation methods for black box models. *Data Mining and Knowledge Discovery*, 37(5):1719–1778, 2023.
- 3 Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O’connell, Terrance Gray, F Maxwell Harper, and Haiyi Zhu. Explaining decision-making algorithms through ui: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–12, 2019.
- 4 Luca Deck, Jakob Schoeffler, Maria De-Arteaga, and Niklas Kühl. A critical survey on fairness benefits of explainable AI. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2024, Rio de Janeiro, Brazil, June 3-6, 2024*, pages 1579–1595. ACM, 2024.
- 5 Luca Deck, Astrid Schomäcker, Timo Speith, Jakob Schöffler, Lena Kästner, and Niklas Kühl. Mapping the potential of explainable artificial intelligence (xai) for fairness along the ai lifecycle. In Mattia Cerrato, Alesia Vallenias Coronel, Petra Ahrweiler, Michele Loi, Mykola Pechenizkiy, and Aurelia Tamò-Larrieux, editors, *Proceedings of the 3rd European Workshop on Algorithmic Fairness, Mainz, Germany, July 1st to 3rd, 2024*, volume 3908 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2024.
- 6 Jay DeYoung, Sarthak Jain, Nazneen Fatema Rajani, Eric Lehman, Caiming Xiong, Richard Socher, and Byron C. Wallace. ERASER: A benchmark to evaluate rationalized NLP models. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 4443–4458. Association for Computational Linguistics, 2020.
- 7 Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995.
- 8 Mica R Endsley. Direct measurement of situation awareness: Validity and use of sagat. In *Situational awareness*, pages 129–156. Routledge, 2017.
- 9 Peter Hase and Mohit Bansal. Evaluating explainable AI: which algorithmic explanations help users predict model behavior? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, 2020.
- 10 Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. Measures for explainable ai: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance. *Frontiers in Computer Science*, 5:1096257, 2023.
- 11 Tobias Huber, Maximilian Demmler, Silvan Mertes, Matthew L. Olson, and Elisabeth André. Ganterfactual-rl: Understanding reinforcement learning agents’ strategies through visual counterfactual explanations. 2023.
- 12 Spencer C. Kohn, Ewart J. de Visser, Eva Wiese, Yi-Ching Lee, and Tyler H. Shaw. Measurement of trust in automation: A narrative review and reference guide. *Frontiers in Psychology*, 12, 2021.
- 13 Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. Tell me more? the effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 1–10, 2012.
- 14 Q Vera Liao, Milena Pribić, Jaesik Han, Sarah Miller, and Daby Sow. Question-driven design process for explainable ai user experiences. *arXiv preprint arXiv:2104.03483*, 2021.
- 15 Q Vera Liao, Yunfeng Zhang, Ronny Luss, Finale Doshi-Velez, and Amit Dhurandhar. Connecting algorithmic research and usage contexts: a perspective of contextualized evaluation for explainable ai. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 10, pages 147–159, 2022.
- 16 Tim Miller. Are we measuring trust correctly in explainability, interpretability, and transparency research? *arXiv preprint arXiv:2209.00651*, 2022.
- 17 Stephen H Muggleton, Ute Schmid, Christina Zeller, Alireza Tamaddoni-Nezhad, and Tarek Besold. Ultra-strong machine learning: comprehensibility of programs learned with ilp. *Machine Learning*, 107:1119–1140, 2018.

- 18 Raja Parasuraman and Victor Riley. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2):230–253, 1997.
- 19 Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of cognitive engineering and decision making*, 2(2):140–160, 2008.
- 20 Lindsay Sanneman and Julie A Shah. The situation awareness framework for explainable ai (safe-ai) and human factors considerations for xai systems. *International Journal of Human–Computer Interaction*, 38(18-20):1772–1788, 2022.
- 21 Timo Speith, Barnaby Crook, Sara Mann, Astrid Schomäcker, and Markus Langer. Conceptualizing understanding in explainable artificial intelligence (XAI): an abilities-based approach. *Ethics Inf. Technol.*, 26(2):40, 2024.
- 22 Neville A Stanton, Paul M Salmon, Laura A Rafferty, Guy H Walker, Chris Baber, and Daniel P Jenkins. *Human factors methods: a practical guide for engineering and design*. CRC Press, 2017.
- 23 Stefano Teso and Kristian Kersting. Explanatory interactive machine learning. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 2019.
- 24 Xinru Wang and Ming Yin. Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*, pages 318–328, 2021.
- 25 Cedric Waterschoot, Raciél Yera Toledo, Francesco Barile, and Nava Tintarev. With friends like these, who needs explanations? evaluating user understanding of group recommendations. In *UMAP (to appear)*, 2025.
- 26 Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20*, page 295–305, New York, NY, USA, 2020. Association for Computing Machinery.

Participants

- Elisabeth André
Universität Augsburg, DE
- Jaesik Choi
KAIST – Daejeon, KR
- Peter Clark
Allen Institute for AI –
Seattle, US
- Elizabeth M. Daly
IBM Research – Dublin, IE
- Peter Flach
University of Bristol, GB
- Jasmina Gajcin
IBM Research – Dublin, IE
- Tobias Huber
TH Ingolstadt, DE
- Eda Ismail-Tsaous
bidt – München, DE
- Patricia Kahr
TU Eindhoven, NL
- Francesca Naretto
University of Pisa, IT
- Talya Porat
Imperial College London, GB
- Daniele Quercia
Nokia Bell Labs –
Cambridge, GB
- Lindsay Sanneman
Arizona State University –
Tempe, US
- Ute Schmid
Universität Bamberg, DE
- Kacper Sokol
ETH Zürich, CH
- Timo Speith
Universität Bayreuth, DE
- Wolfgang Stammer
TU Darmstadt, DE
- Simone Stumpf
University of Glasgow, GB
- Stefano Teso
University of Trento, IT
- Nava Tintarev
Maastricht University, NL

