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.

Report from Dagstuhl Seminar 21361

Extending the Synergies Between SAT and Description Logics

Edited by

Joao Marques-Silva[1], Rafael Peñaloza[2], and Uli Sattler[3]

1   CNRS – Toulouse, FR, `joao.marques-silva@irit.fr`
2   University of Milano-Bicocca, IT, `rafael.penaloza@unimib.it`
3   University of Manchester, GB, `uli.sattler@manchester.ac.uk`

──── **Abstract** ────────────────────────────

This report documents the program and the outcomes of Dagstuhl Seminar 21361 "Extending the Synergies Between SAT and Description Logics". Propositional satisfiability (SAT) and description logics (DL) are two successful areas of computational logic where automated reasoning plays a fundamental role. While they share a common core (formalised on logic), the developments in both areas have diverged in their scopes, methods, and applications. The goal of this seminar was to reconnect the SAT and DL communities (understood in a broad sense) so that they can benefit from each other. The seminar thus focused on explaining the foundational principles, main results, and open problems of each area, and discussing potential avenues for collaborative progress.

1   Executive Summary

Rafael Peñaloza (University of Milano-Bicocca, IT)
Joao Marques-Silva (CNRS – Toulouse, FR)
Uli Sattler (University of Manchester, GB)

About the Seminar

Propositional satisfiability (SAT) and Description Logics (DL) are two successful areas of computational logic where automated reasoning plays a fundamental role. Seen from a very abstract level, they can be thought as being part of the same family of logical formalisms attempting to represent knowledge from an application domain, and differentiated only by their expressivity and correspondent trade-off in reasoning complexity. However, the evolution of the two areas has diverged, mainly due to differences in their underlying goals and methods. While the DL community focused on introducing and fully understanding new constructors capable of expressing different facets of knowledge, the SAT community built highly-optimised solvers targeted for industrial-size problems.

Some recent work has permeated the boundaries between the two communities. It has been shown that some DL reasoning problems could be reduced to known SAT-related tasks. In turn, these reductions motivated new optimisations targeted to the specific shape of the problems constructed by them. The goal of this seminar was to bring together researchers from both communities to foster a deeper collaboration and mutual development. The primary goals were (i) to understand the tasks and methods from one community which could benefit the other, and (ii) to discuss the policies used within the communities to encourage specific advancements.

A relevant issue considered is how to promote the development and testing of DL reasoners. To try to answer this, we discussed the status of benchmarks in both communities, and the success stories from SAT competitions. A salient point was the issue, from the DL point of view, of the many variants that should be evaluated – from the different languages, to the reasoning tasks considered. However, recent SAT competitions have also successfully handled many categories. One possible explanation for the wide availability of solvers capable of handling practical extensions of SAT (like MaxSAT and QBF) is the existence of solvers like MiniSAT, which allow for fast prototyping using SAT solvers as oracles. No analogous tool is available for DL reasoners.

The remaining of the seminar focused on novel and timely tasks which are currently under development in both communities, and where the best possibilities for collaborations are foreseen. Among them, we can mention methods for explaining the result from a solver, and proofs which can be used to automatically verify their correctness. We noted that the notion of an *explanation* is too wide, allowing for different interpretations which were presented as talks during the seminar. Each of these interpretations gives rise to distinct techniques. But interestingly, the core ideas are not necessarily specific to SAT or DLs. This last observation can lead to collaborations studying the problems from both points of view.

In addition to the longer talks whose abstracts accompany this document, other impromptu presentations were triggered by the previous discussions. One clear conclusion which can be taken from these engagements is that the potential for synergic growth between the areas is large and worth exploring.

Format

Due to the COVID-19 situation, the seminar had to be held in a hybrid format. While this had the obvious disadvantage of limiting the social interactions and offline scientific discussions that characterise Dagstuhl seminars, it also allowed the participation of many who, by distance or travel limitations, would have not been able to attend.

Overall, the hybrid format meant having a more structured and linear program than originally planned for the seminar, but as mentioned already the results are promising.

2 Table of Contents

3   Overview of Talks

3.1   SHAP Explanations with Booleans Circuit Classifiers

Leopoldo Bertossi (Adolfo Ibáñez University – Santiago, CL)

The presentation turns around the subject of explainable AI. More specifically, we deal with attribution numerical scores that are assigned to features values of an entity under classification, to identify and rank their importance for the obtained classification label.

We concentrate on the popular SHAP score [2] that can be applied with black-box and open models. We show that, in contrast to its general #P-hardness, it can be computed in polynomial time for classifiers that are based on decomposable and deterministic Boolean decision circuits. This class of classifiers includes decision trees and ordered binary decision diagrams. This result was established in [1]. The presentation illustrates how the proof heavily relies on the connection to SAT-related computational problems.

References
1   Arenas, M., Barceló, P., Bertossi, L. and Monet, M. The Tractability of SHAP-scores over Deterministic and Decomposable Boolean Circuits. Proc. AAAI 2021, pp. 6670-6678.
2   Lundberg , S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S-I. From Local Explanations to Global Understanding with Explainable AI for Trees. Nature Machine Intelligence, 2020, 2(1):56-67.

3.2   SGGS decision procedures for fragments of first-order logic

Maria Paola Bonacina (University of Verona, IT)

SGGS (Semantically-Guided Goal-Sensitive reasoning) is an attractive theorem-proving method for decision procedures, because it generalizes the Conflict-Driven Clause Learning (CDCL) procedure for propositional satisfiability, and it is model-complete in the limit, so that SGGS decision procedures are model-constructing. After summarizing the foundations of SGGS as a theorem-proving method, this talk presents recent and ongoing work on SGGS decision procedures for fragments of first-order logic. This includes both negative and positive results about known decidable fragments: for example, SGGS decides the stratified fragment, and hence Effectively PRopositional logic (EPR). SGGS also allows us to discover several new decidable fragments based on well-founded orderings. For most of these new fragments the small model property holds, as the cardinality of SGGS-generated models can be upper bounded, and membership can be tested by applying termination tools for rewriting. A report on experiments with the prototype theorem prover Koala, which is the first implementation of SGGS, closes the presentation. (SGGS is joint work with David Plaisted; SGGS decision procedures are joint work with Sarah Winkler, who is the author of Koala).

3.3 Clauses and Beyond: On Fast Prototyping with SAT Oracles

Alexey Ignatiev (Monash University – Clayton, AU)

This talk overviews SAT-based modeling capabilities offered by the PySAT toolkit. The toolkit aims at providing a simple and unified interface to a number of state-of-the-art Boolean satisfiability (SAT) solvers as well as to a variety of cardinality and pseudo-Boolean encodings. The purpose of PySAT is to enable researchers working on SAT and its applications and generalizations to easily prototype with SAT oracles in Python while exploiting incrementally the power of the original low-level implementations of modern SAT solvers. The toolkit can be helpful when solving problems in NP but also beyond NP that admit either direct clausal or non-clausal representation.

3.4 Modeling and Solving Problems with SAT

Jean-Marie Lagniez (CNRS, CRIL – Lens, FR)

SAT solvers are nowadays used to tackle a large panel of combinatorial problems. In this talk we highlight three different situations where we show that it is possible to significatively improve the SAT solver effectiveness when considering the problems' nature. First, in [1], we show that by playing on clause database cleaning, assumptions managements and other classical parameters, it was possible to immediately and significantly improve an intensive assumption-based incremental SAT solving task: Minimal Unsatisfiable Set. Second, in [2], we show that it is possible to overcome the difficulty of a problem by encoding it incrementally. We experimentally demonstrate that, by using this trick, the Zykov's encoding can be advantageously leveraged to tackle the graph coloring problem. Finally, we discussed solving the Team Formation problem (TF) with SAT technology. We show that for this problem which consists in solving a set cover problem with a large cardinality constraint, it is more advantageous to leverage a MaxSAT solver rather than a SAT solver. This clearly demonstrates the inefficiency of SAT solvers to deal properly with large cardinality constraints.

References

1 Gilles Audemard, Jean-Marie Lagniez, Laurent Simon: *Improving Glucose for Incremental SAT Solving with Assumptions: Application to MUS Extraction.* SAT 2013: 309-317
2 Gael Glorian, Jean-Marie Lagniez, Valentin Montmirail, Nicolas Szczepanski: *An Incremental SAT-Based Approach to the Graph Colouring Problem.* CP 2019: 213-231
3 Nicolas Schwind, Emir Demirovic, Katsumi Inoue, Jean-Marie Lagniez: *Partial Robustness in Team Formation: Bridging the Gap between Robustness and Resilience.* AAMAS 2021: 1154-1162

3.5   Proofs, Proof Logging, Trust, and Certification

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

Modern combinatorial optimization has had a major impact in science and industry. However, the problems considered are computationally very challenging, requiring increasingly sophisticated algorithm design, and there is a poor scientific understanding of how these complex algorithms, called combinatorial solvers, work. More importantly, even mature commercial solvers are known to sometimes produce wrong results, which can be fatal for some types of applications.

One way to address this problem is to try to enhance combinatorial solvers with proof logging, meaning that they output not only solutions but also proofs of correctness. One can then feed the problem, solution, and proof to a dedicated proof checker to verify that there are no errors. Crucially, such proofs should require low overhead to generate and be easy to check, but should supply 100% guarantees of correctness.

In addition to ensuring correctness, such proof logging could also provide strong development support in that it can quickly flag errors during solver software development. And since the proofs give detailed information about what reasoning steps were performed, this opens up new opportunities for in-depth performance analysis and for identifying potential for further improvements. Finally, it enables auditability by third parties without access to the solver used, and furnishes a stepping stone towards making results explainable.

In this presentation, we review proof logging as it has been adopted by the Boolean satisfiability (SAT) community, and discuss some of the challenges that lie ahead if we want to extend proof logging techniques to more general paradigms in combinatorial optimization. Along the way, we discuss what is meant by a "proof" in a formal sense, and the trade-offs involved between maximizing the efficiency of verification methods and minimizing the need for trust in such methods.

3.6   Provenance in Description Logics

Ana Ozaki (University of Bergen, NO)

Joint work of Camille Bourgaux, Diego Calvanese, Davide Lanti, Ana Ozaki, Rafael Peñaloza, Livia Predoiu, Guohui Xiao
Main reference Diego Calvanese, Davide Lanti, Ana Ozaki, Rafael Peñaloza, Guohui Xiao: "Enriching Ontology-based Data Access with Provenance", in Proc. of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, pp. 1616–1623, ijcai.org, 2019.
URL http://dx.doi.org/10.24963/ijcai.2019/224

We address the problem of handling provenance information in description logic ontologies [1, 2, 4, 3, 5]. We consider a setting for ontology-based data access in the classical DL-Lite$_R$ ontology language [3] and a setting for ontology-mediated access for the $\mathcal{ELH}^r$ ontology language [2]. Our works are based on semirings and extend the notion of data provenance in database theory. Here ontology axioms and mappings are also annotated with provenance tokens. A consequence inherits the provenance of the axioms involved in deriving it, yielding a provenance polynomial as annotation. We analyse the semantics for the already mentioned ontology languages DL-Lite$_R$ and $\mathcal{ELH}^r$ and investigate the problems of

computing provenance and of determining whether a given expression correctly represents the provenance information of a query. In particular, we show that the presence of conjunctions poses various difficulties for handling provenance, some of which are mitigated by assuming multiplicative idempotency of the semiring. We also analyse the problem of computing the set of relevant axioms for a consequence in the $\mathcal{ELH}^r$ case.

References

1 Camille Bourgaux and Ana Ozaki. Querying attributed DL-Lite ontologies using provenance semirings. In *AAAI*, 2019.

2 Camille Bourgaux, Ana Ozaki, Rafael Peñaloza, and Livia Predoiu. Provenance for the description logic elhr. In Christian Bessiere, editor, *IJCAI*, pages 1862–1869. ijcai.org, 2020.

3 Diego Calvanese, Davide Lanti, Ana Ozaki, Rafael Peñaloza, and Guohui Xiao. Enriching ontology-based data access with provenance. In Sarit Kraus, editor, *IJCAI*, pages 1616–1623. ijcai.org, 2019.

4 Ana Ozaki and Rafael Peñaloza. Provenance in ontology-based data access. In Magdalena Ortiz and Thomas Schneider, editors, *DL*, volume 2211 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

5 Rafael Peñaloza. An upper bound for provenance in elhr. In Martin Homola, Vladislav Ryzhikov, and Renate A. Schmidt, editors, *DL*, volume 2954 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.

3.7 ASP, Beyond NP, and Debugging for Explanations?

Francesco Ricca (University of Calabria, IT)

Answer Set Programming (ASP) is a logic programming paradigm featuring a purely declarative language with comparatively high modeling capabilities. ASP can model problems in NP in a compact and elegant way. The availability of efficient implementations, supporting API for programmers makes it suitable for developing applications. ASP implementations are based on SAT technology, and ASP is also a good candidate tool for implementing complex reasoning with Description logics. ASP is a worthy option for modeling several tasks related to explainabilty, which usually require complex modeling capabilities, with comparatively high computational complexity, often beyond NP. However, modeling problems beyond NP with ASP is known to be complicated, on the one hand, and limited to problems in Σ_2^P on the other. Inspired by the way Quantified Boolean Formulas extend SAT formulas to model problems beyond NP, we proposed an extension of ASP that introduces quantifiers over stable models of programs, called ASP(Q) [2]. The definition of ASP(Q) allows for disjunctive programs, thus all the features of the basic language are retained. However, by limiting to normal (or HCF) programs (extended with aggregates and other useful modeling constructs) in ASP(Q), one can take advantage of the classic generate-define-test modular programming methodology and other modeling techniques developed for these best-understood classes of programs to model any problem in the Polynomial Hierarchy. Indeed, the presence of quantifiers allows one to model complex properties in a direct way, and the solutions follow

directly from the definition in the natural language of the problem at hand. Despite that ASP features a simple syntax and intuitive semantics, errors are common during the development of ASP programs. For this reason, we proposed a novel debugging approach allowing for interactive localization of bugs in non-ground programs [1]. The debugging approach points the user directly to a set of non-ground rules involved in the bug, which might be refined (up to the point in which the bug is easily identified) by asking the programmer a sequence of questions on an expected answer set. Our debugger exploits techniques that are related to MUS search, and can be a starting point for developing methods for explaining the outcome of reasonings that can be cast in rule-based form.

References
1   Carmine Dodaro, Philip Gasteiger, Kristian Reale, Francesco Ricca, and Konstantin Schekoti-
     hin. Debugging non-ground ASP programs: Technique and graphical tools. *Theory Pract.
     Log. Program.* 19(2):290–316, 2019.
2   Amendola, G.; Ricca, F.; and Truszczynski, M. Beyond NP: quantifying over answer sets.
     Theory Pract. Log. Program. 19(5-6):705–721, 2019

3.8   Parameterised Complexity of SAT and related problems

Stefan Szeider (TU Wien, AT)

It is well understood that not all instances of the propositional satisfiability problem (SAT) have the same computational hardness. The hardness of instances depends on their structural properties. There are mainly two approaches to mathematically capture structure in SAT instances: (A: Correlation) to capture structure that statistically correlates with CDCL SAT solvers' running time, and (B: Causation) to capture structure that provides a rigorous running time guarantee for a SAT algorithm. In this talk, I will discuss the pros and cons of both approaches. I will then survey the main findings on approach B, covering structure based on graphical and syntactic concepts and on hybrid concepts that combine them.

References
1   Marko Samer and Stefan Szeider. Fixed-Parameter Tractability: in Handbook of Satisfiability,
     Chapter 17, 2nd Edition, 2021, pp. 693–736, ISBN 978-1-64368-160-3.

3.9   Scaling SAT/MaxSAT encodings to large instances with SLIM

Stefan Szeider (TU Wien, AT)

Encoding a combinatorial problem into SAT, to solve it with a SAT solver, is a compelling approach for solving NP-hard problems. However, the encoding often causes a blowup in encoding size, which limits the approach to small instances. The SAT-based Local Improvement Method (SLIM) overcomes this limitation by applying SAT (or MaxSAT)

encodings locally to a heuristically computed global solution. In this talk, I will present the general idea of SLIM and illustrate it with two recent applications to Bayesian network structure learning [1] and the induction of small decision trees [2].

References
1   Vaidyanathan Peruvemba Ramaswamy, Stefan Szeider: Turbocharging Treewidth-Bounded Bayesian Network Structure Learning. AAAI 2021: 3895-3903
2   André Schidler, Stefan Szeider: SAT-based Decision Tree Learning for Large Data Sets. AAAI 2021: 3904-3912

3.10   Existing Benchmarks from Description Logics

David Tena Cucala (University of Oxford, GB)

This talk lists several benchmarks for evaluating empirically the performance of reasoners for expressive Description Logics. The first part of the talk discusses curated repositories of ontologies available in the Web, including NCBO BioPortal, AgroPortal, the Oxford Ontology Repository, the Manchester OWL Corpus, and the repository for the OWL Reasoner Evaluation Competition. The second part of the talk describes several synthetic ontology generators, such as LUBM, UOBM, OntoBench, and OWL2Bench. Finally, the talk discusses some limitations of current benchmarks, and expresses some properties desirable in future testing frameworks for DL reasoners.

References
1   Ma, L., Yang, Y., Qiu, Z .and Xie, G., Pan, Y., Liu, S.: Towards a Complete OWL Ontology Benchmark. In: The Semantic Web: Research and Applications. pp. 125–139. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
2   NCBO BioPortal: `https://bioportal.bioontology.org`
3   AgroPortal: `http://agroportal.lirmm.fr`
4   Oxford Ontology Repository:
     `http://www.cs.ox.ac.uk/isg/ontologies/`
5   Manchester OWL Corpus:
     `http://mowlrepo.cs.manchester.ac.uk/datasets/mowlcorp/`
6   Protegé Ontology Library:
     `https://protegewiki.stanford.edu/wiki/ProtegeOntologyLibrary`
7   Lehigh University Benchmark: `http://swat.cse.lehigh.edu/projects/lubm/`
8   UOBM generator: `https://www.cs.ox.ac.uk/isg/tools/UOBMGenerator/`
9   OntoBench: `http://ontobench.visualdataweb.org`
10   OWL2Bench: `https://github.com/kracr/owl2bench`
11   OWL Reasoner Evaluation Competition: `http://dx.doi.org/10.5281/zenodo.18578`

Participants

- Meghyn Bienvenu
CNRS & University of Bordeaux
– Bordeaux, FR
- Maria Paola Bonacina
University of Verona –
Verona, IT
- Stefan Borgwardt
TU Dresden – Dresden, DE

- Camille Bourgaux
CNRS & ENS – Paris, FR
- David Carral
INRIA – Sophia Antipolis, FR
- Aaron Eberhart
Kansas State University –
Manhattan, US

- Jean-Marie Lagniez
CNRS, CRIL – Lens, FR
- Joao Marques-Silva
CNRS – Toulouse, FR
- Jordi Planes
University of Lleida – Lleida, ES
- Anni-Yasmin Turhan
TU Dresden – Dresden, DE

Remote Participants

- Carlos Ansotegui
University of Lleida – Lleida, ES
- Leopoldo Bertossi
Adolfo Ibáñez University –
Santiago, CL
- Olaf Beyersdorff
Universität Jena – Jena, DE
- Emmanuel Hebrard
LAAS – Toulouse, FR
- Ian Horrocks
University of Oxford –
Oxford, GB
- Alexey Ignatiev
Monash University –
Clayton, AU
- Matti Järvisalo
University of Helsinki –
Helsinki, FI
- Oliver Kullmann
Swansea University, GB

- Carlos Mencía
University of Oviedo –
Oviedo, ES
- Thomas Meyer
University of Cape Town –
Cape Town, ZA
- Jakob Nordström
University of Copenhagen, DK &
Lund University, SE
- Magdalena Ortiz
TU Wien – Vienna, AT
- Ana Ozaki
University of Bergen –
Bergen, NO
- Rafael Peñaloza
University of Milano-Bicocca –
Milano, IT
- Francesco Ricca
University of Calabria –
Rende, IT

- Uli Sattler
University of Manchester –
Manchester, GB
- Renate Schmidt
University of Manchester –
Manchester, GB
- Stefan Szeider
TU Wien – Vienna, AT
- David Tena Cucala
University of Oxford –
Oxford, GB
- Cesare Tinelli
University of Iowa –
Iowa City, US
- Renata Wassermann
University of Sao Paulo –
Sao Paulo, BR

Report from Dagstuhl Seminar 21362

Structure and Learning

Edited by

Tiansi Dong[1], Achim Rettinger[2], Jie Tang[3], Barbara Tversky[4], and Frank van Harmelen[5]

1   **Universität Bonn, DE,** `tdong@uni-bonn.de`
2   **Universität Trier, DE,** `rettinger@uni-trier.de`
3   **Tsinghua University – Beijing, CN,** `jietang@tsinghua.edu.cn`
4   **Columbia University – New York, US,** `btversky@stanford.edu`
5   **VU University Amsterdam, NL,** `frank.van.harmelen@vu.nl`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 21362 "Structure and Learning", held from September 5 to 10, 2021. Structure and learning are among the most prominent topics in Artificial Intelligence (AI) today. Integrating symbolic and numeric inference was set as one of the next open AI problems at the Townhall meeting "A 20 Year Roadmap for AI" at AAAI 2019. In this Dagstuhl seminar, we discussed related problems from an interdiscplinary perspective, in particular, Cognitive Science, Cognitive Psychology, Physics, Computational Humor, Linguistic, Machine Learning, and AI. This report overviews presentations and working groups during the seminar, and lists two open problems.

1   Executive Summary

Tiansi Dong (Universität Bonn, DE)
Achim Rettinger (Universität Trier, DE)
Jie Tang (Tsinghua University – Beijing, CN)
Barbara Tversky (Columbia University – New York, US)
Frank van Harmelen (VU University Amsterdam, NL)

Deep Learning systems are the hope of the fifth industrial revolution. However, recent studies have found that Deep Learning systems can be easily manipulated, i.e. in Natural Language Understanding, Object Recognition. How to introduce structures into Deep Learning systems to improve reliability and performance has become a hot topic in Natural Language Processing (NLP), Machine Learning (ML), Semantic Web (SW) communities around the world. The aim of the seminar is to bring together interdisciplinary researchers around the world for constructive discussions on this theme, in particular, it intends to establish international collaborations to promote computational Humor, with the hope to let AI bring more joy, more

Except where otherwise noted, content of this report is licensed
under a Creative Commons BY 4.0 International license
Structure and Learning, *Dagstuhl Reports*, Vol. 11, Issue 08, pp. 11–34
Editors: Tiansi Dong, Achim Rettinger, Jie Tang, Barbara Tversky, and Frank van Harmelen
   Dagstuhl Reports
   Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

laugh into the world, and do more good for the society. The hybrid seminar is structured in the form of Talks, Working Groups, and Open questions. The seminar started with the talk "Hybrid AI for Humor". The dynamic semantics of humor is beyond the reach of the classic symbolic AI, the deep learning paradigm, and current neural-symbolic integration methods, but can be captured by the neural geometric embedding, in terms of rotating sphere embedding. This novel embedding is rooted in Qualitative Spatial Representation (QSR) in symbolic AI and Learning Representation (LR) in neural ML. The former tries to symbolically delineate the basic spatial knowledge that humans have and possible ways that this knowledge can be used as a reference for abstract knowledge in other domains. LR aims at learning latent feature knowledge from data. The motivation and a geometric approach to realizing the unification were introduced in the talk "Rotating Spheres – A New Wheel for Neuro-Symbolic Unification". The motion of rotating spheres in high-dimensional space is served as a computational model to simulate (1) the motion of the physical world, (2) the circular interaction among the mind, the body, and the world (called *spraction* – a contraction of space, action, and abstraction, in which actions in space create abstractions).

The motion of the physical world is vividly explained in the talk "Rotating Spheres in the Milky Way". This *spraction* process is explained in the talk "Thinking with the Body and the World", which can guide the design of novel cognitive robots, and promote novel cognitive architectures. Two topics were covered by the talk "Learning about Language and Action for Robots", and the talk "Neural-Symbolic Models, Dual-Process Theories, and Cognitive Architectures".

In primates, the same brain structures that support spatial thinking also support conceptual thinking, Single cells in hippocampus gather multi-media information from different memories in the brain to represent places in space, events in time, ideas in conceptual spaces. Update-to-date research of neural simulation is introduced by Volker Tresp with the talk "Knowledge Graph and Cognitive Learning: from Perception to Memory Embedding", which maps embedding models to various cognitive memory functions, in particular to semantic and concept memory, episodic memory, sensory memory, short-term memory, and working memory.

Spatial thinking is multi-modal and established and distorted by our actions and perceptions of the spaces we interact in. This raises two questions: What are good representations for video understanding? and how to compute symbolic rules that the models have learned from the training data? Juergen Gall introduced holistic video understanding and argued the potential of hybrid approaches that combine neural networks with symbolic AI for video understanding and reasoning. Cuenca Grau, Bernado gave the talk "Characterizing Graph Neural Networks Using Logical Rules". He formally defines what it means for a set of logical rules to characterize the behavior of a model and proposes a GNN-based architecture that admits a characterization in terms of Datalog rules.

Spatial thinking is evident in the ways we think and the ways we externalize thought, for example, through words. Our words act on thought the way we act on objects. The philosophy of spatial thinking challenges the computational approach to natural language processing and understanding. Roberto Navigli argued that Natural Language Understanding (NLU) is particularly challenging, as this requires the machine to go beyond processing strings to reach a semantic level. Recent developments and challenges were discussed through three key tasks in NLU, namely Word Sense Disambiguation, Semantic Role Labeling, and Semantic Parsing. Zhiyuan Liu argued that knowledge (including symbols, embeddings, or models) is the key to a deeper understanding of human languages and that big pretrained language models can be regarded as the most advanced approach to model knowledge and to capture knowledge

(including commonsense) from plain text and that the key challenge is how to incorporate both open data and structural knowledge. Alexander Mehler reviewed problems of neural network-based language learning, suggested to introduce the concept of cognitive maps and spatial information processing, and sketched a synergistic model that relates the dynamics of distributed information processing to bias interaction. Jie Tang introduced Wu-Dao, China's first homegrown super-scale intelligent model system, with the goal of building an ultra-large-scale cognitive-oriented pretraining model to focus on essential problems in general artificial intelligence from a cognitive perspective. Wu-Dao substantially outperforms BERT on the SuperGLUE natural language understanding benchmark with the same amount of pre-training data. Alam Mehwish discusses the characteristics of the existing benchmark datasets for the task of KG Completion, and limitations of the existing benchmark datasets and targets those issues in the generation of LiterallyWikidata.

Another externalization of spatial thinking is through graphics. In the talk "Semi-Riemannian Graph Convolutional Networks", Steffen Staab introduced their new geodesic tools that allow for extending neural network operations into geodesically disconnected semi-Riemannian manifolds. Thomas Liebig introduced using p-adic coding and computation for structured domains or domains with inherent granularity.

The ultimate form of spatial thinking is comics (a form of humor, the most creative form of storytelling), which typically show bodies acting in space. Humor is used as a testbed and lighthouse for the development of AI and machine learning. In the talk "Ethics of AI Humor" Kiki explained how humor has frustrated symbolic and statistic AI approaches; in the talk "Knowledge and Inferences Needed for Humor" Julia Rayz introduced recent advances in transformer-based approaches, and raised open questions.

Working groups are the main components of the seminar. The hybrid seminar provides an excellent chance to practice the situation that participants can continue to work together after this seminar, which is the main outcome of this seminar.

The seminar ended with the discussion "Boxology for Hybrid Learning and Reasoning Systems" chaired by Frank van Harmelen.

2 Table of Contents

3 Overview of Talks

3.1 LiterallyWikidata – A Benchmark for Knowledge Graph Completion using Literals

Mehwish Alam (FIZ Karlsruhe, DE)

Recently many efforts have been made in automatically constructing Knowledge Graphs (KGs) from heterogeneous data sources such as text, image, etc. More specifically, cross-domain open KGs such as Freebase, Dbpedia, Wikidata, etc. are either extracted automatically from structured data, generated using heuristics, or are human-curated. Due to the Open World Assumption, KGs are never complete, i.e., there are always some facts missing. In order to solve this problem, recently different KG embedding models have been proposed for automated KG Completion.

This talk discusses the characteristics of the existing benchmark datasets for the task of KG Completion. It further discusses a set of benchmark datasets extracted from Wikidata and Wikipedia, named LiterallyWikidata. It also takes into account the limitations of the existing benchmark datasets and targets those issues in the generation of LiterallyWikidata.

LiterallyWikidata has been prepared with the main focus on providing benchmark datasets for multimodal KG Embedding (KGE) models, specifically for models using numeric and/or text literals. Hence, the benchmark is novel as compared to the existing datasets in terms of properly handling literals for those multimodal KGE models. LiterallyWikidata contains three datasets that vary both in size and structure. These datasets are analyzed based on their connectivity, density, and diameter. Moreover, the datasets also include textual information about the entities in multiple languages (in addition to English). Finally, the results of the benchmarking experiments on the task of link prediction were conducted on LiterallyWikidata.

Currently, LiterallyWikidata does not consider image literals. Moreover, the current results report the performance of existing models on the task of head, tail, and relation prediction. More experiments need to be conducted for the task of entity classification. As a future perspective, these points will be considered along with the bias analysis of this benchmark dataset.

3.2 Online Perceptual Learning and Natural Language Acquisition for Autonomous Robots

Anthony Cohn (University of Leeds, GB)

To operate effectively, and to collaborate with humans, robots need to know much about the world, including the kinds of objects in the world, their properties, the spatial relationships between them and actions that can be performed on them, as well as how language is used to

describe these things. In this work[1], the problem of bootstrapping knowledge in language and vision for autonomous robots is addressed through novel techniques in grammar induction and word grounding to the perceptual world.In particular, we demonstrate a system, called OLAV, which is able, for the first time, to (1) learn to form discrete concepts from sensory data; (2) ground language (n-grams) to these concepts; (3) induce a grammar for the language being used to describe the perceptual world; and moreover to do all this incrementally, without storing all previous data. The learning is achieved in a loosely supervised manner from raw linguistic and visual data. Moreover, the learnt model is transparent, rather than a black-box model and is thus open to human inspection. The visual data is collected using three different robotic platforms deployed in real-world and simulated environments and equipped with different sensing modalities, while the linguistic data is collected using online crowdsourcing tools and volunteers. The analysis performed on these robots demonstrates the effectiveness of the framework in learning visual concepts, language groundings and grammatical structure in these three online settings.

3.3   Characterising Graph Neural Networks Using Logical Rules

Bernado Cuenca Grau (University of Oxford, GB)

There are many practical situations where we would like to learn a function that completes a given dataset in graph format (over a fixed set of unary and binary predicates) with additional facts; these include, for instance, knowledge graph completion and recommendation systems.

Graph Neural Networks (GNNs) are a family of ML models that have proved well-suited for such situations; however, as it is the case with other ML models, it is difficult to explain its predictions, and there is a growing interest in computing "general patterns" (or rules) that the models have learnt from the training data as a form of symbolic explanation.

In this work, we formally define what it means for a set of logical rules to characterise the behaviour of a model, and propose a GNN-based architecture that admits a characterisation in terms of Datalog rules. Our architecture consists of three main elements: (1) an encoder, which transforms the input dataset into a graph annotated with numeric feature vectors; (2) a Monotonic Graph Neural Network (a GNN variant satisfying a property akin to that of monotonicity under homomorphisms of First Order Logic); and (3) a decoder, which transforms the result of GNN application into the output dataset.

Our architecture can be successfully trained in practice for tasks such as knowledge graph completion; furthermore, the corresponding set of rules can be extracted algorithmically from the trained model. Our experiments on well-known knowledge graph completion benchmarks show competitive performance with that of state-of-the-art rule learning methods such as AnyBURL and DRUM.

3.4 Rotating Spheres: A New Wheel for Neuro-Symbolic Unification

Tiansi Dong (Universität Bonn, DE)

The distinction between discrete symbolic representations and continuous vector embeddings (i.e. subsymbolic representation) separates AI researches into two seemingly incompatible paradigms. Cognitively, both of them are products of our minds. How can discrete symbolic representations and rigorous symbolic reasoning be carried out by our neural mind? Sun (1994) suggested a dual-process theory [1]: *...cognitive processes are carried out in two distinct "levels" with qualitatively different mechanisms. Each level encodes a ... set of knowledge for its processing, and the coverage of the two sets ... overlaps substantially. Two different "levels" can potentially work together synergistically, complementing and supplementing each other.*

Given a discrete tree structure and the vector embeddings of its nodes, we can promote these vectors into spheres, and let the containment relations among spheres capture the discrete tree structure. Following this intuition, I argue for a novel neuro-geometric approach for neuro-symbolic unification [2] as follows: (1) vector embeddings from classic neural-networks can be promoted into spheres in higher dimensional space; (2) symbolic structures shall be precisely encoded as topological relations among these spheres. To support this argument, I show the empirical experiments with tree structures and their vector embeddings [3, 4], and neural Euler diagram embeddings for syllogistic reasoning [5].

By representing features as rotating axes, I introduce the term *Rotating Spheres* as a neuro-symbolic building block, and illustrate how they can be used to computationally interpret symbolic humor theory [8] and to simulate "Spatial Humor" [6, 7] which helps to explore how spatial thinking can be computationally linked to non-spatial thinking [11]. This starts from a topological representation of spaces and events and moves on to represent expectations (as another space in mind) and emotions (as a rotating axis). The violation of expectation in humor [9, 10] is computationally simulated as the flip of a rotating axis of a sphere. The flip works like the turning of a button, which triggers the mind machine to laugh.

Continuous vector embeddings and discrete symbolic rules are images seen from the two traditional eyes of AI. Rotating spheres serve as basic building block to unify the two approaches in the mind of AI and to computationally embody the way of thinking. This shapes a new style of AI.

References
1   Ron Sun. *Integrating Rules and Connectionism for Robust Commonsense Reasoning.* John Wiley and Sons, New York, 1994
2   Tiansi Dong. *A Geometric Approach to the Unification of Symbolic Structures and Neural Networks.* Studies in Computational Intelligence 910, Springer, 2021.
3   Tiansi Dong, Christian Bauckhage, Hailong Jin, Juanzi Li, Olaf Cremers, Daniel Speicher, Armin B. Cremers, Joerg Zimmermann. *Imposing Category Trees Onto Word-Embeddings Using A Geometric Construction.* ICLR-19
4   Tiansi Dong, Zhigang Wang, Juanzi Li, Christian Bauckhage, Armin B. Cremers. *Triple Classification Using Regions and Fine-Grained Entity Typing.* AAAI-19
5   Tiansi Dong, Chengjiang Li, Christian Bauckhage, Juanzi Li, Stefan Wrobel, Armin B. Cremers. *Learning Syllogism with Euler Neural-Networks.* arXiv:2007.07320

6   Siba Mohsen, Tiansi Dong. *A Neuro-symbolic approach to Joke Understanding*. At Workshop "AI4Jokes: towards a Neuro-Symbolic Approach for the Analysis and the Generation of Jokes" at the 11th Humor Research Conference.

7   Bitasta Biswas, Tiansi Dong. *Using Blender for Joke Understanding*. At Workshop "AI4Jokes: towards a Neuro-Symbolic Approach for the Analysis and the Generation of Jokes" at the 11th Humor Research Conference.

8   Tiansi Dong. *Jokes, Collages, and AI*. Webinar Humor and Artificial Intelligence at the ISHS (The International Society for Humor Studies) Spring 2021 Webinar Series

9   Salvatore Attardo, Christian F Hempelmann, and Sara Di Maio. *Script Oppositions and Logical Mechanisms: Modeling Incongruities and their Resolutions*. Humor, 15(1):3 – 46, 2002.

10   Julia Taylor Rayz. *Scripts in the Ontological Semantic Theory of Humor*. In: Salvatore Attardo, editor, Script-Based Semantics. 2020.

11   Barbara Tversky. *Mind in Motion*. Basic Books, New York, USA, 2019.

3.5   What are Good Representations for Video Understanding?

Jürgen Gall (Universität Bonn, DE)

In this talk, I will give an overview of some representations that have been used for video understanding. The representation ranges from knowledge-based representations that have been used for robotics applications, fine-hierarchies for sports like gymnastics, joint representations for language and video, and verb-object representations. While all these representations have advantages for specific applications, it is still an open research problem how a universal representation of actions can be defined. In the second part of the talk, I introduce holistic video understanding. Instead of just representing a video by action labels or captions, holistic video understanding provides are more rich representation which contains labels for actions, objects, scenes, attributes, events, and concepts. In order to study this problem, we released the HVU dataset (`https://holistic-video-understanding.github.io/`). It consists of over 570k videos with over 9m annotations of 3142 different semantic labels. I will show a few examples like video retrieval, video captioning, and action recognition that demonstrate the benefits of having such rich semantic descriptions of the videos. Nevertheless, it is an open research questions how relations between objects, attributes, actions, and events can be best utilized for video understanding. In the last part of the talk, I will describe a hybrid approach that combines recurrent neural networks, hidden Markov models, and a context-free grammar for temporal action segmentation. I will show some results that demonstrate the advantage of combining grammars with neural networks and give a few examples for weakly supervised learning. The talk concludes that there is a large potential for hybrid models that combine neural networks with symbolic AI for video understanding and reasoning, which is a promising research direction.

References

1   A. Diba, M. Fayyaz, V. Sharma, M. Paluri, J. Gall, R. Stiefelhagen, L. van Gool. *Large Scale Holistic Video Understanding*. European Conference on Computer Vision (ECCV'20), Springer, LNCS 12350, 593-610, 2020.

2   S. Li, Y. Abu Farha, Y. Liu, M.-M. Cheng, J. Gall. *MS-TCN++: Multi-Stage Temporal Convolutional Network for Action Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

3   H. Kuehne, A. Richard, J. Gall. *A Hybrid RNN-HMM Approach for Weakly Supervised Temporal Action Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 4, 765-779, 2020.

4   A. Richard, H. Kuehne, J. Gall. *Action Sets: Weakly Supervised Action Segmentation without Ordering Constraints.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18), 5987-5996, 2018.

5   A. Richard, H. Kuehne, J. Gall. *Weakly Supervised Action Learning with RNN based Fine-to-Coarse Modeling.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17), 1273-1282, 2017.

3.6   Ethics of AI Humor

Christian Hempelmann (Texas A&M University – Commerce, US) and Max Petrenko (Amazon Web Services – Seattle, US)

To the layperson and self-delusional expert, AI has become frighteningly powerful and humanlike, not least in its application to natural language processing. This leads to ethical issues coming to the fore as to who is responsible for the output of these systems at various levels. General AI systems have to make life-and-death decisions when coupled with self-driving cars and weapons systems. Language-generating AI systems produce racist and sexist output reflecting the human-generated data the systems have learned from. These AI systems are also used to classify as well as generate humor, which raises the same general, but also specific ethical issues. The latter stem from the specific meaning constellations in humor leading not least to the deniability of its messages. We aim to outline the relevant key points to initiate a discussion that we think needs to happen now. I will present with an overview of past approaches to generating and analyzing humor computationally up to 2015. As in its parent discipline, computational linguistics, early approaches were symbolic, rule- and resource-based. Since the 1990s, the methodology came increasingly from computer sciences and was probabilistic, up to the unexplainable algorithms of machine learning.

3.7   Rotating spheres in the Milky Way

Michael Kramer (MPI für Radioastronomie, DE)

My research focuses on the study of "pulsars", i.e. rotating dense cores of exploded stars that emit a lighthouse beam of radio emission which makes them detectable as pulsating radio sources. Their rotation – and motion! – in the Milky Way can be tracked by their regular beacons that stand out from the usually irregular, random, and stochastic processes in the Milky Way and the Universe as a whole. This allows us to use them as reference points, charting our position and motion relative to them. They come with rotation frequencies

from about once every second to almost thousand times a second. The physics is extreme, as the rotational speed at the equator of these about 25 km large objects reaches significant fractions of the speed of light. Finding and exploiting them allows us to use them as precise cosmic clocks, for instance, to study and test the predictions of general relativity. In the past, we had armies of Ph.D. students sifting through our telescope data. Today, by enlarging the parameter space and increasing our sensitivity, we get millions and millions of candidates, so that we need to deploy artificial intelligence and machine learning methods nowadays. One main area of concern is how to avoid throwing out or ignoring the discovery of a new type of signal, only because we didn't know its properties before. In 1967, when pulsars were discovered, it was a bright female Ph.D. student who noticed the unusual signature of the signal. Would artificial intelligence be as clever? If we make enough discoveries, we can expand our network (or "array") of pulsars to convert the Milky Way into a galaxy-sized gravitational wave detector. We can do this by comparing the arrival time of different pulsars and detecting tiny variations, correlated in direction on the sky, which are caused by a background of gravitational waves filling the Universe from past mergers and collisions of galaxies.

3.8   *p*-adic Coding & Computation for Structured Data

Thomas Liebig (TU Dortmund, DE)

For centuries, physical observations and sensor readings are foremost represented by real numbers. This particular choice of data representation and model selection poses assumptions on the geometry of the observed feature space.

Many currently applied models of physical processes suffer from the representation bias posed by the physical origin of most process models, which start modeling with ordinary or partial differential equations on the field of real numbers. Currently, lots of domain and model-specific literature exists, how to incorporate expert knowledge on processes, correlations, or physical behaviour in these models. But some of the data generating processes exhibit chaotic behaviour and we observe that

1) depending on granularity it is cumbersome to model bursts, and

2) scale matters.

Consider, as an example, a traffic flow prediction model. While it appears simple to predict the average daily traffic volumes on a street segment from surrounding observations, it gets hard at fine granularities, since traffic is controlled by external semaphores, and it is not the same predicting few cars more or less on a street when the total traffic flow is high or almost empty.

For such structured domains or domains with inherent granularity, *p*-adic coding and computations overcome the assumptions and provide a natural framework for structured data. We highlight approaches and challenges of *p*-adic modeling.

3.9 Knowledgeable Learning for Natural Language Processing

Zhiyuan Liu (Tsinghua University – Beijing, CN)

In this talk, we argue that knowledge is the key to deeper understanding of human languages. Knowledge can be represented in appropriate ways including symbols, embeddings or models. Natural language processing can be formalized as the acquisition, representation, and application of complicated knowledge for language understanding. Big pretrained language models can be regarded the most advanced approach to model knowledge and to capture knowledge from plain text including commonsense. The key challenge is how to incorporate existing knowledge and make PLMs learn from both open data and structural knowledge. In this talk, we summarize various promising approaches to knowledgeable learning for NLP, including knowledge augmentation over input, knowledge framework over neural architecture, and knowledge regularization over learning objectives. Prompt Tuning seems promising to stimulate model knowledge for diverse downstream tasks.

3.10 Learning Linguistic Representations: Some Challenges and Opportunities

Alexander Mehler (Goethe-Universität – Frankfurt am Main, DE)

In this short talk, I briefly review some of the problems of neural network-based language learning and suggestions for overcoming them. This includes an account of gaps arising from symbolic learning resources. This concerns, first, a so-called algorithmization bias, according to which the same corpus looks very different from the point of view of the output distributions of a set of NLP routines focusing on the same task (e.g., sentiment analysis), so that the application of these routines becomes predictable. Beyond that, polymorphic structuring of fragmented texts (using Twitter data as an example), aspects of distributed authorship and readership, and biased information processing are exemplified. To overcome problems related with these scenarios, the presentation builds on the concept of cognitive maps and spatial information processing. To this end, research on biases is combined with concepts of context-sensitive language learning (e.g., regarding the salience of landmarks, hierarchization effects, localization effects, and Zipfian tripartivity). This analysis is then used to distinguish between two roles of biases in distributional semantics: as OUT parameters for the purpose of hypothesis testing (asking which biases are reflected by which resource), and as IN parameters that constrain the generation of language representation models to reflect particular biases. The paper ends by sketching a synergetic model that relates the dynamics of distributed information processing to bias interaction.

3.11 The challenges of bringing together NLU, multilinguality and KG

Roberto Navigli (Sapienza University of Rome, IT)

Natural Language Processing (NLP) has seen an explosion of interest in recent years, with many industrial applications relying on key technological developments in the field. However, **Natural Language Understanding (NLU)** – which requires the machine to get beyond processing strings and involves a semantic level – is particularly challenging due to the **pervasive ambiguity of language**. In this talk I will present the **recent developments and challenges** of three key tasks in NLU, namely Word Sense Disambiguation, Semantic Role Labeling and Semantic Parsing.

Word Sense Disambiguation, the task of associating a word in context with its most appropriate sense from a predefined sense inventory, is one of the hardest tasks in NLP, however the advent of Deep Learning has led significant improvements, also thanks to the integration of explicit knowledge in the form of lexical knowledge graphs, leading to performances above the hard-to-beat threshold of 80% F1 on standard test sets. Still, several issues are open, including the availability of training data in languages other than English and the granularity of sense inventories. I will also mention the option of dropping the need of a sense inventory, an approach we called Generationary.

I will then move to sentence-level semantics, which is also hampered by the lack of large-scale annotated data. **Semantic Role Labeling**, aimed at extracting the predicate-argument structure of a sentence and identifying the semantic relationships between a predicate and its arguments, suffers from the existence of different, heterogeneous framesets for each language. Recently, we put forward a unifying neural architecture which, trained on data from different languages, outputs predicate senses and roles according to all the available inventories, and enables the use of previously unseen languages and the creation of a network of predicate-argument meanings. Finally, I will discuss the issues of **Semantic Parsing**, which moves from the predicate-argument structure to the overall structure of a sentence, typically as a semantic graph. I will also introduce two recent approaches to generative semantic parsing, based on graph linearization techniques and a pretrained encoder-decoder architecture, and cross-lingual parsing where we address the pervasive data paucity issue with the production of high-quality silver training data.

3.12 Hybrid Humor AI

Max Petrenko (Amazon Web Services – Seattle, US) and Christian Hempelmann (Texas A&M University – Commerce, US)

AI-oriented research on humor has been evolving in a pattern that can also be observed in the history of AI as a field. Case- and problem-specific projects, informed by specific methods, led to results that, while appropriate for the case in point, offered limited room for generalization beyond the discussed cases. These findings also struggled to advance the understanding of humor as a general purpose faculty of intelligence and what kind of AI systems would

be required to model humor comprehension and generation. Similarly to the conceptual space, methodology-wise, much of the existing AI research on humor tends to adopt (often in an exclusionary fashion) a statistical/learning or a symbolic/knowledge-oriented view, which has yielded fragmentary dividends. We will offer insights on the properties of hybrid AI systems required to support general purpose humor research. We will first discuss the assumptions that AI research needs to treat humor in light of artificial general (as opposed to special purpose) intelligence theories, and that a hybrid, or neuro-symbolic, framework is an appropriate framework to pursue such research. We will then focus on the symbolic component of the AI research on humor and discuss the benefits (and challenges) of ontology modeling for the design and implementation of the symbolic component. We will work through an example of a joke modelled with the ontology of humor as an access point to its general architecture.

3.13 Knowledge and Inferences Needed for Humor

Julia Rayz (Purdue University – West Lafayette, US)

Humor is an interesting phenomenon in the context of AI because it relies on inferences and unstated knowledge. The state of the art in natural language processing can claim some (limited) successes in knowledge acquisition from text. It is also possible to access some information from text thought other means, such as asking questions and getting answers to them. This works for many applications but is not sufficient for humor, which – according to various theories – requires realization and resolution of something unexpected. How much, exactly, can we resolve or even approximate the needed unexpected situation with large corpora alone? This talk starts with a brief overview of linguistic theories of humor. We then look at one of the jokes extensively analyzed in the humor literature and discuss the type of knowledge (scripts) needed to understand and detect this joke and project this knowledge and inferences to the (relatively) recent advances in transformer-based approaches, which demonstrate activation of other scripts. The talk also covers some of the recent papers in (non-humorous) natural language processing that are relevant to humor processing. The question that remains unanswered is: would it be possible to retrieve low frequency but relevant information from enormous text corpora by rephrasing and extending the text of the joke or transferring it to a question/answering task? More importantly, if such low-frequency information could be retrieved, would it correspond to typical human-level script retrieval/analysis?

3.14 Semi-Riemannian Graph Convolutional Networks

Steffen Staab (Universität Stuttgart, DE)

Graph Convolutional Networks (GCNs) are typically studied through the lens of Euclidean geometry. Non-Euclidean Riemannian manifolds provide specific inductive biases for embedding hierarchical or spherical data, but cannot align well with data of mixed topologies. We consider a larger class of semi-Riemannian manifolds with indefinite metric that generalize hyperboloid and sphere as well as their submanifolds. We develop new geodesic tools that allow for extending neural network operations into geodesically disconnected semi-Riemannian manifolds. As a consequence, we derive a principled Semi-Riemannian GCN that first models data in semi-Riemannian manifolds of constant nonzero curvature in the context of graph neural networks. Our method provides a geometric inductive bias that is sufficiently flexible to model mixed heterogeneous topologies like hierarchical graphs with cycles. Empirical results demonstrate that our method outperforms Riemannian counterparts when embedding graphs of complex topologies.

3.15 Neural-Symbolic Models, Dual-Process Theories, and Cognitive Architectures

Ron Sun (Rensselaer Polytechnic – Troy, US)

In this talk, I address neural-symbolic models, dual-process theories, and cognitive architectures – their relationships and their relevance to each other. I provide some historical backgrounds and argue that dual-process theories have significant implications for developing neural-symbolic models. Computational cognitive architectures can help disentangle complex issues concerning dual-process theories and thus neural-symbolic models.

The notion of neural-symbolic models harkens back to the 1990s when such models first emerged (see, e.g., Sun & Bookman, 1994). There have been many different ways of structuring such models; the question remains: how should we best structure them? I argue that they should be structured in a cognitively motivated/justified way, based on human cognitive architecture. In particular, they should take into account dual-process theories concerning human cognitive architecture.

The distinction between "intuitive" and "reflective" thinking (i.e., system 1 and system 2) has been, arguably, one of the most important distinctions in cognitive science. There are currently many dual-process theories out there. One such theory was proposed early on in Sun (1994), where the two systems were characterized as follows: "... cognitive processes are carried out in two distinct "levels" with qualitatively different mechanisms. Each level encodes a ... set of knowledge for its processing, and the coverage of the two sets ... overlaps substantially." (Sun, 1994, p.44). That is, the two "levels" encode somewhat similar or overlapping content. But they encode their contents in different ways: Symbolic and

subsymbolic representation are used, respectively. Different mechanisms are thus involved at these two "levels". It was hypothesized that these two different "levels" can potentially work together synergistically, complementing and supplementing each other.

However, some details of more recent dual-process theories are more questionable. Although the distinction is important, the terms involved have often been ambiguous. Not much fine-grained analysis has been done, especially not in a precise, mechanistic, process-based way. Therefore, we need a conceptual and computational framework in this regard. The Clarion cognitive architecture (Sun, 2002, 2016) may be used at a theoretical level as a conceptual tool for generating interpretations and explanations. Many empirical and simulation studies have been conducted within the Clarion framework that shed light on relevant issues. Based on the framework, we re-interpret some common folk psychological notions, to give them more clarity and precision.

In summary, dual-process theories have important implications for neural-symbolic models. If cognitive-psychological realism is what one wants to achieve in developing computational models, dual-process theories must be taken into consideration. However, some issues involved in dual-process theories are more complex than often assumed. These issues are crucial for developing cognitive architectures, and in turn cognitive architectures can help in disentangling these and other theoretically important issues. Together they can lead to better neural-symbolic models. Dual-process theories serve as the theoretical basis and justifications for many cognitively motivated neural-symbolic models.

References

1   Sun, R.. *Integrating Rules and Connectionism for Robust Commonsense Reasoning.* John Wiley and Sons, New York, 1994.

2   Sun, R.. *Duality of the Mind.* Lawrence Erlbaum Associates, Mahwah, NJ., 2002.

3   Sun, R.. *Anatomy of the Mind.* Oxford University Press, Oxford, UK., 2016.

4   Sun, R. & L. Bookman (eds.). *Computational Architectures Integrating Neural and Symbolic Processes.* Kluwer, Needham, MA., 1994.

3.16   WuDao: Pretrain the World

Jie Tang (Tsinghua University – Beijing, CN)

Large-scale pre-trained model on web texts have substantially advanced the state of the art in various AI tasks, such as natural language understanding and text generation, and image processing, multimodal modeling. The downstream task performances have also constantly increased in the past few years. In this talk, I will first go through three families: autoregressive models (e.g., GPT), autoencoding models (e.g., BERT), and encoder-decoder models. Then, I will introduce China's first homegrown super-scale intelligent model system, with the goal of building an ultra-large-scale cognitive-oriented pretraining model to focus on essential problems in general artificial intelligence from a cognitive perspective. In particular, as an example, I will elaborate a novel pretraining framework GLM (General Language Model) to address this challenge. GLM has three major benefits: (1) it performs well on classification, unconditional generation, and conditional generation tasks with one single pretrained model; (2) it outperforms BERT-like models on classification due to improved pretrain-finetune consistency; (3) it naturally handles variable-length blank filling which

is crucial for many downstream tasks. Empirically, GLM substantially outperforms BERT on the SuperGLUE natural language understanding benchmark with the same amount of pre-training data.

3.17 Knowledge Graph and Cognitive Learning: from Perception to Memories

Volker Tresp (Siemens – München, DE)

In this talk, I will present our work on learning representation about knowledge in different memories. A variety of cognitive memory functions are simulated, in particular those in semantic and concept memory, episodic memory, sensory memory, short-term memory, working memory, and the way that perception shapes semantic memory. I will also present and discuss our on-going researches, ranging from an integrated theoretical analysis, novel algorithms, to many new experimental results.

3.18 Thinking with the Body and the World

Barbara Tversky (Columbia University – New York, US)

All living things must move and act in space to survive, even plants, Without action in space, life ceases. Spatial thinking is the foundation of thought; not the entire edifice, but the foundation. In primates, the same brain structures that support spatial thinking also support conceptual thinking, Single cells in hippocampus gather multi-media information from all over the brain to represent places in space, events in time, ideas in conceptual spaces. The single cells are spatially arrayed in entorhinal cortex. Spatial thinking is acting in space with the things in space. Spatial thinking is multi-modal and established and distorted by our actions and perceptions of the spaces we interact in: the space of the body, the space immediately around the body; the larger space of navigation that must be pieced together from different multi-modal experiences.

Spatial thinking is evident in the ways we think and the ways we externalize thought, primarily through words, gestures, and graphics, Our minds go from thought to thought the ways our feet go from place to place, real or virtual paths from place to place or thought to thought. Our words and gestures act on thought the way we act on objects. We raise ideas, pull them together, tear them apart; those words are often accompanied by gestures of the physical actions, even though those actions aren't actually performed and don't need to be performed. Gestures and graphics bear more direct relations to meanings than symbolic words. They augment and alter our own thinking and that of others. When people are alone in a room studying complex texts for later testing, their hands often spontaneously create spatial-motor models of what they are trying to learn. When they do so, they remember better. Gestures we watch also augment comprehension and learning, of action, of time, of number, and many other concepts. Diagrams and sketches, both those provided to us and those we create, also augment comprehension and creativity. All of the above claims were

substantiated by experiments. Gestures and graphics communicate more directly than purely symbolic words, but because they are in most cases neither discrete nor componential, are less tractable to AI than language.

Finally, we show that the world we live in is designed by our actions. The designs we create in the world reflect the ways we design our minds, into categories, orders, themes, 1-1 correspondences, symmetries, repetitions. The actions that create the designs are reflected in our words and our gestures; the designs are used to represent and communicate organizations of ideas, tables, networks, graphs. We call this cycle, spraction, a contraction of space, action, and abstraction. Actions in space create abstractions.

4   Working groups

4.1   Categorising Evaluation Instruments

Anthony Cohn (University of Leeds, GB)

I present work done jointly with Jose Hernandez-Orallo, for a project initiated by the OECD on the Future of Skills: Understanding the Educational Implications of AI and Robotics. I present and discuss an approach to categorising benchmarks, competitions, tests and evaluation standards as AI evaluation instruments (EI) via a set of 18 facets, which we believe will be valuable in distinguishing and evaluating different proposals for evaluating AI systems. These facets applied to two example AI evaluation instruments: the Arcade Learning Environment (ALE) and the Winograd Schema Challenge (WSC). We plan to conduct further evaluation on the validity and usefulness of these facets by applying them to many more EIs.

4.2   Spatial Humor

Tiansi Dong (Universität Bonn, DE) and Christian Hempelmann (Texas A&M University – Commerce, US)

Spatial Humor refers to the set of humor, such that spatial scene is the topic of two scripts and the spatial knowledge causes the logic opposition. The stimuli of *Spatial Humor* can either be cartoon without language description, or pure text description, or a mixture of both. We work on a novel neural-geometric computational model with rotating spheres as building blocks. This rotating sphere model serves as the semantic representation of *Spatial Humor* which synergistically unifies neural vector embedding and symbolic relations. Its representation power is demonstrated in the following aspects: (1) an object instance is represented by a rotating sphere in the embedding space; (2) object features are represented by rotating axes; (3) a snapshot view is represented by a configuration of spheres [3]; (4) an event [2] is schematized by a triple of a configuration of spheres, representing starting, middle, and end of the event; (5) the script opposition of a *spatial humor* is represented by two overlapped sphere configurations such that the overlapped spheres switch rotating axes. In this way, the motion of spheres embodies the dynamic process of humor understanding and the Script-based Semantic Theory of Humor [1].

References

1 Attardo, S., Hempelmann, C. F., and Di Maio, S. (2002). Script oppositionsand logical mechanisms: Modeling incongruities and their resolutions. Humor,15(1):3 – 46.

2 Zacks, J. and Tversky, B. (2001). Event structure in perception and conception. Psychological Bulletin, 127:3 – 21.

3 Dong, T. (2012). Recognizing Variable Environment – The Theory of Cognitive Prism. Studies in Computational Intelligence, Vol. 388, Springer.

4.3   Towards a Survey of Meaning Representation

Tiansi Dong (Universität Bonn, DE), Anthony Cohn (University of Leeds, GB), Christian Hempelmann (Texas A&M University – Commerce, US), Kanishka Misra (Purdue University – West Lafayette, , US), Jens Lehmann (Fraunhofer IAIS – Dresden, DE), Alexander Mehler (Goethe-Universität – Frankfurt am Main, DE), Tristan Miller (OFAI – Wien, AT), Siba Mohsen (Universität Bonn, DE), Roberto Navigli (Sapienza University of Rome, IT), Julia Rayz (Purdue University – West Lafayette, US) Stefan Wrobel (Universität Bonn, DE) Ron Sun (Rensselaer Polytechnic – Troy, US), and Volker Tresp (Siemens – München, DE)

After the working group on "What is missing in ML&AI to understanding Jokes?", we discussed the possibility to survey the expressiveness on existing models on meaning representation, contrasted by the forecast of existing theories in cognitive science about what is relevant cognitive activities and processes. Spatial stimuli activate the zoo of spatial cells in hippocampus, forming cognitive map or collage in the memory, producing spatial descriptions in languages. We need to survey existing models on Mental Spatial Representation (MSR) in the literature of cognitive psychology. On the other hand, we need to analyse vector embeddings of spatial entities and relations in the large-scaled pre-train world model, and find the gap between MSR and vector embedding via Machine Learning.

4.4   Rotating Sphere Model for NLP

Roberto Navigli (Sapienza University of Rome, IT), Tiansi Dong (Universität Bonn, DE), Thomas Liebig (TU Dortmund, DE), Yong Liu (Outreach Corporation – Seattle, US), Alexander Mehler (Goethe-Universität – Frankfurt am Main, DE), Tristan Miller (OFAI – Wien, AT), Siba Mohsen (Universität Bonn, DE), and Sven Naumann (Universität Trier, DE)

The existing n-ball embedding approach can precisely encode a large symbolic tree structure into tree node embeddings. In this working group, we discussed how to apply the idea of n-ball to solve NLP tasks, in particular, the Word Sense Disambiguation (WSD). WSD is a fundamental task in Natural Language Processing (NLP), which impacts a variety of downstream NLP applications. WSD determines the intended meaning of words in a context. To tackle the WSD task, researchers have been investigating knowledge-based approaches,

supervised, semi-supervised, and unsupervised machine learning. However, those methods encounter a number of limitations, besides their costly computation. We let n-ball rotate, and result in the Rotating Spheres Model (RSM). Using RSM, embeddings of word senses work like gestures of a word. Given a context, the word chooses the best gesture. The WSD is to determine the best rotating axis in a given context. Each rating axis represents a sense that in the predefined sense inventory.

4.5 Joint deductive and inductive reasoning benchmarks

Achim Rettinger (Universität Trier, DE), Mehwish Alam (FIZ Karlsruhe, DE), Anthony Cohn (University of Leeds, GB), Bernado Cuenca Grau (University of Oxford, GB), Mateja Jamnik (University of Cambridge, GB), Thomas Liebig (TU Dortmund, DE), Roberto Navigli (Sapienza University of Rome, IT), and Steffen Staab (Universität Stuttgart, DE)

Data-driven empirical methods, specifically Deep Learning, has dominated the past decade of AI approaches, often to the extent that they allegedly achieve superhuman performance. In this working group we started from NLP benchmarks like GLUE and SuperGLUE and discussed how suited they are to measure language understanding, specifically in contrast to symbolic reasoning tasks. The goal of the working group was about benchmarks that require both statistical learning and deductive reasoning. The discussions in the group started from concrete benchmark tasks and capabilities of existing systems from all areas of AI, like Winograd Schema, Raven's Progressive Matrices, Digital Aristotle, Wolfram Alpha, Watson, humor detection and attempted to find criteria to describe characteristics and facets of benchmarks. Existing facets in related work on categorising evaluation instruments were discussed and an own list was extracted and compiled. Next, the discussions shifted from the benchmarks perspective to the systems perspective and existing approaches to categorizing hybrid systems were reviewed. While the questions of what makes a system a hybrid system, was not ultimately decided, it was agreed that a combination of the task perspective and the systems perspective would be a valuable contribution to the community. The next steps required to obtain more insights in this working groups topic would be a) an abstraction to classes of tasks, b) an abstraction to classes of concrete systems and c) and the extraction of relations between classes of tasks and classes of systems.

5 Open problems

5.1 What would be an aggregated neural model for syllogistic reasoning?

Tiansi Dong (Universität Bonn, DE), Pietro Lio (University of Cambridge, GB), and Ron Sun (Rensselaer Polytechnic – Troy, US)

Animals like monkeys can do syllogistic reasoning [2]. This suggests that syllogistic reasoning does not need language.

Given enough training data, vanilla neural networks can learn syllogism [3, 4]. Without training data, but equipped with a topological map, we can develop a novel neural network for rigorous syllogistic reasoning [5].

What would be an aggregated neural model of the two models satisfying all criteria in [1], such that it starts from learning patterns, making errors, then involving the capability in syllogistic reasoning. If one simulates System 1 for syllogism, and the other simulates System 2 for syllogistic reasoning. What could be the dual-process model [6, 7, 8, 9] for syllogistic reasoning?

References

1 S. Khemlani and P. N. Johnson-Laird (2012). *Theories of the Syllogism: A Meta-Analysis.* Psychological Bulletin 138(3) 427-457.

2 Ferrigno, Stephen and Huang, Yiyun and Cantlon, Jessica F (2021). *Reasoning Through the Disjunctive Syllogism in Monkeys.* Psychological Science 32(2) 1-9.

3 Wang, D., Jamnik, M., and Li'o, P. (2018). *Investigating diagrammatic reasoningwith deep neural networks.* In Diagrams 2018, pages 390 – 398.

4 Wang, D., Jamnik, M., and Li'o, P. (2020). *Abstract diagrammatic reasoningwith multiplex graph networks.* ICLR.

5 Dong, T., Li, C., Bauckhage, C., Li, J., Wrobel, S., and Cremers, A. B.(2020). *Learning syllogism with euler neural-networks.* arXiv preprintar Xiv:2007.07320.

6 Sun, R.. *Integrating Rules and Connectionism for Robust Commonsense Reasoning.* John Wiley and Sons, New York, 1994.

7 Sun, R.. *Duality of the Mind.* Lawrence Erlbaum Associates, Mahwah, NJ., 2002.

8 Sun, R.. *Anatomy of the Mind.* Oxford University Press, Oxford, UK., 2016.

9 Sun, R. & L. Bookman (eds.). *Computational Architectures Integrating Neural and Symbolic Processes.* Kluwer, Needham, MA., 1994.

5.2   How do we infuse human knowledge and machine learning to transform enterprise sales engagement and intelligence?

Yong Liu (Outreach Corporation – Seattle, US)

The digital transformation is accelerating in the sales engagement domain with the applications of machine learning (ML) and artificial intelligence (AI). Given sales engagement is an inherently human-in-the-loop process, there is an increasing need from both business and technology capability sides to adopt a knowledge-infused learning approach that synthesizes human knowledge and statistical machine learning for explainable and trustworthy intelligence.

However, there are a lot of open questions about how we might realize such a vision. For example, how do we build a sales engagement knowledge graph incrementally by considering three types of data in the sales engagement (3Cs): Contacts (persons involved), Contents (multimedia materials including email messages, voice calls, demo videos, video conferences and transcripts, sales and purchase proposals, legal agreement, and contracts etc.) and Contexts (history of the engagement, customer pain points and success stories etc.)? How do we maintain and grow a temporal dynamic knowledge graph that allows time-traveling of the graph with flexibility to trace and incorporate the knowledge evolution? How do we provide turn-by-turn explainable recommendations along the process to guide the sales engagement? All these open questions are exciting areas to explore in the years to come.

5.3   What is missing in ML&AI to understand Jokes?

Alexander Mehler (Goethe-Universität – Frankfurt am Main, DE), Tiansi Dong (Universität Bonn, DE), Thomas Liebig (TU Dortmund, DE), Tristan Miller (OFAI – Wien, AT), Siba Mohsen (Universität Bonn, DE), and Sven Naumann (Universität Trier, DE)

Why current Machine Learning and AI (ML&AI) techniques cannot understand jokes as we humans do? What is missing? The knowledge that is needed to understand jokes is neither in the joke texts, nor in the neural networks. Acquisition and reasoning with commonsense knowledge is still an open problem for Machine Learning and AI. The meaning representation based on embeddings is insufficient. We need meaning representation formats that are beyond vector representations. Vectors are only shadows. Information processing and meaning understanding are embodied. The discussion guides us to develop novel embodied ML&AI techniques to understand *Spatial Jokes* first.

5.4   Can We Diagram the Understanding of Humour?

Tristan Miller (OFAI – Wien, AT), Anthony Cohn (University of Leeds, GB), Tiansi Dong (Universität Bonn, DE), Christian Hempelmann (Texas A&M University – Commerce, US), Siba Mohsen (Universität Bonn, DE), and Julia Rayz (Purdue University – West Lafayette, US)

Cartoons can be understood without language. That is, a suitably arranged scene of simple objects, with no accompanying text, is often enough to make us laugh – evidence that thinking (mental activity) happens before language [4]. This raises the question of non-linguistic diagrammatic representation of spatial humour, along with the mechanism of neural computation. In particular, we raise following questions: (1) How can we diagrammatically formalise spatial humour? (2) How can these diagrammatic formalisms be processed by neural networks? (3) How can this neural computation deliver high-level schema that are similar to the script-opposition semantic theory of humour [2, 1, 3]?

The spatial knowledge encoded in the scene can activate the necessary spatial and non-spatial knowledge. By what neural associative mechanism or process of reasoning do we put this all together to "get" the joke? During the seminar, we aimed to make some headway towards establishing (1) exactly what sort of scene-specific and common-sense knowledge is required to understand any given cartoon, (2) what part of this knowledge could in principle be acquired by existing machine learning (ML) techniques, and which could be acquired or encoded through symbolic structures, (3) what activation process acquires the rest of the knowledge required to interpret the humour, and (4) whether there is a unified representation that could represent this knowledge in a computer's working memory.

References

1   Salvatore Attardo. The General Theory of Verbal Humor. In Salvatore Attardo, editor, *The Routledge Handbook of Language and Humor*. Routledge, New York, NY, 2017.
2   Salvatore Attardo, Christian F. Hempelmann, and Sara Di Maio. Script oppositions and logical mechanisms: Modeling incongruities and their resolutions. *HUMOR: International Journal of Humor Research*, 15(1):3–46, 2002.
3   Julia Taylor Rayz. Scripts in the Ontological Semantic Theory of Humor. In Salvatore Attardo, editor, *Script-Based Semantics: Foundations and Applications*, pages 229–256. De Gruyter Mouton, Berlin, Boston, 2020.
4   Barbara Tversky. *Mind in Motion.* Basic Books, New York, NY, 2019.

Participants

- Tiansi Dong
  Universität Bonn, DE

- Thomas Liebig
  TU Dortmund, DE

- Alexander Mehler
  Goethe-Universität –
  Frankfurt am Main, DE

- Sven Naumann
  Universität Trier, DE

- Achim Rettinger
  Universität Trier, DE

- Steffen Staab
  Universität Stuttgart, DE

- Volker Tresp
  Siemens – München, DE

Remote Participants

- Mehwish Alam
  FIZ Karlsruhe, DE
- Christian Bauckhage
  Fraunhofer IAIS –
  Sankt Augustin, DE
- Sven Behnke
  Universität Bonn, DE
- Jian-Jia Chen
  TU Dortmund, DE
- Anthony Cohn
  University of Leeds, GB
- Armin B. Cremers
  Universität Bonn, DE
- Bernado Cuenca Grau
  University of Oxford, GB
- Asja Fischer
  Ruhr-Universität Bochum, DE
- Xiaoming Fu
  Universität Göttingen, DE
- Jürgen Gall
  Universität Bonn, DE
- Christian Hempelmann
  Texas A&M University –
  Commerce, US
- Mateja Jamnik
  University of Cambridge, GB

- Markus Knauff
  Universität Gießen, DE
- Michael Kramer
  MPI für Radioastronomie, DE
- Jens Lehmann
  Fraunhofer IAIS – Dresden, DE
- Juanzi Li
  Tsinghua University –
  Beijing, CN
- Pietro Lio
  University of Cambridge, GB
- Yong Liu
  Outreach Corporation –
  Seattle, US
- Zhiyuan Liu
  Tsinghua University –
  Beijing, CN
- Xiaoxing Ma
  Nanjing University, CN
- Tristan Miller
  OFAI – Wien, AT
- Siba Mohsen
  Universität Bonn, DE
- Adithya Murali
  University of Illinois –
  Urbana-Champaign, US

- Roberto Navigli
  Sapienza University of Rome, IT
- Max Petrenko
  Amazon Web Services –
  Seattle, US
- Julia Rayz
  Purdue University –
  West Lafayette, US
- Ron Sun
  Rensselaer Polytechnic –
  Troy, US
- Jian Tang
  Mila – Montreal, CA
- Jie Tang
  Tsinghua University –
  Beijing, CN
- Barbara Tversky
  Columbia University –
  New York, US
- Frank van Harmelen
  VU University Amsterdam, NL
- Ran Yu
  Universität Bonn, DE

Integrated Deduction

Edited by

Maria Paola Bonacina[1], Philipp Rümmer[2], and Renate A. Schmidt[3]

1   Università degli Studi di Verona, IT, `mariapaola.bonacina@univr.it`
2   Uppsala University, SE, `philipp.ruemmer@it.uu.se`
3   University of Manchester, GB, `renate.schmidt@manchester.ac.uk`

--- **Abstract** ---

Logical reasoning plays a key role in fields as diverse as verification and synthesis, programming language foundations, knowledge engineering, and computer mathematics. Logical reasoning is increasingly important in intelligent systems, such as decision support systems, agent programming environments, and data processing systems, where deduction may provide explanation, course of action, and the capability of learning from missing information. Problem formalization in these domains typically involves multiple mathematical theories, knowledge bases, and ontologies, all of which may be very large. Problem solving requires both efficient automation and sophisticated human-machine interaction. The thrust of this seminar was that the key to unleash the power of computerized logical reasoning is *integration*, at different abstraction levels.

This seminar offered a forum to discuss the issues related to *integration of deduction* in a diverse range of applications. In terms of reasoning procedures, the presence of both theories and quantifiers in problems from many contexts calls for methodologies to *integrate state-of-the-art SMT solvers and automated theorem provers*. This leads to investigate techniques such as *model-based reasoning* and *semantic guidance*, that were presented and discussed at the seminar. Similarly, the *integration of inference rules for higher-order reasoning* in inference systems that were born for first-order reasoning, such as *superposition*, was prominent among the topics debated at the seminar. At the architectural level, the sheer difficulty of the problems calls for the *integration of provers and solvers* into *interactive reasoning environments*. These range from *higher-order proof assistants* with background reasoners as *hammers*, to *interactive program verifiers*, both widely covered at the seminar in talks and discussions.

The seminar showed how the application of deduction to intelligent systems necessitates the integration of deduction with other paradigms, such as *probabilistic reasoning* and *statistical inferences*. In fact, it emerged from the seminar that even systems that are not natively deductive, such as agent programming environments and industrial tools for ontology-based processing, benefit significantly from the integration of deduction. A clear and shared uptake from the seminar was that *scalability* and *usability* are crucial challenges at all levels of integration. The seminar fully succeded in promoting the exchange of new ideas and suggestions for future research.

1 Executive Summary

Maria Paola Bonacina
Philipp Rümmer
Renate A. Schmidt

This report contains the program and outcomes of the Dagstuhl Seminar 21371 on *Integrated Deduction* that was held at Schloss Dagstuhl, Leibniz Center for Informatics, during September 12–17, 2017. It was the fourteenth in a series of Dagstuhl Deduction seminars held biennially since 1993.

The motivation for this seminar was the following. Automated deduction has developed a wide and diverse range of methods and tools for logico-deductive reasoning. They include SAT solvers,[1] SMT solvers,[2] automated theorem provers, aka ATP systems, proof assistants, aka interactive theorem provers (ITP), as well as libraries of formalized mathematics and formalized knowledge. These methods and tools have found successful application in computing fields as diverse as the *analysis, verification, and synthesis of systems*, *programming language design*, *knowledge engineering*, and *computer mathematics*. However, no method, tool, paradigm, or even reasoning style can solve all problems, or respond to all demands coming from even a single field of application. Therefore, the next grand challenge for automated deduction is *integration*.

Integration occurs and is needed at different abstraction levels. Within deduction itself, integration of deductive engines allows us to build more powerful, more flexible, more expressive reasoners, that can solve more problems with fewer resources, meaning not only memory and computing time, but also human time and human expertise, the latter two often being the most precious of resources. Next, deductive reasoners get integrated into other tools, such as *automated test generators*, *verifying compilers*, or *program synthesizers*, just to name a few. Yet another level of integration occurs when *logico-deductive reasoning* is integrated with other forms of automated reasoning, such as *probabilistic reasoning* and *statistical inference*. This leads to the integration of deduction within *intelligent systems*, such as *decision support systems*, *agent programming environments*, and *data processing systems*. Here deduction may provide explanation, course of action, and the capability of learning from missing information; it may also aid modelling and facilitate agent communication.

The seminar on *Integrated Deduction* successfully covered as many as possible of these integration issues, including:

- Integration of deductive engines into more general automated deductive systems;
- Integration of automated deductive systems into interactive proof assistants;
- Integration of deduction into formal methods tools;
- Integration of deduction for knowledge processing; and
- Integration of deduction into intelligent systems such as agent-based systems.

Furthemore, the seminar investigated a number of key technological and human-related issues, that are largely orthogonal to most integration contexts, affecting both feasibility and deployment of integrated deduction. Examples of such issues are:

- The development of interfaces for integration;

---

[1] SAT solvers are solvers for satisfiability queries in propositional logic, known as the SAT problem.
[2] SMT stands for satisfiability modulo theories.

- The generation of continuous feedback during the run of deductive tools, including also information from intermediate or unsuccessful states;
- The reproducibility of results in the presence of tool updates or imposed resource limits (e.g., available computation time or memory) that may introduce non-determinism; and
- Advanced tradeoff's between performance and expressivity as well as between specialization and genericity.

Practical challenges around integrated deductive systems, including collaboration with non-expert users or access to data sets, were also discussed.

The seminar brought together a diverse audience, including both researchers working in deduction and researchers working in neighbouring areas that make use of deduction. Many participants have experience in building, using, and applying systems with integrated deduction.

Following the tradition of the Dagstuhl *Seminars on Deduction*, most of the program consisted of contributed talks by participants on their research. In this manner, the bottom-up style of the Dagstuhl experience was preserved, allowing for spontaneous contributions as they emerged during the seminar.

However, this seminar was also innovative with respect to tradition, in that it featured *five invited tutorials* on key topics in integrated deduction. These tutorials were valuable in highlighting the state-of-the-art in the integration of deduction systems and in fomenting discussions on challenges and open problems.

The program also featured a hike in the woods and a social dinner in a nearby village, that helped establishing or strengthtening collaborations.

The following section contains the abstracts for most of the talks and tutorials listed in alphabetical order.

<div style="background:#f7b500;">**2**</div> **Table of Contents**

3.1 Higher-order superposition in action (Tutorial)

Alexander Bentkamp (VU University Amsterdam, NL), Jasmin Christian Blanchette (VU University Amsterdam, NL), and Sophie Tourret (INRIA Nancy – Grand Est – Villers-lès-Nancy, FR & MPI für Informatik – Saarbrücken, DE)

Following Blanchette's talk about higher-order superposition, this tutorial delved into higher-order logic topics (syntax, semantics, unification problem), the Zipperposition prover, and the lambda-superposition calculus. To clarify the fine points of the calculus, we ran Zipperposition on actual problems and studied the generated proof diagrams.

3.2 Integrating Optimization Solvers into Proof Assistants

Alexander Bentkamp (VU University Amsterdam, NL)

Optimization is used extensively in engineering, industry, and finance, and various methods are used to transform problems to the point where they are amenable to solution by numerical methods. I present progress towards developing a framework, based on the Lean interactive proof assistant, for designing and applying such reductions in reliable and flexible ways.

3.3 Integrating higher-order reasoning into superposition

Jasmin Christian Blanchette (VU University Amsterdam, NL)

I described our journey, in the past five years, from first-order to higher-order superposition, focusing on the key design decisions, including our focus on a graceful generalization and on refutational completeness. I presented the three main milestones along the way and hinted at some ongoing work to optimize the calculus further.

3.4 Semantically-guided goal-sensitive reasoning: theorem proving and decision procedures (Tutorial)

Maria Paola Bonacina (Università degli Studi di Verona, IT)

SGGS (Semantically-Guided Goal-Sensitive reasoning) is an attractive theorem-proving method for decision procedures, because it generalizes the Conflict-Driven Clause Learning (CDCL) procedure for propositional satisfiability, and it is model-complete in the limit, so that SGGS decision procedures are model-constructing. After summarizing the foundations of SGGS as a theorem-proving method, this tutorial presents recent and ongoing work on SGGS decision procedures for fragments of first-order logic. This includes both negative and positive results about known decidable fragments: for example, SGGS decides the stratified fragment, and hence Effectively PRopositional logic (EPR). SGGS also allows us to discover several new decidable fragments based on well-founded orderings. For most of these new fragments the small model property holds, as the cardinality of SGGS-generated models can be upper bounded, and membership can be tested by applying termination tools for rewriting. A report on experiments with the prototype theorem prover Koala, which is the first implementation of SGGS, closes the presentation.

(SGGS is joint work with David Plaisted; SGGS decision procedures are joint work with Sarah Winkler, who is the author of Koala).

3.5 Proofs in SMT (Tutorial)

Pascal Fontaine (University of Liège, BE)

Proofs have an increasing importance for automated reasoning, and in particular for integration of deduction engines. In this talk, we present our efforts to produce detailed, checkable proofs in the context of SMT solving. This includes producing proofs for the underlying SAT solver, the various theories, quantifier instantiation and formula processing. We conclude by a short introduction to the Alethe concrete format, an attempt at a versatile, easy to use proof format, in the philosophy of the SMT-LIB format.

This talk mentions the work of many, including Haniel Barbosa, Jasmin Blanchette, Mathias Fleury, and Hans-Jörg Schurr.

References
1 Haniel Barbosa, Jasmin Christian Blanchette, and Pascal Fontaine. Scalable fine-grained proofs for formula processing. In Leonardo de Moura, editor, *Automated Deduction – CADE*

26 - 26th International Conference on Automated Deduction, Gothenburg, Sweden, August 6-11, 2017, Proceedings, volume 10395 of *Lecture Notes in Computer Science*, pages 398–412. Springer, 2017.

2   Hans-Jörg Schurr, Mathias Fleury, Haniel Barbosa, and Pascal Fontaine. Alethe: Towards a generic SMT proof format (extended abstract). *CoRR*, abs/2107.02354, 2021.

3   Hans-Jörg Schurr, Mathias Fleury, and Martin Desharnais. Reliable reconstruction of fine-grained proofs in a proof assistant. In André Platzer and Geoff Sutcliffe, editors, *Automated Deduction – CADE 28 – 28th International Conference on Automated Deduction, Virtual Event, July 12-15, 2021, Proceedings*, volume 12699 of *Lecture Notes in Computer Science*, pages 450–467. Springer, 2021.

3.6   AProVE as a Platform for Integrated Deduction

Carsten Fuhs (Birkbeck, University of London, GB)

This talk addresses the following points:

- How does AProVE use other deduction tools? (SAT/SMT solvers, tools for termination and complexity analysis for specific formats, compilers, proof checkers, ...)
- How do other deduction tools use AProVE? (Proof checkers, Knuth-Bendix completion tools, higher-order termination analysis tools, ...)
- Work in Progress: Using Complexity Bounds for Automated Scheduling (joint work with Thaïs Baudon and Laure Gonnord)

References

1   Jürgen Giesl, Cornelius Aschermann, Marc Brockschmidt, Fabian Emmes, Florian Frohn, Carsten Fuhs, Jera Hensel, Carsten Otto, Martin Plücker, Peter Schneider-Kamp, Thomas Ströder, Stephanie Swiderski, René Thiemann. *Analyzing Program Termination and Complexity Automatically with AProVE*. Journal of Automated Reasoning 58(1):3-31, 2017. `https://doi.org/10.1007/s10817-016-9388-y`

2   Thaïs Baudon, Carsten Fuhs, Laure Gonnord. *Parallel Complexity of Term Rewriting Systems*. In Proc. 17th Workshop on Termination (WST'21), pages 39 – 44, 2021. `https://costa.fdi.ucm.es/wst2021/WST2021_proceedings.pdf`, `https://hal.archives-ouvertes.fr/hal-03418400`

3.7   From MCSAT to CDSAT and beyond

Stéphane Graham-Lengrand (SRI – Menlo Park, US)

We present the model-constructing satisfiability approach (MCSAT) to SMT-solving, and illustrate it with the theory of linear arithmetic. We then describe an abstract framework for integrating multiple reasoning modules, called CDSAT for Conflict-Driven Satisfiability, which in particular generalizes MCSAT, CDCL(T), and the equality sharing scheme for disjoint theory combination. CDSAT comes with soundness, completeness, and termination results based on the individual reasoners satisfying appropriate conditions. We discuss proof production for the UNSAT answers of CDSAT. We also present a new algorithm that leverages conflict-driven reasoning to solve quantified satisfiability problems for complete theories; this was implemented in the form of the YicesQS solver, which entered the SMT competition in the BV and NRA logics.

3.8   Efficient local reductions to basic modal logic

Ullrich Hustadt (University of Liverpool, GB)

We present the model-constructing satisfiability approach (MCSAT) to SMT-solving, and illustrate it with the theory of linear arithmetic. We then describe an abstract framework for integrating multiple reasoning modules, called CDSAT for Conflict-Driven Satisfiability, which in particular generalizes MCSAT, CDCL(T), and the equality sharing scheme for disjoint theory combination. CDSAT comes with soundness, completeness, and termination results based on the individual reasoners satisfying appropriate conditions. We discuss proof production for the UNSAT answers of CDSAT. We also present a new algorithm that leverages conflict-driven reasoning to solve quantified satisfiability problems for complete theories; this was implemented in the form of the YicesQS solver, which entered the SMT competition in the BV and NRA logics.

3.9 Conjecture Synthesis, Lemma Discovery and Reasoning

Moa Johansson (Chalmers University of Technology – Göteborg, SE)

Formulating interesting novel conjectures about a new problem domain is a key component of mathematical reasoning. How could this be done by a machine? It is neither a purely deductive problem, nor is it easily solved by data driven machine learning methods. Theory exploration is a technique which tries to address this problem, by combining heuristic search over possible terms, with automated testing to evaluate terms and form equational conjectures. A key for tractability is to start with smaller and more general terms, and exclude any terms which can be reduced by already discovered ones. We demonstrate the QuickSpec system for conjecture generation, and its combination with several theorem provers through the TIP-interface.

References

1   Solrun Halla Einarsdottir, Nicholas Smallbone and Moa Johansson. *Template-based Theory Exploration: Discovering Properties of Functional Programs by Testing.* Proceedings of IFL'20, 2021.
2   Moa Johansson, Koen Claesson and Maximilian Algehed. *Quick Specifications for the Busy Programmer.* Nicholas Smallbone, Journal of Functional Programming, 2017.
3   Moa Johansson and Nicholas Smallbone. *Conjectures, Tests and Proofs: An Overview of Theory Exploration.* Invited paper, Proceedings of 9th Workshop on Verification and Program Transformation, 2021.
4   Moa Johansson. *Automated Theory Exploration for Interactive Theorem Proving.* Conference on Interactive Theorem Proving (ITP), p. 1-11, 2017.

3.10 Using Deduction within Methods for Non-Standard Reasoning in Description Logics

Patrick Koopmann (TU Dresden, DE)

Central reasoning tasks for description logic ontologies such as consistency testing, subsumption checking and classification require deductive reasoning, in the sense that logical consequences from an ontology have to decided or computed. In addition to these standard-reasoning tasks, non-deductive reasoning is useful for many applications of ontologies such as in diagnosis, privacy, as well as for debugging and managing large complex ontologies.

Examples for those non-standard reasoning tasks include abduction, repair, module extraction and forgetting, but also proof generation in the absence of a reasoning calculus. While not directly deduction problems, methods for solving these tasks often use deduction internally. We look at a class of such methods where deduction is used to saturate a set of sentences. That means, in order to solve a non-standard reasoning task, we first translate part of the problem into an appropriate logic, compute all entailments within some class of logical sentences, and then use the saturated set of sentences to compute the solution to our problem. Often, the challenge is to define such a class of sentences that is both bounded by the input and sufficient for constructing the solution, since otherwise our method would either not terminate or produce an incomplete solution. We illustrate this by presenting three examples of such methods, two for solving variants of abduction in description logics [1, 2], and one for ABox repair and anonymization [3], including both published results and current research, which integrate different deduction systems such as the theorem prover SPASS, the datalog engine VLog, as well as a newly developed calculus dedicated to the problem at hand.

References

1 Patrick Koopmann, Warren Del-Pinto, Sophie Tourret, Renate A. Schmidt. *Signature-Based Abduction for Expressive Description Logics.* Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR 2020, pages 592–602, 2020. `https://doi.org/10.24963/kr.2020/59`

2 Fajar Haifani, Patrick Koopmann, Sophie Tourret. *Abduction in $\mathcal{EL}$ via Translation to FOL.* Proceedings of the Second Workshop on Second-Order Quantifier Elimination and Related Topics (SOQE 2021) associated with the 18th International Conference on Principles of Knowledge Representation and Reasoning (KR 2021), pages 46–58, 2021. `http://ceur-ws.org/Vol-3009/paper2.pdf`

3 Franz Baader, Patrick Koopmann, Francesco Kriegel, Adrian Nuradiansyah. *Computing Optimal Repairs of Quantified ABoxes w.r.t. Static $\mathcal{EL}$ TBoxes.* Automated Deduction – CADE 28 – 28th International Conference on Automated Deduction, pages 309–326, 2021. `https://doi.org/10.1007/978-3-030-79876-5_18`

3.11   Verified Proof Checkers

Magnus Myreen (Chalmers University of Technology – Göteborg, SE)

In this talk, I described my work on program verification and particularly my focus on proving functional correctness down to the machine code that runs the programs, including for proof checkers. I have worked on and have supervised work on several checkers, but only talked about (1) my old work on proving end-to-end correctness of Jared Davis's Milawa prover, and (2) recent work on a LPR/LRAT checker for UNSAT proofs. My talk included a description of the CakeML project, which was the context of (2). I used demos to show what the tools look like when running.

References

1   Yong Kiam Tan, Marijn J. H. Heule, and Magnus O. Myreen. `cake_lpr:` *Verified Propagation Redundancy Checking in CakeML.* Tools and Algorithms for the Construction and Analysis of Systems (TACAS), 2021.

2   Jared Davis, Magnus O. Myreen.. *The self-verifying Milawa theorem prover is sound (down to the machine code that runs it).* Journal of Automated Reasoning (JAR), 2015.

3   Yong Kiam Tan, Magnus O. Myreen, Ramana Kumar, Anthony Fox, Scott Owens, and Michael Norrish. *The verified CakeML compiler backend.* Journal of Functional Programming (JFP), 2019.

3.12   On reasoning about multisets (Tutorial)

Ruzica Piskac (Yale University – New Haven, US)

When reasoning about container data structures that can hold duplicate elements, multisets are the obvious choice for representing the data structure abstractly. However, the decidability and complexity of constraints on multisets has been much less studied and understood than for constraints on sets. In this presentation, we outline an efficient decision procedure for reasoning about multisets with cardinality constraints. We describe how to translate, in linear time, multisets constraints to constraints in an extension of quantifier-free linear integer arithmetic, which we call LIA*. LIA* extends linear integer arithmetic with unbounded sums over values satisfying a given linear arithmetic formula. We show how to reduce a LIA* formula to an equisatisfiable linear integer arithmetic formula. However, this approach requires an explicit computation of semilinear sets and in practice it scales poorly even on simple benchmarks. We then describe a recent more efficient approach for checking satisfiability of LIA*. The approach is based on the use of under- and over-approximations of LIA* formulas. This way we avoid the space overhead and explicitly computing semilinear sets. Finally, we report on our prototype tool which can efficiently reason about sets and multisets formulas with cardinality constraints.

3.13   Constrained Horn Clauses in Verification: 11 Years later

Philipp Rümmer (Uppsala University, SE)

Constrained Horn Clauses (CHC) have over the last decade emerged as a uniform framework for reasoning about different aspects of software safety. CHCs form a fragment of first-order logic, modulo various background theories, in which models can be constructed effectively with the help of model checking algorithms. In the talk I have given an overview of CHCs, including their use in program verification and the recently established competition CHC-COMP [2, 1]. I have then presented some of our work on the development of CHC solvers that can handle relevant theories such as integers, bit-vectors, and ADTs [3], and outlined challenges remaining in the area.

References

1 Grigory Fedyukovich and Philipp Rümmer. Competition report: CHC-COMP-21. *CoRR*, abs/2109.04635, 2021.

2 Philipp Rümmer. Competition report: CHC-COMP-20. In Laurent Fribourg and Matthias Heizmann, editors, *Proceedings 8th International Workshop on Verification and Program Transformation and 7th Workshop on Horn Clauses for Verification and Synthesis, VPT/H-CVS@ETAPS 2020, Dublin, Ireland, 25-26th April 2020*, volume 320 of *EPTCS*, pages 197–219, 2020.

3 Hossein Hojjat and Philipp Rümmer. The ELDARICA Horn solver. In Nikolaj Bjørner and Arie Gurfinkel, editors, *2018 Formal Methods in Computer Aided Design, FMCAD 2018, Austin, TX, USA, October 30 - November 2, 2018*, pages 1–7. IEEE, 2018.

3.14 Development and integration of deduction for the medical ontology SNOMED CT

Renate Schmidt (University of Manchester, GB)

In my presentation I talked about our experiences and ongoing work in a Partnership with SNOMED Intl to develop bespoke content extraction software for the medical ontology SNOMED CT. SNOMED CT is a large knowledge base of standardised, precise definitions of clinical terms and medical codes for use in health care systems in many countries. It has long been an aim to have the capability to compute smaller self-contained extracts of the ontology that are restricted to a narrow focus, for example, kidney diseases, dentistry or nursing practice. Such subontologies would make it easier to reuse and share content, to assist with new ontology creation, quality assurance, ontology update and debugging. In addition, reasoning tasks such as querying and classification take less time to execute over a smaller extract than over the original ontology. Our subontology builder takes as input a set of focus concepts, which could be a reference set that the user is interested in, and generate a subontology that can be used in place of SNOMED CT in a specific application. The idea is that the subontology equivalently captures the semantics of these focus concepts and their relationship to other concepts in a certain abstracted form. Because the subtype relationships between concepts are so important for SNOMED enabled search in patient data, a further requirement is that, the concept hierarchy over the focus and supporting concepts in SNOMED CT must be included in the subontology. Subontologies computed for a collection of standard lists of clinical concepts and reference sets for specialities can be viewed in the new subontology browser here:

        `https://iaa.snomed.tools`

and compared with SNOMED CT here:

        `https://browser.ihtsdotools.org/`

The research was undertaken in EPSRC IAA Project 290 funded by the UK EPSRC, the University of Manchester and IHTSDO.

3.15   Tackling Commonsense Reasoning Problems with a First-Order Logic Reasoner

Claudia Schon (Universität Koblenz-Landau, DE)

Joint work of Ulrich Furbach, Teresa Krämer, Claudia Schon
Main reference Ulrich Furbach, Teresa Krämer, Claudia Schon: "Names Are Not Just Sound and Smoke: Word Embeddings for Axiom Selection", in Proc. of the Automated Deduction – CADE 27 – 27th International Conference on Automated Deduction, Natal, Brazil, August 27-30, 2019, Proceedings, Lecture Notes in Computer Science, Vol. 11716, pp. 250–268, Springer, 2019.
URL https://doi.org/10.1007/978-3-030-29436-6_15

This talk reports on our experiences using a first-order logic reasoner to solve commonsense reasoning benchmark problems like the Choice of Plausible Alternatives (COPA) Challenge. Most approaches in this area rely on pre-trained language models and do not use reasoning. In contrast, we combine a deductive theorem prover with large background knowledge bases and machine learning. Since these background knowledge bases represent a very large amount of knowledge from a wide variety of domains, they cannot be used by the reasoner as a whole. Selection methods are used to select suitable background knowledge for a specific task. In the area of commonsense reasoning, these selection methods can benefit from the integration of statistical techniques such as word embeddings. We show that incorporating word embeddings into the selection process enables the selection of background knowledge that is thematically appropriate to a commonsense reasoning task. This approach is implemented and we present experimental results.

3.16   Deduction as a Service

Stephan Schulz (Duale Hochschule Baden-Württemberg – Stuttgart, DE)

Joint work of Mohamed Hassona, Stephan Schulz
Main reference Mohamed Hassona, Stephan Schulz: "Deduction as a Service", in Proc. of the 5th Workshop on Practical Aspects of Automated Reasoning co-located with International Joint Conference on Automated Reasoning (IJCAR 2016), Coimbra, Portugal, July 2nd, 2016, CEUR Workshop Proceedings, Vol. 1635, pp. 32–40, CEUR-WS.org, 2016.
URL http://ceur-ws.org/Vol-1635/paper-04.pdf

Traditionally, problems for automated theorem provers were small, tightly specified, and often with long, complicated proofs. In contrast, many more recent problems are automatically generated from large artefacts, or posed over large common-sense or mathematical knowledge bases, but often with rather simple and short proofs.

In these cases, the total time for a proof attempt is often dominated by the overhead of parsing and premise selection. Offering deduction over a large, persistent knowledge base as a service can amortise this overhead, reducing the time of single proof attempts to a level acceptable even for interactive use.

I describe the architecture of such a system, some of the practical challenges, and the state of the art so far.

3.17 Integrating Machine Learning into Saturation-based ATPs (Tutorial)

Martin Suda (Czech Technical University – Prague, CZ)

Applying the techniques of machine learning (ML) promises to dramatically improve the performance of modern automatic theorem provers (ATPs) and thus to positively impact their applications. The most successful avenue in this direction explored so far is machine-learned clause selection guidance, where we learn to recognize and prefer for selection clauses that look like those that contributed to a proof in past successful runs. In this talk I present Deepire, an extension of the ATP Vampire where clause selection is guided by a recursive neural network (RvNN) for classifying clauses based solely on their derivation history.

3.18 On what is wrong with higher-order SMT and what we are doing to fix it

Sophie Tourret (INRIA Nancy – Grand Est – Villers-lès-Nancy, FR & MPI für Informatik – Saarbrücken, DE)

Recent work has extended ground SMT solvers to higher-order logic (HOL), but SMT solving has yet to show its full power in HOL. It remains to lift quantifier instantiation algorithms to perform higher-order unification. As a consequence, widely used instantiation techniques, such as trigger- and particularly conflict-based instantiation, can only be applied in a limited manner. Congruence closure with free variables (CCFV) is a decision procedure for the E-ground (dis-)unification problem, which is at the heart of these instantiation techniques. Here, as a first step towards fully supporting trigger- and conflict-based instantiation in HOL, we define the E-ground (dis-)unification problem in λ-free higher-order logic (λfHOL), an extension of first-order logic where function symbols may be partially applied and functional variables may occur, and extend CCFV to solve it. To improve scalability in the context of handling higher-order variables, we rely on a SAT encoding of the CCFV search. We present a solution reconstruction procedure so that the propositional models lead to solutions for the respective E-ground (dis-)unification problem. This is instrumental to fully port trigger- and conflict-based instantiation to be fully applied in λfHOL.

3.19 Linear-Time Verification of Data-Aware Dynamic Systems with Arithmetic

Sarah Winkler (University of Verona, IT)

Counter systems have been investigated in formal methods, database theory, and AI. Though model checking of such systems is undecidable since two-counter machines can be simulated, different decidable classes have been discovered by restricting the constraint language and/or the control flow. This talk presents a new, abstract criterion for the decidability of linear-time verification of such systems, called finite summary, which guarantees the existence of a faithful finite-state abstraction. We demonstrate that several decidability conditions studied in formal methods and database theory can be seen as concrete, checkable instances of this property. To this end, we exploit results from SMT, and automated reasoning in general. Finally, we show how the finite summary property leads to modularity results: a system enjoys finite summary if it can be partitioned appropriately into smaller systems that possess the property. Our results allow us to analyze systems that cannot be handled by earlier approaches.

Participants

- Erika Abraham
  RWTH Aachen University, DE
- Alexander Bentkamp
  VU University Amsterdam, NL
- Jasmin Christian Blanchette
  VU University Amsterdam, NL
- Maria Paola Bonacina
  University of Verona, IT
- Pascal Fontaine
  University of Liège, BE
- Carsten Fuhs
  Birkbeck, University of
  GLondon, GB
- Stéphane Graham-Lengrand
  SRI – Menlo Park, US
- Ullrich Hustadt
  University of Liverpool, GB

- Moa Johansson
  Chalmers University of
  Technology – Göteborg, SE
- Patrick Koopmann
  TU Dresden, DE
- Magnus Myreen
  Chalmers University of
  Technology – Göteborg, SE
- Ruzica Piskac
  Yale University – New Haven, US
- Philipp Rümmer
  Uppsala University, SE
- Renate Schmidt
  University of Manchester, GB
- Claudia Schon
  Universität Koblenz-Landau, DE

- Stephan Schulz
  Duale Hochschule
  Baden-Württemberg –
  Stuttgart, DE
- Alexander Steen
  University of Luxembourg, LU
- Martin Suda
  Czech Technical University –
  Prague, CZ
- Sophie Tourret
  INRIA Nancy – Grand Est –
  Villers-lès-Nancy, FR &
  MPI für Informatik –
  Saarbrücken, DE
- Sarah Winkler
  University of Verona, IT

Behavioural Types: Bridging Theory and Practice

Edited by

Mariangiola Dezani[1], Roland Kuhn[2], Sam Lindley[3], and Alceste Scalas[4]

1   **University of Turin, IT, `dezani@di.unito.it`**
2   **Actyx AG – München, DE, `roland@actyx.io`**
3   **University of Edinburgh, GB, `sam.lindley@ed.ac.uk`**
4   **Technical University of Denmark – Lyngby, DK, `alcsc@dtu.dk`**

Abstract

Behavioural types specify the way in which software components interact with one another. Unlike data types (which describe the structure of data), behavioural types describe communication protocols, and their verification ensures that programs do not violate such protocols. The behavioural types research community has developed a flourishing literature on communication-centric programming, exploring many directions. One of the most studied behavioural type systems are session types, introduced by Honda et al. in the '90s, and awarded with prizes for their influence in the past 20 and 10 years (by the ESOP and POPL conferences, respectively). Other varieties of behavioural types include typestate systems, choreographies, and behavioural contracts; research on verification techniques covers the spectrum from fully static verification at compile-time to fully dynamic verification at run-time.

In the last decade, research on behavioural types has shifted emphasis towards practical applications, using both novel and existing programming languages (e.g., Java, Python, Go, C, Haskell, OCaml, Erlang, Scala, Rust). An earlier Dagstuhl Seminar, 17051 "Theory and Applications of Behavioural Types" (January 29–February 3, 2017), played an important role in coordinating this effort. Yet, despite the vibrant community and the stream of new results, the use of behavioural types for mainstream software development and verification remains limited.

This limitation is largely down to the rapid pace at which mainstream industrial practice for the design and development of concurrent and distributed systems evolves, often resulting in substantial divergence from academic research. In the absence of established tools to express communication protocols, widely used implementations concentrate solely on scalability and reliability. The flip side is that these systems are either overly loose, supporting any conceivable communication structure (via brokers), or overly restricted, supporting only simple request-response protocols (like HTTP or RPC).

In this seminar, experts from academia and industry explored together how best to bridge the gap between theory and mainstream practice. They tackled challenges that are fundamental in practical systems development, but are rarely or only partially addressed in the behavioural types literature – in particular, *failure handling*, *asynchronous communication*, and *dynamic reconfiguration*. Moreover they explored how the tools of behavioural types and programming languages theory (such as *linearity*, *gradual types*, and *dependent types*) can help to address these challenges.

1 Executive Summary

Mariangiola Dezani (University of Turin, IT)
Roland Kuhn (Actyx AG – München, DE)
Sam Lindley (University of Edinburgh, GB)
Alceste Scalas (Technical University of Denmark – Lyngby, DK)

This seminar followed the earlier Dagstuhl Seminar 17051 "Theory and Applications of Behavioural Types". Whereas Seminar 17051 was quite broad, encompassing both theory and practice across a wide range of areas relating to behavioural types, this seminar was much more focused, concentrating on how best to enable the use of behavioural types for practical programming.

Initial preparations

We gathered initial lists of proposed talks and breakout topics prior to the start of the seminar via an online form. We added to these throughout the week. We scheduled talks and breakout groups daily depending on audience interest and participant availability. The first part of the week was primarily talks, with ample time for stimulating discussions; the second part included more time for breakout sessions.

Hybrid seminar logistics

Due to the SARS-CoV-2 pandemic, the seminar was organised in hybrid format, with both in-person and remote participants. As the virtual participants came from a wide range of time zones (from central US to Japan) we gave special consideration to the time slot 2pm–4pm CEST during which everyone could attend. Those in Europe and Japan were able to attend morning sessions and those in Europe and America to attend further afternoon sessions (and a special evening session on Monday).

In order to run a successful hybrid Dagstuhl seminar, we made essential use of the dedicated equipment available at Dagstuhl: a Zoom-based streaming setup, with multiple cameras and ceiling microphones in the seminar room. All talks were live-streamed to both virtual and in-person participants. Talks were recorded so that virtual participants from incompatible time zones could catch up, then deleted at the end of the week. Larger hybrid breakout sessions were held in the main seminar room, and smaller ones elsewhere using a more ad hoc setup.

Moreover, all participants (local and virtual) were invited to use Zulip (a chat application) to exchange messages and files, pose questions during presentations, and remain informed on the upcoming events, group activities, and schedule updates.

Activities and outcomes

Throughout the seminar, the participants gathered in focused breakout groups: the findings of the breakout groups are described in more detail elsewhere in the report. Here is a brief summary:

Typing non-channel-based models allowed researchers with a wide range of perspectives and backgrounds to exchange their views. A key observation was that modern concurrent systems that coordinate via streams of events are difficult to analyse and verify with using existing approaches, and new formalisms are needed.

Logic-based approaches reviewed the state of the art, and discussed new directions. One of the conclusions is that more research is needed to relate concurrent and distributed systems to a broader range of logics beyond classical and intuitionistic linear logic (which are the focus of most current publications).

Type-informed recovery strategies explored failure handling at different levels (from network to application), and summarised several open questions not addressed in existing work.

Session types with untrusted counter-parties focused on how to ensure that different processes interact under compatible protocols, establishing the beginning of new work on monitoring and adaptation.

Join patterns / synchronisation – the next generation collected a survey of various attempts to integrate join patterns in programming languages, and discussed why they have not yet become mainstream. The discussion highlighted the need for exploring the connections between join patterns and linear logic, and the use of the join calculus as a reference for new implementation attempts.

The participants of several breakout groups have agreed to continue their work and collaboration after the seminar.

In addition to these more structured breakout sessions there were further lively improvised meetings and discussions (especially after dinner) which are not summarised in the report.

Overall, we believe that the seminar activities were a success. Unfortunately the hybrid format did pose a barrier for remote participants, especially those in different time zones. But on the positive side, for many participants this was their first Dagstuhl seminar, and for the in-person participants it was their first in-person scientific gathering after many months of virtual events due to the SARS-CoV-2 pandemic: their feedback has been enthusiastic.

At the end of the seminar the participants agreed to remain in contact to continue the discussions, and foster new collaborations. There was strong enthusiasm for organising a follow-up Dagstuhl seminar in the future, perhaps taking place in about two years time. To enable future collaborations the participants:

1. created a GitHub organisation where all seminar participants (and other researchers invited later) can exchange references and materials;
2. agreed to use the seminar's Zulip chat (mentioned above) as a starting point to set up a more permanent solution for continuing the interactions and exchanges (e.g., a mailing list);
3. nominated four people who will propose a new seminar, building upon the results of this one.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Session Logical Relations for Noninterference

Stephanie Balzer (Carnegie Mellon University – Pittsburgh, US)

In this talk I introduce the audience to linear session types through the lens of noninterference. Session types, as the types of message-passing concurrency, naturally capture what information is learned by the exchange of messages, facilitating the development of a flow-senstive information flow control (IFC) type system guaranteeing noninterference. Noninterference ensures that an observer (adversary) cannot infer any secrets from made observations. I will explain the key ideas underlying the development of the IFC type system as well as the construction of the logical relation conceived to prove noninterference. The type system is based on intuitionistic linear logic and enriched with possible worlds to impose invariants on run-time configurations of processes, leading to a stratification in line with the security lattice. The logical relation generalizes existing developments for session-typed languages to open configuration to allow for a more subtle statement of program equivalence.

3.2 Session Types for Runtime Verification

Christian Bartolo Burló (Gran Sasso Science Institute, IT)

Communication is central to present day computation. The expected communication protocols between parties can be formalised as session types, serving as specifications which systems can be verified against. We present our work on the runtime verification of communicating systems using session types, where we investigate their *monitorability* qualities in [1] and augment them with probabilities in [2].

From the work in [1], we show that it is impossible to achieve both *sound* (*i.e.,* only flag ill-typed processes) and *complete* (*i.e.,* flag all ill-typed processes) monitors for verifying the interaction between black-box components. Correspondingly, we prove that our autogenerated session monitors are sound and *weakly*-complete: *i.e.*, the monitors get stuck upon certain violations to the session type. On the practical side, we present a Scala toolkit, `STMonitor` [3], for the automatic generation of session monitors following our formal model. These executable monitors can be used as proxies to instrument communication across black-box processes written in any programming language. We also present the results of a series of benchmarks, showing that the synthesised monitors only introduce limited overheads.

Finally, we present a tool-based methodology from [2] that extends `STMonitor` by synthesising monitors from *probabilistic session types*. These types have each choice point augmented with a probability distribution describing how often each choice should be taken.

The synthesised monitors infer the probabilistic behaviour of a system at runtime, and based solely on the evidence observed up to the current point of execution, issue warnings when the observed behaviour deviates from the one specified by the type.

References

1   Christian Bartolo Burlò, Adrian Francalanza, and Alceste Scalas. *On the Monitorability of Session Types, in Theory and Practice.* ECOOP, LIPIcs, vol. 194, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, pp. 20:1–20:30.

2   Christian Bartolo Burlò, Adrian Francalanza, Alceste Scalas, Catia Trubiani, and Emilio Tuosto. *Towards Probabilistic Session-Type Monitoring.* COORDINATION, Lecture Notes in Computer Science, vol. 12717, Springer, 2021, pp. 106–120.

3   Christian Bartolo Burlò, Adrian Francalanza, and Alceste Scalas. *On the Monitorability of Session Types, in Theory and Practice (Artifact).* Dagstuhl Artifacts Ser. 7 (2021), no. 2, 02:1–02:3.

3.3   A Model of Actors and Grey Failures

Laura Bocchi (University of Kent – Canterbury, GB) and Laura Voinea (University of Kent – Canterbury, GB)

We report on ongoing work on defining a model of failures for distributed systems, as a first step towards better detection and recovery. Looking at failures along an unpredictability axis, at the two extremes of the spectrum we find fail-stop (a component either works correctly or stops, and this latter case can be recognised and thus dealt with) and byzantine failure (a component behaves arbitrarily). In practice, systems can experience behaviours that lie somewhere between these two extremes, and this is often called grey failure: the system appears to be functional, but its overall performance is degraded in some way that may anticipate the full, fail-stop, failure of the system or some of its components.

In the last decade, several kinds of grey failures have been studied, such as transient failures (e.g., a component is down at periodic intervals), partial failures (only some subcomponents are affected), and slowdowns [2].

The symptoms of a grey failure tend to be subtle and ambiguous, involve any layer of the stack, and be signalled by different parts of the system having different perceptions of the health of some component, i.e., differential observation [1].

We present a model of grey failures for actor-based systems. Our model of failures consists of three inter-dependent models: (1) an (actor-based) systems model based on a process calculus, (2) a "curse" model of injected failures, and (3) a model awareness that components of the system have of each other, based on monitoring.

In (2) each curse is analogous to a trace or test of the system. Interesting developments include:

- establishing links to probabilistic functions (e.g., modelling patterns of failure distributions in real systems) as well as generalising to symbolic notions of curse to make model checking more tractable.
- appropriate definition of quality of a diagnosis, such as soundness, completeness, and timeliness, with respect to the injected failures.
- appropriate definition of recovery.

References

1 Huang et al. *Gray Failure: The Achilles' Heel of Cloud-Scale Systems.* In Proc. HotOS. Association for Computing Machinery, Whistler, BC, Canada, 2017

2 Gunawi et al. *Fail-Slow at Scale: Evidence of Hardware Performance Faults in Large Production Systems.* ACM Trans. Storage, vol. 14, no. 3, 2018

3.4   Quantitative Types in Idris 2

Edwin Brady (University of St Andrews, GB)

Dependent types allow us to express precisely what a function is intended to do. Recent work on Quantitative Type Theory (QTT) extends dependent type systems with linearity, also allowing precision in expressing when a function can run. This is promising, because it suggests the ability to design and reason about resource usage protocols, such as we might find in distributed and concurrent programming, where the state of a communication channel changes throughout program execution. As yet, however, there has not been a full-scale programming language with which to experiment with these ideas. Idris 2 is a new version of the dependently typed language Idris, with a new core language based on QTT, supporting linear and dependent types. This talk described Idris 2, and how QTT has influenced its design. I gave examples of the benefits of QTT in practice including: expressing which data is erased at run time, at the type level; and, resource tracking in the type system leading to type-safe concurrent programming with session types.

3.5   Global Types and Event Structure Semantics for Asynchronous Multiparty Sessions

Ilaria Castellani (INRIA – Sophia Antipolis, FR), Mariangiola Dezani (University of Turin, IT), and Paola Giannini

We propose an interpretation of asynchronous multiparty sessions as Flow Event Structures. We also introduce a new notion of type for asynchronous multiparty sessions, ensuring the expected properties for sessions, including progress.

Our types, which reflect asynchrony more directly than standard global types and are more permissive, are themselves interpreted as Prime Event Structures.

The main result is that the Event Structure interpretation of a session is equivalent, when the session is typable, to the Event Structure interpretation of its type.

3.6   An Overview of Explicit Cancellation

Simon Fowler (University of Glasgow, GB)

Safely implementing behavioural types typically requires some form of linearity, typically in the form of a linear type system, in order to rule out errors such as using an endpoint more than once or failing to complete a set of actions.

Unfortunately, linear type systems are difficult to integrate with exceptions or failures, which are inevitable in real-world applications. This talk gives an overview of explicit cancellation as introduced by Mostrous & Vasconcelos in 2014, and how the idea has since been applied to support exception handling in functional languages with applications in web programming, graphical user interfaces, and actor systems.

3.7   The STARDUST project: Session Types for Reliable Distributed Systems

Simon Gay (University of Glasgow, GB)

The STARDUST project (Session Types for Reliable Distributed Systems) is funded by the UK EPSRC (grants EP/T014512/1, EP/T014628/1, EP/T014709/1) from 1st October 2020 to 30th September 2024. It is a collaboration between the University of Glasgow, the University of Kent and Imperial College London. The key objective is to combine the communication-structuring mechanism of session types with the scalability and fault-tolerance of actor-based software architectures. The result will be a well-founded theory of reliable actor programming, supported by a collection of libraries and tools, and validated on a range of case studies. Key aims are to deliver tools that provide lightweight support for developers – e.g. warning of potential issues – and to allow developers to continue to use established idioms. By doing so we aim to deliver a step change in the engineering of reliable distributed software systems.

3.8 A Multiparty Session Typing Discipline for Fault-Tolerant Event-Driven Distributed Programming

Raymond Hu (Queen Mary University of London, GB)

This paper presents a formulation of multiparty session types (MPSTs) for practical fault-tolerant distributed programming. We tackle the challenges faced by session types in the context of distributed systems involving asynchronous and concurrent partial failures – such as supporting dynamic replacement of failed parties and retrying failed protocol segments in an ongoing multiparty session – in the presence of unreliable failure detection. Key to our approach is that we develop a novel model of event-driven concurrency for multiparty sessions. Inspired by real-world practices, it enables us to unify the session-typed handling of regular I/O events with failure handling and the combination of features needed to express practical fault-tolerant protocols. Moreover, the characteristics of our model allow us to prove a global progress property for well-typed processes engaged in multiple concurrent sessions, which does not hold in traditional MPST systems.

To demonstrate its practicality, we implement our framework as a toolchain and runtime for Scala, and use it to specify and implement a session-typed version of the cluster management system of the industrial-strength Apache Spark data analytics framework. Our session-typed cluster manager composes with other vanilla Spark components to give a functioning Spark runtime; e.g., it can execute existing third-party Spark applications without code modification. A performance evaluation using the TPC-H benchmark shows our prototype implementation incurs an average overhead below 10%.

3.9 Papaya: Global Typestate Analysis of Aliased Objects

Mathias Jakobsen (University of Glasgow, GB) and Ornela Dardha (University of Glasgow, GB)

Typestates are state machines used in object-oriented programming to specify and verify correct order of method calls on an object. To avoid inconsistent object states, typestates systems often enforce linear typing, which eliminates – or at best limits – aliasing. However, aliasing is an important feature in programming, and the state-of-the-art on typestates is too restrictive if we want typestates to be adopted in real-world software systems.

In this talk, we present a type system for an object-oriented language with typestate annotations, which allows for unrestricted aliasing, and as opposed to previous approaches it does not require linearity constraints. The typestate analysis is global and tracks objects throughout the entire program graph, which ensures that well-typed programs conform to and complete the declared protocols.

3.10 Session Types as Program Logics

Eduard Kamburjan (University of Oslo, NO)

Main reference Eduard Kamburjan: "Behavioral Program Logic", in Proc. of the Automated Reasoning with
Analytic Tableaux and Related Methods – 28th International Conference, TABLEAUX 2019,
London, UK, September 3-5, 2019, Proceedings, Lecture Notes in Computer Science, Vol. 11714,
pp. 391–408, Springer, 2019.
URL https://doi.org/10.1007/978-3-030-29026-9_22

This talk was based on a conference presentation at TABLEAUX 2019. We present Behavioral Program Logic (BPL), a dynamic logic for trace properties that incorporates concepts from behavioral types and allows reasoning about nonfunctional properties within a sequent calculus. BPL uses behavioral modalities, to verify statements against behavioral specifications. Behavioral specifications generalize both postconditions and behavioral types. They can be used to specify other static analyses, e.g., data flow analyses. This enables deductive reasoning about the results of multiple analyses on the same program, potentially implemented in different formalisms. Our calculus for BPL verifies the behavioral specification gradually, as common for behavioral types. This vastly simplifies specification, calculus and composition of local results. We present a sequent calculus for object-oriented actors with futures that integrates a pointer analysis and bridges the gap between behavioral types and deductive verification.

3.11 Priorities as a Graded Monad

Wen Kokke (University of Edinburgh, GB) and Ornela Dardha (University of Glasgow, GB)

Main reference Wen Kokke, Ornela Dardha: "Prioritise the Best Variation", in Proc. of the Formal Techniques for
Distributed Objects, Components, and Systems – 41st IFIP WG 6.1 International Conference,
FORTE 2021, Held as Part of the 16th International Federated Conference on Distributed
Computing Techniques, DisCoTec 2021, Valletta, Malta, June 14-18, 2021, Proceedings, Lecture
Notes in Computer Science, Vol. 12719, pp. 100–119, Springer, 2021.
URL https://doi.org/10.1007/978-3-030-78089-0_6

In this talk, I will present PGV, a variant of Wadler's GV which decouples channel creation from thread spawning, and restores deadlock freedom by adding priorities. Notably, I will show how PGV embeds deadlock free communication and concurrency primitives with priorities in a standard linear functional language using a graded monad.

References

1 Kokke W., Dardha O. (2021) Prioritise the Best Variation. In: Peters K., Willemse T.A.C. (eds) Formal Techniques for Distributed Objects, Components, and Systems. FORTE 2021. Lecture Notes in Computer Science, vol 12719. Springer, Cham. `https://doi.org/10.1007/978-3-030-78089-0_6`
2 Wen Kokke and Ornela Dardha. 2021. Deadlock-free session types in linear Haskell. Proceedings of the 14th ACM SIGPLAN International Symposium on Haskell. Association for Computing Machinery, New York, NY, USA, 1–13. DOI:`https://doi.org/10.1145/3471874.3472979`

3.12 Asymmetric Replicated State Machines

Roland Kuhn (Actyx AG – München, DE), Hernán Melgratti (University of Buenos Aires, AR), and Emilio Tuosto (Gran Sasso Science Institute, IT)

A factory hall is a place where many humans and machines work together to turn input materials into finished goods. The efficiency of this collaboration – structured into many loosely coupled cells – is of vital importance to the business, hence our system favours availability and local progress over (global) correctness. Conflicts are allowed and can be recognised and compensated.

We present a formal model for machines emitting events to their local logs, where communication occurs eventually by shipping log prefixes between machines. A global type governs the desired protocol and can – if well-formed – be projected to local Mealy machines who will then faithfully realise the protocol.

3.13 Choreographic Programming in Choral

Fabrizio Montesi (University of Southern Denmark – Odense, DK)

This talk is an introduction to Choral (`https://www.choral-lang.org`), the first language for programming choreographies (multiparty protocols) based on mainstream programming abstractions: in Choral, choreographies are objects [2].

Choral's interpretation of choreographies is made possible by new types that enhance standard object types with a notion of locality. Every object is located at some roles (Alice, Bob, etc.), which denotes that the object is implemented collaboratively by them. Thus, objects become choreographic.

Choral is a choreographic programming language [1]: given a choreography that defines interactions among some roles, an implementation for each role in the choreography is automatically generated by a compiler. These implementations are libraries in pure Java, which developers can modularly compose in their own programs to participate correctly in choreographies. Crucially, Choral gives back to the programmer control over the APIs exposed to the users of the generated libraries. For the first time in the application of choreographic languages, this feature enables the generation of libraries that support information hiding, in the sense that the generated libraries hide the communication behaviour that they enact. An important consequence is that updates to the communication behaviour of a choreography might not alter the APIs of the generated libraries, avoiding the need for updating the client code that uses the (code generated from the) choreography.

Leveraging the interpretation of choreographies as objects, Choral brings higher-order composition to choreographic programming. The key novelty is that Choral allows for higher-order composition without the need for global synchronisations or central coordination, which is required by other current models. Choral's foundations have been recently modelled as an extension of the λ-calculus, Chorλ [3]. Chorλ has new reduction and rewriting rules that formalise the principles that underpin decentralised higher-order composition of choreographies.

References
1   Fabrizio Montesi. *Choreographic Programming*. PhD Thesis, IT University of Copenhagen, 2013.
2   Saverio Giallorenzo and Fabrizio Montesi and Marco Peressotti. *Choreographies as Objects*. CoRR abs/2005.09520, 2020.
3   Luís Cruz-Filipe and Eva Graversen and Lovro Lugović and Fabrizio Montesi and Marco Peressotti. *Choreographies as Functions*. CoRR abs/2111.03701, 2022.

3.14   Effpi: verified message-passing programs in Scala 3

Alceste Scalas (Technical University of Denmark – Lyngby, DK)

I will talk about Effpi: an experimental toolkit for strongly-typed concurrent and distributed programming in Scala 3.

Effpi addresses a main challenge in the development of concurrent programs: errors like protocol violations, deadlocks, and livelocks are often spotted late, at run-time, when applications are tested or (worse) deployed. Effpi aims at finding such errors early, when code is written and compiled.

Effpi provides: (1) a set of Scala classes for describing communication protocols as types; (2) an embedded DSL for concurrent programming (reminiscent of Akka actors); (3) a Scala compiler plugin to verify whether protocols and programs enjoy desirable properties, such as deadlock-freedom; and (4) an efficient run-time system for executing Effpi programs.

The combination of (1) and (2) allows the Scala 3 compiler to check whether an Effpi program implements a desired protocol/type; and this, together with (3), means that many concurrent programming errors are found and reported at compile-time. Further, (4) allows for running highly concurrent Effpi programs with millions of interacting processes/actors, by scheduling them on a limited number of CPU cores.

In this talk, I will give an overview of Effpi, illustrate its design and main features, and explain how it leverages the capabilities of Scala 3. I will also discuss ongoing and future work.

References
1   Scalas, A., Yoshida, N., & Benussi, E. (2019). *Verifying Message-Passing Programs with Dependent Behavioural Types*. In Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation (pp. 502–516). Association for Computing Machinery. https://doi.org/10.1145/3314221.3322484
2   Scalas, A., Yoshida, N., & Benussi, E. (2019). Effpi: Verified Message-Passing Programs in Dotty. In Proceedings of the Tenth ACM SIGPLAN Symposium on Scala (pp. 27–31). Association for Computing Machinery. https://doi.org/10.1145/3337932.3338812

3.15 Algebraic Session Types

Peter Thiemann (Universität Freiburg, DE) and Vasco T. Vasconcelos (University of Lisbon, PT)

Current session type systems rely on recursive types to model recursive protocols. This approach requires nontrivial algorithms for checking type equivalence. For traditional recursive session types, type equivalence can be reduced to equivalence of finite automata. For context-free session types, it amounts to equivalence of deterministic context-free languages.

The system of algebraic session types overcomes these problems while keeping the expressivity of context-free session types. By modeling recursive protocols as an extension of recursive algebraic datatypes, we can replace a structural approach to recursive types by a nominal one. This shift in perspective avoids the expensive algorithms for type equivalence and replaces them with a linear-time test in the size of the type.

We demonstrate with examples that algebraic session types enable new ways of parameterizing protocols, which were difficult to achieve with previous systems.

3.16 Polymorphic Context-free Session Types

Peter Thiemann (Universität Freiburg, DE) and Vasco T. Vasconcelos (University of Lisbon, PT)

Context-free session types provide a typing discipline for recursive structured communication protocols on bidirectional channels. They overcome the restriction of regular session type systems to tail recursive protocols. This extension enables us to implement serialisation and deserialisation of tree structures in a fully type-safe manner. We present the theory underlying the language FreeST 2, which features context-free session types in an extension of System F with linear types and a kind system to distinguish message types and channel types. The system presents some metatheoretical challenges, which we address, contractivity in the presence of polymorphism, a non-trivial equational theory on types, and decidability of type equivalence. We also establish standard results on type preservation, progress, and a characterisation of erroneous processes.

3.17   A Joyful Empirical Study on Session Types

Nobuko Yoshida (Imperial College London, GB)

Rust is a modern systems language focused on performance and reliability. Complementing Rust's promise to provide "fearless concurrency", asynchronous message passing is widely used thanks to its efficient and intuitive communication model—although it is also vulnerable to many concurrency errors such as deadlocks. Particularly, developers frequently exploit asynchronous message reordering, where sends and receives are reordered to maximise computation-communication overlap. Unfortunately, these kinds of optimisations open up a Pandora's box of further subtle concurrency bugs.

To guarantee deadlock-freedom by construction, we present rumpsteak: a new Rust framework based on session types. Previous session type implementations in Rust are either (1) built upon synchronous and blocking communication, which incurs a substantial performance cost; and/or (2) limited to two-party interactions, which risks introducing deadlocks. Crucially, none of these implementations can support the safe message reordering we seek.

rumpsteak instead uses multiparty session types and targets asynchronous applications using async/await code. Its unique feature is the ability to practically offer asynchronous message reordering while preserving deadlock-freedom. For this, rumpsteak incorporates two recent advanced session type theories: (1) k-multiparty compatibility (k-MC), which globally verifies safety properties for a set of participants and (2) asynchronous multiparty session subtyping, which locally verifies optimisations in the context of a single participant. Specifically, we propose a novel algorithm for asynchronous subtyping that is both sound and decidable.

3.18   Monitoring Protocol Conformance with Multiparty Session Types and OpenTelemetry

Fangyi Zhou (Imperial College London, GB) and Nobuko Yoshida (Imperial College London, GB)

In this talk, we demonstrate our ongoing work of monitoring protocol conformance using multiparty session types and OpenTelemetry. OpenTelemetry is a new observability framework for distributed cloud software, and we utilise its distributed tracing functionality to validate traces against a prescribed global type.

3.19 Statically Verified Refinements for Multiparty Protocols

Fangyi Zhou (Imperial College London, GB), Francisco Ferreira, Raymond Hu (Queen Mary University of London, GB), Rumyana Neykova (Brunel University – Uxbridge, GB), and Nobuko Yoshida (Imperial College London, GB)

With distributed computing becoming ubiquitous in the modern era, safe distributed programming is an open challenge. To address this, multiparty session types (MPST) provide a typing discipline for message-passing concurrency, guaranteeing communication safety properties such as deadlock freedom.

While originally MPST focus on the communication aspects, and employ a simple typing system for communication payloads, communication protocols in the real world usually contain constraints on the payload. We introduce refined multiparty session types (RMPST), an extension of MPST, that express data dependent protocols via refinement types on the data types.

We provide an implementation of RMPST, in a toolchain called Session*, using Scribble, a multiparty protocol description toolchain, and targeting F*, a verification-oriented functional programming language. Users can describe a protocol in Scribble and implement the endpoints in F* using refinement-typed APIs generated from the protocol. The F* compiler can then statically verify the refinements. Moreover, we use a novel approach of callback-styled API generation, providing static linearity guarantees with the inversion of control. We evaluate our approach with real world examples and show that it has little overhead compared to a naive implementation, while guaranteeing safety properties from the underlying theory.

4 Working groups

4.1 Breakout Group: Typing Non-Channel-Based Models

Gul Agha (University of Illinois – Urbana-Champaign, US), Mariangiola Dezani (University of Turin, IT), Simon Fowler (University of Glasgow, GB), Philipp Haller (KTH Royal Institute of Technology – Kista, SE), Raymond Hu (Queen Mary University of London, GB), Eduard Kamburjan (University of Oslo, NO), Roland Kuhn (Actyx AG – München, DE), Hernán Melgratti (University of Buenos Aires, AR), Alceste Scalas (Technical University of Denmark – Lyngby, DK), and Peter Thiemann (Universität Freiburg, DE)

We started with the talk by Eduard Kamburjan (see materials), which describes a typing discipline and implementation for active objects based on asynchronous remote invocations. Thereafter, we briefly introduced the execution model later presented by Roland Kuhn and Emilio Tuosto in the plenary session on Thursday, namely evaluating state machines over an eventually consistent global log that is merged from locally produced append-only logs.

In such a system we have no channels to which session types can be attached. Instead, we need to characterise the event traces produced by the execution of a global protocol, where additional difficulty arises from the fact that the local behaviour can produce a greater variety of event traces due to local state machines having only a partial view of the global log – event dissemination is the mechanism by which non-determinism is introduced in this model.

The resulting "event soup" prompted the question of whether join calculus is a suitable model to tame such a system. The issue with this is that the setting is meant to guarantee 100% availability while join calculus treats messages as linear values that can only be consumed once; implementing this implies a consensus protocol between different participants that would ruin availability during network partitions. Events in the Actyx system have quantity omega, so join calculus is too strict.

The Actyx implementation currently has no mechanism for preventing conflicts (a consensus-based facility could be added in an opt-in fashion to coordinate important decisions). Instead, it allows conflicts to be detected since they are clearly visible in the event traces. This is okay for all cases where compensating actions can adequately fix the situation.

At this point we concluded that the discussed system is clearly distinct from the known body of previous work, so it is hard to transfer existing models or mechanisms to this setting.

We then switched to Actors as the other non-channel system that is widely used. Immediately after Eduard's talk we had already briefly discussed that removing the response (and thus the usage of Futures) from that Active Objects model would result in a very similar setting to the Actor model. We then looked at the join calculus again but moved on to the question of whether to type the behaviour of the processes or the contents of the mailbox.

There are two approaches: either use local types to govern the actions performed by the actor (which is done by Neykova and Yoshida's work on Multiparty Session Actors, and by Harvey et al.'s work on EnsembleS, for example); or to type the contents of the mailbox. The naive way of implementing the latter, namely typing a sequence of types to be received, is too restrictive to use in practice. Instead, work by de'Liguoro and Padovani considers a mailbox type system, where mailboxes are given a type described by a commutative regular expression (unordered sequencing, replication, and choice), which can rule out errors such as unhandled messages and deadlocks.

Gul Agha pointed out that in the absence of a typed channel it is more difficult to figure out failure cases because it is less obvious what happens when a given message does not arrive. This might require some dynamic (i.e. symbolic) analysis in addition to the static typing judgement, i.e. finding an undesirable configuration and working one's way backwards to find how this configuration could arise. Another issue is that the absence of a message cannot be clearly attributed to a single role, so the static analysis will need to model for example whether there can be at least one participant of a given role in a state that allows the needed interaction. Symbolic reasoning can then be used to figure out whether such state is actually realised.

Roland Kuhn pointed out that completeness of such a system is not required and will probably not even lead to the most useful system, because an application-dependent unhandled failure rate is acceptable in systems where humans can be called upon to fix things – which includes most systems today.

Since also such systems cannot statically prevent all conflicts, it is important to keep track of the local knowledge at the time a decision was made (like a causality trace) to be able to figure out the correct compensating actions. While this has a cost and is therefore

usually not done in real-world systems, the cost can be kept at a minimum for only tracking a single bit for each message that is expected within a given session (i.e. a version vector where each counter is 0 or 1).

One important message from Gul Agha was that we have nice means to separate

- how (which is the private implementation)
- what (which is the method, function signature, message, . . . that selects the operation)
- when (which is governed by a session type)
- who (which can be pub-sub, static, assignment from a pool, etc.)

By keeping these well separated we get higher modularity and reuse, e.g. just needing to change a session type and reusing all methods and implementations.

4.2 Breakout Group: Logic-based approaches

Marco Carbone (IT University of Copenhagen, DK), Stephanie Balzer (Carnegie Mellon University – Pittsburgh, US), Ornela Dardha (University of Glasgow, GB), Wen Kokke (University of Edinburgh, GB), Sam Lindley (University of Edinburgh, GB), Fabrizio Montesi (University of Southern Denmark – Odense, DK), J. Garrett Morris (University of Iowa – Iowa City, US), Jorge A. Pérez (University of Groningen, NL), Bernardo Toninho (NOVA School of Science and Technology – Lisbon, PT), and Philip Wadler (University of Edinburgh, GB)

In the last decade, behavioural types, in particular Session Types, have been connected to Linear Logic. Back in 2010, Caires and Pfenning [1] proposed a proposition-as-types connection between intuitionistic linear logic and a session-typed variant of the pi calculus. Later, Wadler [2] used the same idea to draw a connection to classical linear logic. The two results have laid the basis for a stream of contributions in the area aiming at tightening the connection between behavioural type systems and linear logic. Thanks to this approach, it has been possible to represent some problems in behavioural types logically, solve them, and then map them back.

The goal of this breakout-group was not only to discuss the current state of the art, but also the directions for future research on the topic. Such directions can be summarised as follows:

- The community is interested in further exploiting the proposition-as-types approach for better understanding session/behavioural types.
- Different results are present in the literature: it is time that we also try to relate them formally so that we can transfer strengths from one approach to another.
- We should not restrict to Classical/Intuitionistic Linear Logic, but explore the relationship with other logics. This discussion has spawned a further discussion on what the minimum requirements are for a system to be a logic. An explicit and clear answer to "what is a logic?" can be found in Henry De Young's thesis [3], specifically on pag. 37, section 3.2.

In order to address the points above, people present at the meeting proposed to:

- Set up an online reading group in order to discuss relationships between single approaches
- Set up a sharing platform where results can actually be shared, and new collaborations can be started.

References

1 Caires, L., Pfenning, F.: Session types as intuitionistic linear propositions. In: CONCUR. pp. 222–236 (2010)

2 Wadler, P.: Propositions as sessions. In: ICFP. pp. 273–286 (2012)

3 Henry De Young. *Session Typed Ordered Logical Specifications*. PhD Thesis, 2020. `https://www.cs.cmu.edu/~hdeyoung/assets/papers/thesis20.pdf`

4.3 Breakout Group: Type-Informed Recovery Strategies

Fabrizio Montesi (University of Southern Denmark – Odense, DK), Laura Bocchi (University of Kent – Canterbury, GB), Marco Carbone (IT University of Copenhagen, DK), Ornela Dardha (University of Glasgow, GB), Mariangiola Dezani (University of Turin, IT), Philipp Haller (KTH Royal Institute of Technology – Kista, SE), Mathias Jakobsen (University of Glasgow, GB), Sam Lindley (University of Edinburgh, GB), J. Garrett Morris (University of Iowa – Iowa City, US), Philip Munksgaard (University of Copenhagen, DK), Laura Voinea (University of Kent – Canterbury, GB), Philip Wadler (University of Edinburgh, GB), and Fangyi Zhou (Imperial College London, GB)

Managing failures is important in distributed systems, but related research on behavioural types is still in the early stages. We are particularly interested in how types could help in the programming of strategies for recovering from failures.

The problem of failure recovery is multifaceted, because there are different categories of failures that can be encountered at runtime. These include crashes, message losses, and wrong ordering of messages or actions. Furthermore, depending on the level that software operates at (recall, for example, the layers of the OSI model), assumptions and focus might change.

For example, if we wish to write a low-level protocol like Ethernet or a distributed agreement protocol, then it might be desirable to use a fine-grained model that (a) exposes failures in detail, like single failures in the communication of each network packet, "alive" timeouts, etc., and (b) allows for programming recovery strategies to handle such failures, e.g., retransmission.

Differently, if we are reasoning about code designed for the application level, it is typical to make stronger reliability assumptions. For example, we might assume that network transmissions are reliable, delegating to the lower levels to deal with relevant failures there. A failure raised from the lower levels would then be managed using more abstract constructs on the application level, leading to more coarse recovery strategies (e.g., restart of the entire protocol or reconnection).

The multifaceted and multilevel nature of failure recovery is reflected by past and current research, including: types for managing fallible interactions and message loss in choreographic languages [3]; work on ensuring that data to be communicated can be meaningfully marshalled [4]; and exceptions for session types, where a reliable network is assumed [2, 1].

There are several open questions related to the principles of recovery strategies, including:

- How can we model the principles (for failure recovery) used in real-world software on different levels? How should we then interface lower-level models (with weak reliability assumptions) with higher-level models (with stronger reliability assumptions)?

- Can we modularly encode recovery strategies in high-level languages into lower-level languages that make weaker reliability assumptions?
- Can we extend session types to reason usefully about failures of different kinds?
- Can we make a fundamental calculus of failures? Is there a useful link to ongoing research on effect handlers (e.g., [5])?

References
1   Paul Harvey and Simon Fowler and Ornela Dardha and Simon J. Gay. *Multiparty Session Types for Safe Runtime Adaptation in an Actor Language*. In 35th European Conference on Object-Oriented Programming, ECOOP 2021, LIPIcs Vol. 194, 10:1–10:30, 2021.

2   Marco Carbone and Kohei Honda and Nobuko Yoshida. *Structured Interactional Exceptions in Session Types*. In CONCUR 2008 – Concurrency Theory, 19th International Conference, Lecture Notes in Computer Science, Vol. 5201, pp. 402–417, Springer, 2008.

3   Fabrizio Montesi and Marco Peressotti. *Choreographies meet Communication Failures*. CoRR abs/1712.05465, 2018.

4   Heather Miller and Philipp Haller and Martin Odersky. *Spores: A Type-Based Foundation for Closures in the Age of Concurrency and Distribution*. In 28th European Conference on Object-Oriented Programming, ECOOP 2014, Lecture Notes in Computer Science, Vol. 8586, pp. 308–333, Springer, 2014.

5   Sam Lindley. *Handler calculus*. Draft, 2021. `https://homepages.inf.ed.ac.uk/slindley/papers/handler-calculus-draft-may2021.pdf`

4.4   Breakout Group: Session types with untrusted counter-parties

Philip Munksgaard (University of Copenhagen, DK), Christian Bartolo Burló (Gran Sasso Science Institute, IT), Marco Carbone (IT University of Copenhagen, DK), Mariangiola Dezani (University of Turin, IT), Simon Fowler (University of Glasgow, GB), Mathias Jakobsen (University of Glasgow, GB), Roland Kuhn (Actyx AG – München, DE), Fabrizio Montesi (University of Southern Denmark – Odense, DK), Alceste Scalas (Technical University of Denmark – Lyngby, DK), Peter Thiemann (Universität Freiburg, DE), Emilio Tuosto (Gran Sasso Science Institute, IT), and Fangyi Zhou (Imperial College London, GB)

There is no question that session types and behavioral types are useful in situations where you control all the nodes in the network. In practice, however, that is not always the case, eg. in peer-to-peer networks or on the internet in general. Still, it would be helpful to be able to use session types to model the communication protocols in such cases.

Attempting to do so presents a range of problems, including:

- How do we know what protocol the counterparty is using?
- Can we trust that the counterparty will adhere to that protocol?
- How should we handle failures, both in the protocol and in the transport layer?

Starting from the end, exhaustive work has been performed in the area of *monitors*. Given a session type, a monitor sits between the user and an untrusted counterparty, mediating and making sure that the user only sees messages that adhere to the specified protocol. Depending on the particular implementation, it can handle timeouts, protocol errors, dropped connections and so on. With such a monitor, we can safely implement our side of the session type, without having to think about these problems.

However, what about a well-meaning counterparty whose protocol differs slightly from ours? A simple example would be a login server that accepts two strings, a username and a password (in that order), while your client expects to send a password and a username. Without additional information from the server, you would not be able to correctly interface with the server, even with a monitor.

For such a case to work, we would need to have some sort of meta-protocol for talking about the session type. In essence, each party should be able to tell the other what session type it expects to be using, and we should have a way of handling discrepancies in the protocol. This line of reasoning lead to some interesting debates at the seminar, and a stated intention from several attendees to continue work after the seminar.

4.5 Breakout Group: Join Patterns / Synchronization – The Next Generation

Claudio Russo (Dfinity – Cambridge, GB), Gul Agha (University of Illinois – Urbana-Champaign, US), Philipp Haller (KTH Royal Institute of Technology – Kista, SE), Eduard Kamburjan (University of Oslo, NO), Emilio Tuosto (Gran Sasso Science Institute, IT), Laura Voinea (University of Kent – Canterbury, GB), and Philip Wadler (University of Edinburgh, GB)

Intro

The attendees were a mixture of people with considerable experience of join patterns and the join calculus and those curious to hear learn more about them.

We kicked off with an introduction to join patterns as:

- a mechanism for pattern matching over the empty/non-empty state of co-located, (asynchronous) message queues.
- a special case of more general Logic Programming constructs with arbitrary predicates on channel contents from Concurrent Prolog.

Some stressed that the main advantage of joins is expressing synchronization & coordination, not necessarily pattern matching over the content of queues enabled by Concurrent Prolog.

We then re-iterated some of the history of join patterns – which have come in several guises from language extensions to libraries, for niche and mainstream languages. A (by no-means exhaustive) list includes the Join Calculus, JoCaml, Funnel (the precursor to Scala), Polyphonic C#, JErlang, JoinJava, Comega, the C# Joins library, Concurrent Basic, Scalable Joins, Scala, Akka, and Sharpie.

The different presentations of join patterns typical adopt one of :

- recursive function declarations with alternatives of conjoined function headers (JoCaml) and selective "return"s to different functions within the same body.
- object methods with alternatives of conjoined method headers (Polyphonic C#) and single returns.
- separate channels declarations with patterns over them (Joins library, Concurrent Basic)
- designs disguised as (library) extensions of pattern matching (Scala)
- various degrees of support for synchronous channels in patterns (on, sometimes several).

Limits to Adoption

Despite the plethora of implementations and designs, join patterns have never made into the mainstream the way, say, lambda abstraction and garbage collection have. Eisenbach has expressed similar difficulties [5]. We came up with following potential reasons:

- The early presentation of joins were all asynchronous, sometimes continuation based, and thus less approachable.
- Alternatives such as actors and futures are easier to explain, though likely less expressive (e.g. n-ary synchronization is difficult to achieve with usual primitives).
- Features such as lambdas and GC took several decades to gain wide-spread adoption. Perhaps join patterns just need more time.

Typing

We then briefly turned to discussing the typing aspects of join patterns. Phil Wadler asked "Is there a Curry-Howard isomorphism" for the join calculus as there is for linear types and process-calculi, citing Frank Pfenning's work on linear logic for typing processes. Since the inspiration for join patterns comes from chemistry (a restriction of the chemical abstract machine), not logic, uncovering a Curry-Howard isomorphism might prove challenging. Nevertheless, the atomic consumption of linear resources inherent to joins suggests a strong connection to linear typing: [1] applies separation logic to verify some join based programs, while the more recent work [2, 3, 4] on linear typing of joins might shed further light.

Expressivity

Discussion briefly turned to the question of expressivity: how do join patterns compare to say, pi-calculus or actors? Cedric Fournet's thesis gave a translation between pi and joins but we could not recall if it was modular or whole-program. Regardless of expressivity, there is strong reason to believe that join-calculus is easier to implement than, say, pi-calculus, especially in a distributed setting. The reason for this is that the scheduling decisions needed to be made in joins can be resolved locally, at a receiver, while typical implementations of the pi-calculus rely on reaching agreement between distributed parties and are thus more obviously suited to shared-memory, non-distributed implementations.

References

1 Kasper Svendsen, Modular specification and verification for higher-order languages with state, PhD thesis, `https://cs.au.dk/~birke/phd-students/SvendsenS-thesis.pdf`.
2 Silvia Crafa, Luca Padovani, The Chemical Approach to Typestate-Oriented Programming, ACM Transactions on Programming Languages and Systems, vol. 39(3), pages 13:1-13:45, 2017.
3 Luca Padovani, A Type Checking Algorithm for Concurrent Object Protocols, Journal of Logical and Algebraic Methods in Programming, vol. 100, pages 16-35, 2018.
4 Luca Padovani, Deadlock-Free Typestate-Oriented Programming, Programming Journal, vol. 2(3), pages article 15, 2018.
5 Susan Eisenbach, personnel communication with Russo, Microsoft Research, 2018.

Participants

- Marco Carbone
  IT University of
  Copenhagen, DK
- Simon Fowler
  University of Glasgow, GB
- Philipp Haller
  KTH Royal Institute of
  Technology – Kista, SE
- Mathias Jakobsen
  University of Glasgow, GB

- Eduard Kamburjan
  University of Oslo, NO
- Roland Kuhn
  Actyx AG – München, DE
- Sam Lindley
  University of Edinburgh, GB
- Fabrizio Montesi
  University of Southern Denmark –
  Odense, DK

- Philip Munksgaard
  University of Copenhagen, DK
- Alceste Scalas
  Technical University of Denmark
  – Lyngby, DK
- Peter Thiemann
  Universität Freiburg, DE
- Emilio Tuosto
  Gran Sasso Science Institute, IT

Remote Participants

- Gul Agha
  University of Illinois –
  Urbana-Champaign, US
- Stephanie Balzer
  Carnegie Mellon University –
  Pittsburgh, US
- Christian Bartolo Burló
  Gran Sasso Science Institute, IT
- Laura Bocchi
  University of Kent –
  Canterbury, GB
- Edwin Brady
  University of St Andrews, GB
- Ilaria Castellani
  INRIA – Sophia Antipolis, FR

- Tzu-Chun Chen
  Evonik Industries – Hanau, DE
- Ornela Dardha
  University of Glasgow, GB
- Mariangiola Dezani
  University of Turin, IT
- Simon Gay
  University of Glasgow, GB
- Raymond Hu
  Queen Mary University of
  London, GB
- Atsushi Igarashi
  Kyoto University, JP

- Jules Jacobs
  Radboud University
  Nijmegen, NL
- Wen Kokke
  University of Edinburgh, GB
- Dimitrios Kouzapas
  University of Cyprus –
  Nicosia, CY
- Hernán Melgratti
  University of Buenos Aires, AR
- J. Garrett Morris
  University of Iowa –
  Iowa City, US

- Rumyana Neykova
Brunel University –
Uxbridge, GB

- Marco Peressotti
University of Southern Denmark –
Odense, DK

- Jorge A. Pérez
University of Groningen, NL

- Claudio Russo
Dfinity – Cambridge, GB

- Guido Salvaneschi
Universität St. Gallen, CH

- Bernardo Toninho
NOVA School of Science and
Technology – Lisbon, PT

- Vasco T. Vasconcelos
University of Lisbon, PT

- Laura Voinea
University of Kent –
Canterbury, GB

- Philip Wadler
University of Edinburgh, GB

- Nobuko Yoshida
Imperial College London, GB

- Fangyi Zhou
Imperial College London, GB
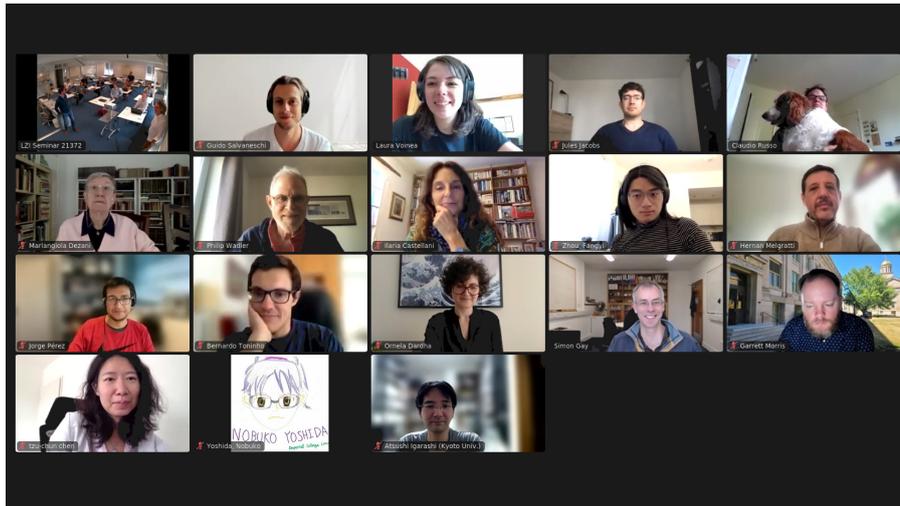
Conversational Agent as Trustworthy Autonomous System (Trust-CA)

Edited by

Effie Lai-Chong Law[1], Asbjørn Følstad[2], Jonathan Grudin[3], and Björn Schuller[4]

1   **Durham University, GB,** `lai-chong.law@durham.ac.uk`
2   **SINTEF – Oslo, NO,** `asbjorn.folstad@sintef.no`
3   **Microsoft – Redmond, US,** `jgrudin@microsoft.com`
4   **Universität Augsburg, DE,** `schuller@informatik.uni-augsburg.de`

──── **Abstract** ────

This report documents the program and the outcomes of Dagstuhl Seminar 21381 "Conversational Agent as Trustworthy Autonomous System (Trust-CA)". First, we present the abstracts of the talks delivered by the Seminar's attendees. Then we report on the origin and process of our six breakout (working) groups. For each group, we describe its contributors, goals and key questions, key insights, and future research. The themes of the groups were derived from a pre-Seminar survey, which also led to a list of suggested readings for the topic of trust in conversational agents. The list is included in this report for references.

1   Executive Summary

Effie Lai-Chong Law (Durham University, GB)
Asbjørn Følstad (SINTEF – Oslo, NO)
Jonathan Grudin (Microsoft – Redmond, US)
Björn Schuller (Universität Augsburg, DE)

The overall goal of the Dagstuhl Seminar 21381 "Conversational Agent as Trustworthy Autonomous System" (Trust-CA) was to bring together researchers and practitioners, who are currently engaged in diverse communities related to Conversational Agents (CA), to explore challenges in maximising the trustworthiness of and trust in conversational agents as AI-driven autonomous systems – an issue deemed increasingly significant given their widespread uses in every sector of life – and to chart a roadmap for the future conversational agent research. The three main challenges we identified were:

- How do we develop trustworthy conversational agents?
- How do we build people's trust in them?
- How do we optimise human and conversational agent collaboration?

The Seminar Trust-CA took place on 19-24 September 2021 in a hybrid mode. Out of 50 invitees, 19 attended in person and the rest joined online from all over the world, including Brazil, Canada, France, Germany, Greece, Ireland, Netherlands, Norway, Poland, South Korea, Sweden, Switzerland, UK and USA.

The four-day scientific programme started by unpacking the notion of "trust in conversational agent" with a panel discussion. Each of the four seminar organisers expressed their views on the notion. Jonathan Grudin presented a list of ten species of trust that can be applied to conversational agents, for instance, "Trust that a CA will correctly interpret my question or request; will deliver relevant, reliable, useful information." Asbjørn Følstad first presented an overview of the six themes derived from a pre-Seminar survey (details are in Overview of Working Groups) and then described his recent work on the effect of human likeness of a conversational agent on trust. Björn Schuller presented factors influencing trust in humans, such as being reliable, ethical, moral and charismatic, and in conversational agents, such as being explainable, interpretable and transparent. He also discussed how to measure trust reliably and the danger of overtrust. Effie Law discussed the notion of trust with reference to multidisciplinary theory of trust (e.g. psychological, social, historical), beyond the use of questionnaires to evaluate trust, and identifying applications where agents are of high practical value. Some attendees commented on the ideas shared, e.g., the elusiveness of trust.

The scientific programme comprised two major parts – Talks and Breakout Groups. There were altogether 20 talks, covering a range of topics (see Abstracts). Nine of the talks were delivered in person and the rest online. There were six Breakout Groups with each discussing one of the six themes: Group 1 – Scope of Trust in CA; Group 2 – Impact of CA; Group 3 – Ethics of CA; Group 4 – AI and Technical Development; Group 5 – Definition, Conceptualisation and Measurement of Trust; Group 6 – Interaction Design of CA. Group 1, 3 and 4 had one team each whereas Group 2, 5 and 6 had two teams each. To ease collaboration, individual teams were either in-person or online (except for Group 4 which was in hybrid mode). Each group had three two-hour working sessions . In the evening, each group reported progress and invited feedback for shaping subsequent sessions.

The group discussions led to intriguing insights that contributed to addressing the main challenges listed above and stimulated future collaborations (see the Workgroup Reports). Here we highlight one key insight of each group. Group 1 developed a dynamic model of trust with three stages, Build-Maintain-Repair, which evolve over time. Group 2 drafted a code of ethics for trustworthy conversational agents with eight provisions. Group 3 explored the ethics challenge of transparency from the perspective of conversational disclosure. Group 4 called for increased collaboration across research communities and industries to strengthen the technological basis for trust in conversational agents. Group 5 proposed a framework for integrating measurement of trusting beliefs and trusting behaviour. Group 6 analysed several aspects of multimodality to understand their possible effects on trust in conversational agents. Apart from the scientific programme, the Seminar organised several social events, including after-dinner wine and cheese gatherings, hiking in a nearby historic site, and a music event.

Overall, our Dagstuhl Seminar Trust-CA was considered a success. The major outputs were derived from the pre-Seminar survey (six research themes and a recommended reading list), twenty talks, and six multi-session breakout groups. Thanks must go to the enthusiastic involvement of all attendees in analysing various aspects of the burgeoning topic of conversational agents. Of course, the Seminar could only take place with the generosity of *Schloss Dagstuhl – Leibniz Center for Informatics*. The efficiency and friendliness of the scientific and administrative staff of Schloss Dagstuhl was much appreciated by the organisers and all attendees.

 Table of Contents

Working groups

Open problems

3.1   Chatbots and Voice Assistants from the Perspective of Machine Ethics and Social Robotics

Oliver Bendel (FH Nordwestschweiz – Windisch, CH)

As a discipline, machine ethics examines the possibilities and limits of moral and immoral machines. Social robotics researches and builds robots that interact with, communicate with, are close to, and map features of humans and animals. In doing so, they have a specific use, such as care, support, or entertainment. In his talk, Oliver Bendel outlined the fundamentals of machine ethics and social robotics and presented conversational agents and complementary systems that have emerged from these disciplines. The GOODBOT (2013) is a chatbot that responds morally adequately to problems of the user. The LIEBOT (2016), also a chatbot, can lie systematically, using seven different strategies. The BESTBOT (2018) is a chatbot that recognizes certain problems and conditions of the user with the help of text analysis and facial recognition and reacts morally to them. In 2019, Oliver Bendel and his team developed the MOME (the name stands for "morality menu"). With the help of sliders, you can transfer your moral beliefs to the chatbot MOBO, which then formulates and responds accordingly. The most recent project to date was SPACE THEA (2021), a voice assistant that demonstrates empathy and is designed to accompany astronauts to Mars. Most of the artifacts earn our trust by recognizing our situation and helping and supporting us. The LIEBOT, on the other hand, systematically lies to us and makes us aware that conversational agents can be designed in an abusive or negative way.

3.2   Interaction with multi-bots

Heloisa Candello (IBM Research – Sao Paulo, BR)

User Evaluation of Multi-party Conversational Systems. Recent advances in artificial intelligence, natural language processing, and mobile computing, together with the rising popularity of chat and messaging environments, have enabled a boom in the deployment of interactive systems based on conversation and dialogue. This talk explores the design and evaluation of conversational interfaces, and it is focused on design and evaluation methods that address specific challenges of interfaces based on multi-party dialogue. I will show two projects. First, Café com os Santiagos is an artwork where visitors conversed with three chatbots portraying characters from a book in a scenographic space recreating a 19th-century coffee table. It was accessed by more than 10.000 users in a public space, resulting in insights to improve the conversation system even more. Second, I will show an experiment with Finch's cognitive investment adviser. Finch interface made a state-of-art artificial conversational governance system accessible for regular users to assist in financial decisions.

3.3 Why should we care about linguistic register? Insights on chatbot language design

Ana Paula Chaves (Federal University of Technology – Paraná, BR)

This talk discusses the relevance of linguistic register as a theoretical framework for chatbot language design. I presented the concept of register and discussed how using register-specific language influences the user's perceptions of their interactions with chatbots. To demonstrate that, I presented a study performed in the context of tourism information search chatbots, where participants evaluated the language appropriateness, credibility, and user experience when facing chatbot utterances in different registers. I argued that the appropriate use of language is relevant to design trustworthy chatbots since it influenced the chatbot's credibility in our study. I also pointed to future research directions

3.4 Towards Personalized Explainable AI

Cristina Conati (University of British Columbia – Vancouver, CA)

The AI community is increasingly interested in understanding how to build artifacts that are accepted and trusted by their users in addition to performing useful tasks. It is undeniable that explainability can be an important factor for acceptance and trust. However, there is still limited understanding of the actual relationship between explainability, acceptance, and trust and which factors might impact this relationship. In this talk, I argue that one such factor relates to the user's individual differences in terms of both long-term, stable traits (e.g., expertise, cognitive abilities, preferences) and short-term transient states (e.g., level of cognitive load, affective state). Namely, given a specific AI application, different types and forms of explanations may work best for different users, and even for the same user at different times, depending to some extent on both their long-term traits and short-term states. As such, our long-term goal is to develop personalized XAI tools that adapt dynamically to the user's needs by taking relevant user factors into account. In this talk, I focus on research investigating the impact of long-term traits, and how they may drive XAI personalization. I present a general methodology to address these questions, followed by an example of how it was applied to ascertain which long-term traits are relevant for personalizing explanations in an intelligent tutoring system (ITS). I discuss how to move forward from these insights, and present research paths that should be explored to make personalized XAI happen.

3.5 In human-likeness we trust? The implications of human-like design on partner models and user behaviour

Benjamin Cowan (University College – Dublin, IE)

In human-likeness we trust? The implications of human-like design on partner models and user behaviour" Abstract: Voice has now become a mainstream interaction modality. Current voice interfaces fundamentally rely on human conversation as an interaction metaphor, using human-like design to support partner model building. My talk will explore how human-like VUI design shapes our beliefs of a machine partner's abilities, how this is potentially crucial to consider in terms of trust in voice interface interaction, and whether this interaction metaphor is actually appropriate as we strive for more trustworthy conversational systems.

3.6 Underestimated Challenges in Developing and Using Conversational Agents

Jonathan Grudin (Microsoft – Redmond, US)

For decades, conversational agents were developed in small, trusting, homogeneous laboratories. Since 1995, commercial internet activity and the web has seen the rise of "bad actors" and a range of grey activity, creating challenges that need to be anticipated by university researchers and conversational agent developers who still work in small, trusting groups where consideration of potential technology misuse is low. The 'virtual companion' artificial general intelligence, reflected in ELIZA and Turing Test contestants, remains a science fiction mainstay but is approaching real-life extinction. The take-down of Tay by trolls led to sophisticated risk-mitigation approaches, but it is an expensive arms race. Amnesic conversational partners are unappealing but privacy considerations inhibit the retention of personal communication. Hugging Face, Zo, Le Luda, Replika, and others disappeared or failed to gain traction. At the brief-conversation extreme, intelligent assistants such as Alexa have encountered concerns about re-enforcing submissive female stereotypes and shaping children to converse with impersonal imperatives. Most work on conversational agents lies between these two. Task-focused chatbot technology can be employed by a range of people with a range of intentions. Facilitating seamless human-in-the-loop can be necessary and wonderful, but concealing that humans are in the loop can be ruinous. Major uses of chatbots include reducing human conversation or more effectively steering behavior. This can yield positive outcomes or negative outcomes. Let's aim high, but periodically consider unintended consequences should our work be misused.

3.7 Designing Conversational Agents for Dyadic and Group Interactions

Soomin Kim (Seoul National University, KR)

The advancements in technology shift the paradigm of how individuals communicate and collaborate. Machines play an active role in human communication. However, we still lack a generalized understanding of how exactly to design effective machine-driven communication and discussion systems. In this paper, I present new interactive systems in the form of a conversational agent, or a chatbot, that facilitate dyadic and group interactions. Specifically, I focus on: 1) a conversational agent to engage users in dyadic communication, 2) a chatbot called GroupfeedBot that facilitates daily social group discussion, 3) a chatbot called DebateBot that enables deliberative discussion. The findings of this thesis are as follows. For a dyadic interaction, participants interacting with a chatbot system were more engaged as compared to those with a static web system. However, the conversational agent leads to better user engagement only when the messages apply a friendly, human-like conversational style. These results imply that the chatbot interface itself is not quite sufficient for the purpose of conveying conversational interactivity. Messages should also be carefully designed to convey such. Unlike dyadic interactions, which focus on message characteristics, other elements of the interaction should be considered when designing agents for group communication. In terms of messages, it is important to synthesize and organize the information given that countless messages are exchanged simultaneously. In terms of relationship dynamics, rather than developing a rapport with a single user, it is essential to understand and facilitate the dynamics of the group as a whole

3.8 Measuring Understanding in Interactions with Embodied Conversational Interfaces: Theory, Studies and Computation

Dimosthenis Kontogiorgos (KTH Royal Institute of Technology – Stockholm, SE)

Research in face-to-face human-robot interaction has focused on developing teaching robots that have little abilities to adapt to users' signals of understanding. Human speakers seem to establish common ground incrementally, the mutual belief of understanding among the conversational partners. When teaching each other new tasks, speakers tend to package pieces of information in small fragments and provide information to the learners incrementally. In this talk, I will present our work investigating how speakers' incremental construction of utterances affect the cognitive resources of the conversational partners during utterance production and comprehension. I will also discuss implications for future empirical research on the design of task-oriented human-robot interactions, and how assistive social robots may benefit from the production of fragmented instructions. Using data from a recent online perception study, I will finally present empirical findings from recent research on how we used mouse movement analysis to detect user uncertainty when guided by a conversational interface.

3.9 Establishing long-term relationships with conversational agents – lessons from prolonged interactions with social robots

Guy Laban (University of Glasgow, GB)

Social robots' cognitive architectures and embodied cognition can elicit socially meaningful behaviours and emotions from humans. These robots can afford valuable opportunities for social engagement with human users, and there is a growing evidence base that documents how social robots might function as autonomous tools to support psychosocial health interventions via establishing meaningful relationships. Since interactions with social robots are novel and exciting for many people, one particular concern is the extent to which people's behavioural and emotional engagement with robots might develop from initial interactions with a robot, when a robot's novelty is especially salient, and be sustained over time. Here we aimed to test the extent to which social robots can elicit emotional expression and disclosures from people, as well as affect their perceptions in a long and intensive period. Through the use of a mediated online experiment, this research was designed to examine the type and extent of expressions people use to communicate with a social robot, how they perceive it, as well as how people disclose information and emotions to a social robot via online video chats across time. Across a period of five weeks, 39 participants engaged in interactions with the social robot Pepper (SoftBank Robotics) via Zoom video chats twice a week. Participants were asked by Pepper about their general everyday experiences in one condition, whereas in the second condition these topics were framed to the COVID-19 pandemic. Our results suggest that people gradually perceived the robot to demonstrate higher degrees of agency and experience across sessions, as well as being friendlier. Moreover, participants perceived the interaction quality and the robot communication competence to be better across sessions. Finally, participants reported positive mood changes due to their interactions with Pepper across all sessions.

3.10 A Cryptocurrency Chatbot and the Social-technical Gap of Trust

Minha Lee (TU Eindhoven, NL)

Cryptocurrencies are proliferating as instantiations of blockchain, which is a transparent, distributed ledger technology for validating transactions. Blockchain is thus said to embed trust in its technical design. Yet, blockchain's technical promise of trust is not fulfilled when applied to the cryptocurrency ecosystem due to many social challenges stakeholders experience. By investigating a cryptocurrency chatbot (Brokerbot) that distributed information on cryptocurrency news and investments, we explored social tensions of trust between stakeholders, namely the bot's developers, users, and the bot itself. We found that trust in Brokerbot and in the cryptocurrency ecosystem are two conjoined, but separate challenges that users and developers approached in different ways. We discuss the challenging, dual-role of a Brokerbot as an object of trust as a chatbot while simultaneously being a mediator of

trust in cryptocurrency, which exposes the social-technical gap of trust. Lastly, we elaborate on trust as a negotiated social process that people shape and are shaped by through emerging ecologies of interlinked technologies like blockchain and conversational interfaces

3.11 Codex as a personal assistant?

Clayton Lewis (University of Colorado – Boulder, US)

The development of language-model based artificial intelligence, as seen in Codex and GPT-3, may open the way to new forms of artificial personal assistants. At present, a trial of Codex produces interesting results for a use case involving keeping track of gifts. The generated code does not work, and would require programming knowledge to repair. But the results suggestion that these models may offer new ways to think about the challenges of cognitive science, including the challenges to cognitive theorizing articulated by Harold Garfinkel. These ways of thinking may also contribute to understanding the mechanisms of trust in interactions with artificial agents.

3.12 Designing Inclusive Conversational Agents that Older Adults Can Trust

Cosmin Munteanu (University of Toronto Mississauga, CA)

Older adults (65+) are at increasing risk of being "digitally marginalized" or "digitally isolated". This is often the result of active or passive, conscious or unconscious bias in how older adults are overlooked in the design of new digital technologies. When design is more actively focused on older adults, this is often reduced to clichés about limited cognitive or physical abilities. The consequences of such approaches are significant, with many seniors having difficulties in transitioning their use of essential services to the online space in several key areas: taking financial decisions, understanding health information or accessing health services, staying connected to families, or simply doing online shopping. This is, paradoxically, exacerbated by the increased use of interfaces that are marketed as "natural", such as voice and conversational agents (chatbots). In this talk I focus on one of the most overlooked barriers toward older adults' trusting of such interfaces: mental models. I am arguing for new methodological approaches that empower older users and put them in the lead for designing novel interactive technologies that assist with reducing their digital marginalization and isolation, and through this, better reflect older adults' mental models of interacting with and trusting of such new technologies.

3.13 An introduction of the UKRI Trustworthy Autonomous Systems Node on Trust

Birthe Nesset (Heriot-Watt University – Edinburgh, GB)

Robots are rapidly gaining acceptance in recent times, where the general public, industry and researchers are starting to understand the utility of robots, for example for delivery to homes or in hospitals. However, it is key to understand how to instil the appropriate amount of trust in the user. One aspect of a trustworthy system is its ability to explain actions and be transparent, especially in the face of potentially serious errors. Here, we study the various aspects of transparency of interaction and its effect in a scenario where a robot is performing triage when a suspected Covid-19 patient arrives at a hospital. Our findings consolidate prior work showing a main effect of robot errors on trust, but also showing that this is dependent on the level of transparency. Furthermore, our findings indicate that high interaction transparency leads to participants making better informed decisions on their health based on their interaction. Such findings on transparency could inform interaction design and thus lead to greater adoption of robots in key areas, such as health and well-being.

3.14 Impact of adaptation mechanisms on user's perception of agent

Catherine Pelachaud (Sorbonne University – Paris, FR)

During an interaction, interlocutors may adapt their behaviors at different levels. We have developed different adaptation mechanisms for a virtual agent interacting with human users. These mechanisms act at the levels of conversational strategy and multimodal behaviors or at the signals level. The former two mechanisms are modeled using reinforcement learning technique. The agent learned the best strategy or multimodal behaviors to display on the fly while interacting with a user. The third model is learned from data of human-human interaction and is modeled using LSTM. These three mechanisms have been integrated within an architecture for human-agent interaction. Lately we have worked on integrating a modality that has not received much attention in human-agent interaction, namely touch. Social touch conveys several functions such as showing emotion, getting the attention, comforting someone. We have developed a decision model, based on the emotional model FAtiMA to endow the agent with the capacity to determine when to touch the user and with which touch. Finally, a third argument I have presented regards our work on simulating laughter in virtual agent, and in particular, how laughter can have an impact on users perception of the agent and on the quality of the interaction.

3.15 To trust or not to trust? What is the use case? Insights from applied research in conversational interaction

Stefan Schaffer (DFKI – Berlin, DE)

Applied research at DFKI provides insights into various application examples of conversational interaction. Three demonstrators for conversational assistant system projects for the use cases of care management, railway security and museum guides are presented and references to the topic of trust are made. The main requirement in the care management use case was to improve the inclusion of visually impaired caregivers. Based on user research, a demonstrator of a conversational ERP tool for care management was developed. A demonstration of the system showed that correctness, which is affected by automatic speech recognition errors, for example, is a key factor in ensuring trust. The security service use case focusses on the issue that soccer fans in rail travel often cause security relevant situations. Based on participatory design, a conversational assistant for efficient input of security relevant information was developed. A focus group revealed that generation of reliable information is crucial for this use case. Regarding trust, the recommendation was derived that appropriate system feedback should be generated for safety-relevant information. In the museum use case, several NLP mechanisms, including the transformer-based model BERT, are implemented to answer fact and open questions. To make the necessary annotation effort manageable, different amounts of training data are annotated for different objects. The assumption is stated that the total amount of meta data enrichment will influence the level of trustworthiness of the system.

3.16 The value of small talk and responsiveness

Ryan Schuetzler (Brigham Young University – Provo, US)

Small changes in the way a CA interacts can influence both behavior and attitudes. Tailoring chatbot messages to reflect active listening demonstrates to users that the bot can understand them, which can improve feelings of social presence and engagement. In an interview chatbot, we manipulated tailoring in the small talk rapport-building phase of the interview to understand the effect it has on perceptions and self-disclosure. In most circumstances, increased social presence and engagement is a good thing. However, we have shown that in discussions of sensitive information, less social presence might be preferable to improve user disclosure. Because users disclosing sensitive information want to feel that they are not being judged, tailoring reduces disclosure. While tailoring and small talk are small manipulations, they can create significant effects in how users perceive and respond to a chatbot.

3.17   How and When should chatbot self-disclose?

Zhou Yu (Columbia University – New York, US)

Social chatbots research has attached much attention lately. However, how and to what extent people respond to chatbot self-disclosure and how self-disclosure can impact task success remain less known. We designed a social chatbot that can perform three different types of self-disclosure: sharing factual information, cognitive opinions, and emotions. The chatbot can conduct small talks and provide relevant recommendations on two topics, movies and COVID-19 best practices. Through a large-scale user study, we found that chatbots' level of self-disclosure correlates with better conversational engagement and warmth towards the chatbot. Chatbots that perform all three types of disclosure also complete the recommendation task more effectively than ones that only perform one or two types of disclosure.

3.18   Democratizing Conversational AI: Challenges and Opportunities of No-Code, Reusable AI

Michelle X. Zhou (Juji Inc. – Saratoga, US)

Creating quality conversational AI agents not only requires deep AI expertise and sophisticated software engineering skills, but also requires large amounts of training data and intensive computational resources. Few organizations have such expertise, let alone the required resources to develop and manage their own version of conversational AI agents. To democratize conversational AI and bridge the potential AI divide, we have been developing an end-to-end, no-code AI platform that enables non-IT professionals to create, deploy, and manage their custom conversational AI agents with no code, and no IT resources required. Such a conversational AI platform has three key characteristics. First, it supports the end-to-end, no-code development of conversational AI agents with cognitive intelligence–AI agents with human soft skills, such as active listening skills and reading between the lines. These human soft skills enable AI agents to interact with their users and complete their tasks responsibly and empathetically. Second, it supports multi-level reuses of pre-built AI components, which then enables rapid customization of a conversational AI agent with no code. Third, it enables real-time conversational AI monitoring and live updates/improvements without interrupting ongoing critical conversations. Our platform has been used by non-IT professionals from multiple domains to create and manage their own conversational AI agents, demonstrating the practical values of no-code, reusable AI.

4 Working groups

4.1 Overview of Working Groups: Origin and Process

All Groups

4.1.1 Origin

Prior to the launch of the Seminar, a web-based survey was conducted to gather the attendees' views on the following issues:

1. What are the main challenges to be addressed for the topic of trust in CA?
2. Which papers (max. three) to be recommended as key background reading for the topic of trust in CA?
3. What topic to be proposed for a PhD student research project related to trust in CA?

Responses from 39 attendees were obtained. Thematic analysis of the data for *Item 1* resulted in six themes as shown in the concept map below. Each of these themes was discussed in a breakout group during the Seminar (Section 4.2 – 4.7). For *Item 2*, a list of references was compiled (Section: Open Problems). For *Item 3*, only a subset of the respondents provided input, results are not presented here.
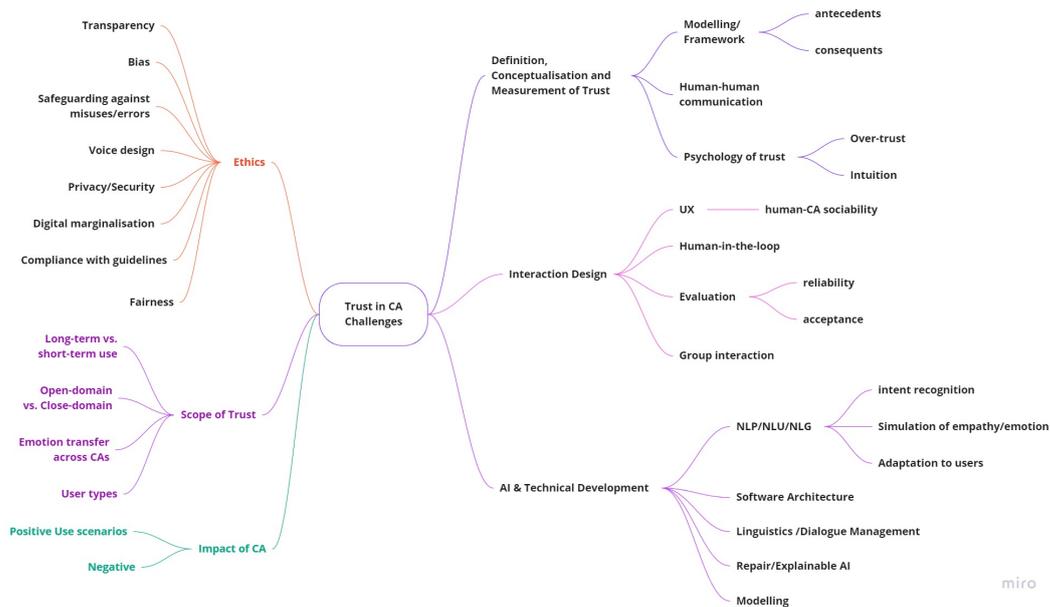
Figure 1 Six themes of Trust in CA.

4.1.2 Process

The aim of conducting breakout groups was to advance state of the art theories, methodologies and practices on trust in CA. The group discussion, which was guided by some key questions, drew on the experiences and expertise of individual members. Each of the breakout groups

had three sessions with each lasting about two hours. Key insights from the discussion were reported back in a plenary meeting in the evening to invite feedback for shaping the direction of the groupwork in the following day.

The 50 attendees were allocated to different breakout groups based on their preferences. An attendee joined one group in the morning and another group in the afternoon. Different group memberships as such encouraged stimulation and collaboration. Outputs of each breakout group are summarised in the following subsections.

4.2   Breakout Group 1: Scope of Trust in CA

Group1

Contributors: Oliver Bendel, Birthe Nesset, Catherine Pelachaud, Guy Laban, Eren Yildiz, Effie Law

4.2.1   Goal and key questions

Goal: To further explore the theoretical and practical basis for trust in CAs.
Key questions:
- How is trust in CAs established and maintained?
- Which are the relevant factors?

Relevant aspects: Usage duration (long-term vs. short term use); Domain specificity (open domain vs. closed domain); Transferability of experience across different CAs; User types

4.2.2   Key insights

A dynamic stage-based trust model with the temporal aspect is the key insight gained from the discussions of Group 1. Specifically, there are three main stages of trust evolvement, namely Build, Maintain and Repair. Depending on the severity of the consequence of broken trust between a CA and its user, interaction strategies deployed in a preceding stage may be invoked.

The Build stage.   A chatbot starts an interaction by building trust between the user and itself. It is crucial that the chatbot knows whether trust is established, and that the user is aware of the scope of the chatbot. To build trust, the following factors should be taken into consideration:
- showing affordances such as abilities, core features, limitations, and purposes
- setting up the right expectations and cost
- personalization and customization
- giving a sense of control to the user
- accommodating the user's mistakes
- recognizing the user's goals, needs and preferences and taking them seriously

When an optimal level of trust is established, the chatbot can switch to the Maintain Stage.
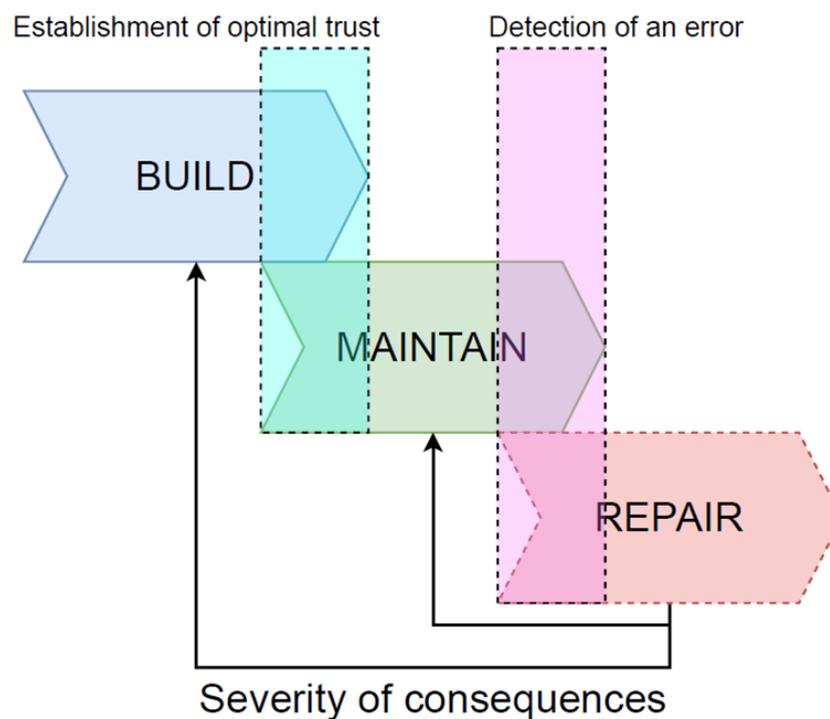
Figure 2 Dynamic Trust Model.

The Maintain stage.　It is crucial to maintain the relationship between the chatbot and its user to ensure the continuity of user engagement. The chatbot can achieve so with the following features:

- being adaptable and reliable
- obtaining ongoing feedback from users to analyse their behaviour, attitude and emotions.
- showing the chatbot's ability to recognise the user with reference to the interaction history

The Maintain stage is the desired stage of the chatbot for demonstrating its trustworthiness. However, when some components of the chatbot fail to accomplish their task, it moves to the Repair Stage.

The Repair Stage.　The chatbot aims to fix trust issues between the user and itself. The following actions are required:

- acknowledgment of error
- identification of error
- apologising
- repair and correction by learning
- reaffirming conforming to the shared goal

However, some errors do not need to be repaired and the interaction can be transferred back to the Maintain stage. The chatbot may decide that a user action, which has broken trust due to the failure of a certain task, is not important. Thus, it is acceptable to continue having trust in that component as it is. On the other hand, if the user detects an error committed by a chatbot and points it out, the chatbot can attempt to repair trust by issuing

an apology. Another condition is that if the chatbot is not confident about the accuracy of a response but delivers it nevertheless with a forewarning, it could mitigate the need to repair. In some cases where the chatbot decides that the broken trust cannot be repaired because it is either too costly to perform the repair or the repair is impossible with the approach used, then the chatbot may decide to build a new relationship for the same goal but with a different approach.

Overall, by assessing the severity of consequences of the broken trust, the chatbot decides whether (i) to repair trust, (ii) switch to the stage of maintaining the relationship, or (iii) build a completely new relationship for the same goal with a different strategy/approach.

Regaining Trust.   The following actions can be undertaken to regain the lost trust:
- adjusting the weights of individual factors of trust; such weights are application-dependent and user-specific
- referring to a taxonomy of CA can help fine-tuning the weights, which can also be supported by participatory design and empirical evaluation
- real-time signal detection and adaptation to allow CA to clarify intents, manage user expectations and update user models as strategies to adjust the weights of CA trust
- resolving mismatch between error performance and mental models (e.g., user verbal and non-verbal behaviours to infer emotion)

Basically, every chatbot type has different weights for individual trust factors, including:
- Inclusiveness (e.g., accessibility, non-discriminativeness)
- Competences
- Availability
- Warmth (e.g., friendliness, empathy)
- Legality
- Engagement
- Reliability (i.e., consistency)
- Professionalism (e.g., type of the language, avoiding typos and grammar mistakes, embodiment appearance)

4.2.3   Future Research

The following questions require further research effort to address:
- The notion of 'modality' entails clarification: Is multimodality integral to CAs or an add-on feature?
- How to define and operationalise the features (the above bullet points) in each of the three stages of the dynamic and temporal trust model?
- Are factors of trust hierarchical? Whether and how they can be prioritised?
- How can machine learning methods be used to determine the weights of individual factors, and adapt them with respect to contextual changes?

4.3 Breakout Group 2: Impact of CA

Group 2

Contributors:
Group 2a (in person): Sebastian Hobert, Ryan Schuetzler, Frode Guribye, Clayton Lewis, Stefan Schaffer, Martin Porcheron
Group 2b (online): Levi Witbaard, Margot van der Goot, Stefan Morana, Ana Paula Chaves, Jonathan Grudin, Heloisa Candello, Christine Liebrecht, Yi-Chieh Lee, Jasper Feine

4.3.1 Goal and key questions

Goal: Trust in CAs through positive social and commercial changes.
Key questions: How may CAs be applied for positive social and commercial impact?
Relevant aspects: Positive use scenarios; Negative use scenarios

4.3.2 Key insights

4.3.2.1 Group 2a

The group started with the basic question "What is social good?". Then they explored the ethical implications of designing CAs, identified research and development areas of CA as well as future work.

- **Definition of Social Good**: It can be defined in many forms, but in our breakout group, we adopted a utilitarian demarcation in which you design with the aim to maximize benefits for the individual users and society at large, while minimizing individual and societal risks. Of course, one can never predict the full consequences of any action, nor can we predict all the ways people will adapt to the affordances provided by our systems. Intent is core to driving a project for social good; designers, developers, organisations must adopt a stance to delivering social good.
- **Draft for a Code of Ethics for trustworthy CAs**: If driving social good is rooted in the design of CAs, as per our definition, we propose the following set of ethical guidelines for designers to consider:
  - Design CAs and their underlying systems to be worthy of users' trust, not just with the appearance of trustworthiness
  - Be open and explicit about the intent of the CA
  - Take care to recognize and design for marginalized groups
  - Consider the possible negative uses of the CA
  - Take responsibility for the intended and unintended consequences of CA use
  - Minimize risk of harm created by inaccurate responses, or through disclosure of private information
  - Consider possible sources of bias, including commercial interests, and be explicit when they might conflict with the users' best interests
  - Do not unnecessarily exploit the humanlike capabilities of a CA to deceive or manipulate users

- **Research and Development Areas for Social Good**: We believe that the strongest possibilities for social good chatbots are present when one or more of the following conditions are met: (1) the humanlike conversational capabilities of a CA that allow it to accomplish its goal better than a traditional system; (2) the CA can do things that people either cannot (either through lack of ability or resources) or are not willing to do. We have identified the following categories for the creation of CAs for social good. For each category, we begin to outline promising areas for development as well as potential pitfalls that must be addressed to maximize social good and minimise the risk of harm.
    - **Mental Health**: CAs can relieve shortages of mental healthcare workers and reach those who may not otherwise have access or may otherwise be hesitant to seek access. The digital nature of CAs can enhance user trust, and especially their trust that they are not being judged or evaluated based on their responses. However, especially in this category of CA, care must be taken to minimize the potential for inaccurate responses causing harm. While a CA can potentially reach and help many people, designers and researchers must do all they can to ensure the appropriate response to crisis situations such as suicidal thoughts.
    - **Virtual Companionship**: Loneliness has become epidemic, across all age groups under different conditions. Our CAs have the potential to relieve loneliness and provide a connection to those that may otherwise not have one. These agents can be designed to help the elderly who tend to experience loneliness more than others. In the design of these virtual companions, designers must take care to ensure individuals do not develop a dependency on the technology and avoid otherwise beneficial human contact in favour of virtual companionship.
    - **Learning**: The main purpose of CAs in educational settings is supporting instructors, teaching assistants or learners in-class, blended- or online settings instead of replacing them. Providing (automated and individualized) feedback, learning materials, or answers to individual questions in a conversational way seems to be useful particularly useful in large-scale settings in which otherwise learning support would not be offered or in small-scale settings in which offering manual support is effortful. This seems valuable to social good as a more educated populace worldwide increases. Some people especially benefit from a social connection associated with learning, as evidenced by the struggles some experienced during the virtual learning of the COVID-19 pandemic.
    - **Healthcare**: With a shortage of healthcare workers worldwide and limited resources to provide appropriate care, CAs have the potential to relieve pressure on strained human resources. By providing automated access to, for instance, screening and informational services, we can enhance availability and access to these vital services. CAs can also reach populations currently unserved or underserved by the healthcare system.

- **Participatory Design Activities to Ensure Trust in CA**: Since there might be a variation in perspectives, preferences, goals and values in terms of how social good is perceived as such we should not rely solely on the perspectives of the designers and developers when creating CAs. One way of ensuring that the design of a CA is properly anchored in the perspectives and values of its future users is to include them in design activities. Providing design environments and authoring tools that are easy to use can be an important step in empowering communities of practice to build their own CAs that meet the particular needs of the community. With available authoring tools for CAs, participants can engage in design activities that are more similar to end-user programming and tailoring of services and skills that can be aimed at meeting such particular needs.

- **Considerations of Trust**: Trust is especially important in "social good" applications because they often, but not always, deal with higher levels of risk. Trust, and trustworthiness, are important determinants of use when risk is high. As social impact applications are often used with emerging markets and unserved or underserved populations, trust and trustworthiness are critically important to not:
  - (Premature) Deployments of ineffective or harmful technology can hamper future research and developments, slowing down new developments for years to come (e.g., Clippy, Tay, Google Glass, or an over-eager deployment of self-driving cars). Over-promising and setting too-high expectations could erode future trust if the technology fails to meet expectations, even if it is better than alternatives. (e.g., even if a self-driving car is better than human drivers, it faces increased scrutiny, and if it kills people, it erodes societal trust and hinders the advance of future, better technology).
  - Because CAs are still somewhat an emerging technology, malicious, ineffective, or harmful use could result in erosion of trust at a general level (e.g., if Amazon was found to be selling information from private conversations near Alexa)

4.3.2.2 Group 2b

Outcomes of the discussion are summarised in Figure 3:
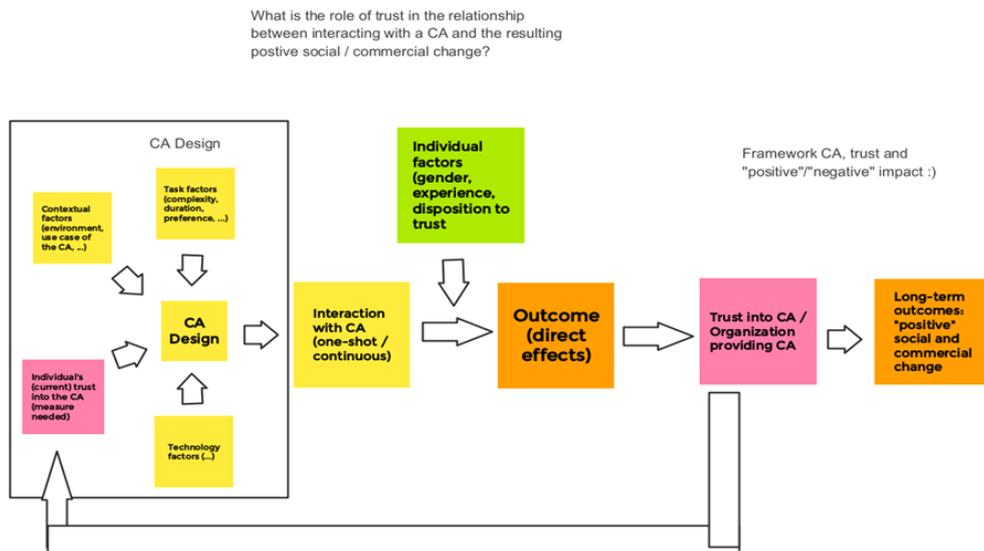
Figure 3 CA design and impact.

Accordingly, CA design is influenced by the contextual, technological and task factors. Interaction with CA (one-shot or continuous) leads to outcome in terms of direct effects, including trust in CA itself and the organisation providing CA.

4.3.3 Future Research

The following research questions need to be explored as future work:

Group 2a:

- Do transparency and explanations support trust in CAs for social good?
- Effectiveness – do our social good CAs actually produce social good, producing better outcomes for all or certain individuals or groups
- How to foster initial trust in CAs (for social good) and how to maintain the impression of trust over time?
- Trustworthiness vs. impression of trustworthiness? What is more important?
- Intent seems to be a core concept. Is intent the only difference related to trust between commercial CAs and CAs for social good?
- How to manage expectations in a new market without previous CA experience?

Group 2b:

- Should the CA remember? Should the CA immediately show that the CA has the history knowledge? To what extent is longer-term interaction needed from the perspective of the user?
- Should CA address stereotypes (e.g. gender stereotype) and how we as a community can contribute to fight against these stereotypes?
- How to continuously/automatically measure trust in CA? Are there approximations for trust rather than using a survey?
- Should the CA be able to measure user trust and adapt itself depending on the user's current level of trust?
- Continuous real-life long-term relationship (e.g. financial trading) is more difficult to study in an experimental context. How to ensure match between real-world case study and study design?

4.4 Breakout Group 3: Ethics of CA

Group 3

Contributors: Minha Lee, Björn Schuller, Elisabeth André, Leigh Clark, Asbjørn Følstad

4.4.1 Goal and key questions

Goal: Trust in CAs through ethical design and implementation.

Key questions:

- Which are key ethical aspects of CAs?
- How to design for ethical CAs use?

Relevant aspects: Transparency, bias, fairness, and digital marginalization, privacy/security, safeguard against misuse and error.

4.4.2   Key insights

Initial considerations – scoping the ethics challenge.   Ethics is a research topic of high relevance to trust in conversational agents. In their overview of future directions in chatbot research, Følstad et al. (2021) identified ethics as an area in need of substantial research efforts. Such future research has a valuable starting point in the existing background on ethics in AI-based systems (Hagendorff, 2020). For example, an EC High level Expert Group has detailed the ethical basis for trustworthy AI-systems in general (EC, 2019). Research on ethical aspects on conversational agents is also emerging (e.g., Ruane et al., 2019).

Given the broadness of research challenges pertaining to ethics in conversational agents, the group converged on a specific research challenge of high importance for ethics in conversational agents: conversational disclosure.

Conversational disclosure – a key ethical challenge in CAs.   Conversational disclosure concerns how to achieve transparency during interaction with conversational agents. Transparency is a key ethical requirement in AI-based systems (EC, 2019; Hagendorff, 2020), and concerns the need to (a) clarify to users that they are interacting with an automated system, not a human, (b) provide insight into how user data are processed and used, and (c) explain the relevant system characteristics and limitations to the user.

Transparency may be a particularly important ethics requirement for conversational agents as the interaction with such agents may easily be confused with interaction with humans, and users may also be inclined to share personal information as part of such interaction – for example as part of using conversational agents for mental health or relational purposes. The forthcoming European legal regulation of AI systems, the AI Act, will likely make transparency in conversational agents a legal requirement (Schaake, 2021) – so that providers are responsible for users understanding they are interacting with a conversational agent and not a human being.

Designing for transparency through conversational disclosure may be achieved by following two different paths: a guidelines-oriented approach and a practice-oriented approach; the two paths corresponding to a deontological vs. virtue-oriented approach to ethics in AI-systems (Hagendorff, 2020). The two paths are detailed below.

Guidelines-oriented approach to conversational disclosure.   In a guidelines-oriented approach, it may be valuable to consider how to provide conversational disclosure at different points in time during a prolonged period of use.

- *initial disclosure*, at the onset of the first interaction
- *routine disclosure*, at predefined milestones
- *requested disclosure*, initiated by the user.

For each of these forms of disclosure, research may address which items to include as part of the disclosure and how to design such disclosure so as to provide a good user experience. Could, for example, routine disclosure be designed so as to provide added value to the user? (e.g. presenting content from previous interactions for evocation or engagement, inspired by approaches to sharing insight based on the users' person information in services like Google Timeline or Strava.).

Practice-oriented approach to conversational disclosure.   A practice-oriented approach to conversational disclosure would concern establishing conversational disclosure as a craft skill. Establishing such a skill would include tackling challenges such as how to provide conversational disclosure without disturbing the flow of interaction.

For example, users of relational conversational agents may wish for such agents to be humanlike in their interaction so as to achieve a desired perceived companionship. At the same time, such agents should also be fully transparent to their users. Nevertheless, the interaction should not be interrupted at inappropriate points in time by having the agent explain itself.

Negotiating the need for human likeness on the one hand and conversational disclosure on the other is a design challenge that may require craft skill rather than guideline adherence. To establish a practice-oriented approach to conversational disclosure, ethics may need to be included in teaching and training on design of conversational agents.

4.4.3   Future Research

Ethics in conversational agents is an area of a broad range of research challenges. In this groupwork we addressed one such challenge, conversational disclosure. To further explore ethics in conversational agents, the groupwork participants have initiated a forthcoming CHI workshop on ethics in conversational user interfaces.

References
1   Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99-120.
2   EC (2019). Ethics Guidelines for Trustworthy Artificial Intelligence. Technical report, EC High level Expert Group on AI.
3   Ruane, E., Birhane, A., & Ventresque, A. (2019, December). Conversational AI: Social and Ethical Considerations. In AICS (pp. 104-115).
4   Schaake M (2021). The European commission's artificial intelligence act. Policy Brief, Stanford HAI

4.5   Breakout Group 4: AI and Technical Development for CA

Group 4

Contributors: Matthias Kraus, Roger K. Moore, Ricardo Usbeck, Ana Paiva, Rolf Pfister, Elayne Ruane.

4.5.1   Goal and key questions

Goal: Towards advancing the technical basis for trust in CAs
Key question: How may the technical basis for CAs be advanced to strengthen trust?
Relevant aspects: NLP – intent recognition, simulation of empathy, adaptation to users, software architecture, linguistics / dialogue management, repair, explainable AI, modelling.

4.5.2   Key insights

The topic of the technological basis for trust in CAs is very broad. The group addressed this by discussing some key issues over the course of the workgroup sessions. Summaries of the discussions are provided below.

The need for strengthened collaboration across research communities. Advancing the technical basis for trust in CAs is challenging due to disconnected nature of research communities. As a starting point for strengthen connections and exchange between communities, a list of research communities was compiled:

- CHI – ACM CHI Virtual Conference on Human Factors in Computing Systems
- CONVERSATIONS – workshop on chatbot research
- CUI – conference on conversational user interfaces
- IUI – conference on intelligent user interfaces
- HAI – conference on human agent interaction
- HRI – conference on human robot interaction
- INTERSPEECH – conference on spoken language processing
- IVA – conference on intelligent virtual agents
- LREC – conference on language resources and evaluation
- SemDial – workshop on the semantics and pragmatics of dialogue
- SIGdial – special interest group on discourse and dialogue
- Dagstuhl Seminar 20021 – SLIVAR – spoken language interaction with virtual agents and robots
- Dagstuhl Seminar 21381 – Trust-CA – conversational agent as trustworthy autonomous system

The need for strengthened collaboration between academic research and industry. There seems to be a disconnect between large commercial vendors and the academic community regarding research of relevance to the technological basis for CAs. Furthermore, a shift may be observed where high profile research increasingly is coming from large technology companies. Along with this, data and computational resources are increasingly isolated within commercial entities.

In consequence of "data as the new code", there is a need for researchers to access data to fully understand or replicate a system or research conducted on a system. However, challenges exist for sharing of data held by industry, including privacy risks and difficulties in sanitizing data at scale. Also, there may be a perceived loss of competitive advantage in sharing data resources.

Furthermore, there seems to be a talent-pipeline challenge in the AI space in which it is difficult for academia to attract and keep PhDs and postdocs due to the imbalance in financial compensation between these positions and the roles available within industry.

The challenge for industrial players to oversee and evaluate CAs. Automatic and comprehensive evaluation of CAs is technically challenging. There is a need for better evaluation methods for CA owners. The availability of tools, frameworks, and platforms has reduced barriers for uptake of CAs in industry. At the same time, there may be a lack of sufficient guidelines for practitioners within industry using these tools, e.g., for intent design and optimization. Hence, while creating a CA may be quick and low-effort, it can be challenging to design and develop a CA of sufficient quality to provide the desired user experience.

Possibly, CA owners relying on tools with insufficient documentation, guidelines, and transparency may be unaware of the limitations of component technologies and thus experience overtrust in those tools.

Investigation is required to establish best practice guidelines in this space. Specific guidelines will vary from one platform to another due to differences in model architecture, training data, and other platform features and modules such as entity recognition or sentiment detection.

Due to the black-box character and evolving nature of platform components, this may be something that needs to be done by platform owners. Furthermore, there is a need for confidence scores for component technologies to allow CA owners to build trust in their systems.

System architecture and complexity. An aspect of CA systems that makes it challenging to manage trust is system architecture and complexity. This complexity concerns, in part, end-to-end systems and large language models. As, for example, seen in challenges of handling bias in data and system output, which is important to a trustworthy system. Due to the inherent complexity in such systems, managing trust may come at a cost (e.g. accuracy). Complex modular systems need to spread trust along the chain of modules but tuning one component might affect another. Possibly, certifications may be developed to handle seemingly competing objectives in complex CA systems.

Ethics and transparency. The AI and technical development underpinning of CAs also entail a range of ethical issues. Tools and approaches such as emotion detection can have great benefit and be used in personalization but while some use cases can be ethical, other scenarios may be ethically questionable. One approach to addressing such ethical issues may be to look towards other fields that have faced similar challenges.

Transparency can increase trust in CAs. A CA's behaviour should be transparent. That is, it needs to be understandable but also to allow for in-depth insights, e.g. into the used data sources. The need for transparency may, however, vary between user groups. Hence, to achieve transparency in CAs user's roles and profiles need to be considered. CAs may also need to afford transparency with different modalities of interaction.

Conversational repair and trust. Conversational repair is important in CAs, to support needed adjustment or adaptation of dialogue to mitigate interpretational issues or misunderstandings. Repair strategies impact user trust and attitudes towards a chatbot. Detecting the need for conversational repair may be challenging and we currently lack sufficient automated approaches – for example in the case of false positive replies in CAs.

4.5.3 Future Research

Relevant next steps in research towards strengthening the technological basis for trust include:
- Strengthen opportunities for collaboration between academic research and industry, including how to share data or metadata when publishing technical research on CAs
- Develop guidelines for design, development, and evaluation of CAs, for use of all human actors in the CA supply chain.
- Research efforts addressing how to manage trust in complex systems enabling current and future CAs
- Research towards transparency and explainability in CAs
- Research addressing automatic conversational repair in CAs

4.6 Breakout Group 5: Definition, Conceptualization and Measurement of Trust

Group 5

Contributors:

Group 5a (in person): Martin Porcheron, Minha Lee, Birthe Nesset, Frode Guribye

Group 5b (online): Margot van der Goot, Roger K. Moore, Ricardo Usbeck, Ana Paiva, Catherine Pelachaud, Elayne Ruane

Group 5c (in person): Björn Schuller, Guy Laban, Dimosthenis Kontogiorgos, Matthias Kraus, Asbjørn Følstad

4.6.1 Goal and key questions

Goal: Enable assessment and measurement of trust in CAs

Key question: How to define, conceptualize and measure trust in CA?

Relevant aspects: Modelling frameworks – antecedents / consequents; basis in knowledge on human-human communication; psychology of trust – over-trust / intuition

4.6.2 Key insights

Defining trust. Trust is addressed in different disciplines, both as a general concept within psychology, sociology, and management research (e.g., Rousseau, 1998; Mayer et al., 1995) and – more recently – as a term of relevance for users' perceptions of technology (e.g., Corritore et al., 2003; McKnight et al., 2011). A range of definitions exists for trust. There is variation in definitions concerning whether trust should be construed as a belief or attitude (Lewis et al., 2018; Lee & See 2004), and the degree to which there is a behavioural element in trust (Söllner et al., 2016; Malle & Ullman, 2021).

For conceptual clarity, it may be useful to consider trust an attitude which may be founded by trusting beliefs, and which may lead to trusting behaviour.

Trusting behaviour is determined by trust and may as such be an indicator of trust – provided users have a choice. Trusting behaviour is also moderated by environment, user group, and use case. An example of trusting behaviour is self-disclosure. Trust may also impact engagement level in behaviour and tendency to repeated use.

Developing trust through conversational interactions. The notion of trust in technology arguably is of particular relevance to CAs, due to their conversational interaction with users. Conversations are humanlike which has implications for users trusting beliefs and behaviours. Furthermore, conversations may be relational, leading to expectations of evolving capabilities in agent. Conversations may also be cooperative, leading to expectations of mutual adaptations in the user and conversational agent to achieve a common goal.

On this background, trust in CA may be considered as gradually built through conversations. In consequence, four trust concepts may be of particular relevance for CAs:

- *Initial trust*: trust required for users to initiate interaction. Initial trust corresponds to the notion of calculating trust in Rousseau et al. (1998)

- *Sharing trust*: trust required for sharing information with chatbot. The relevance of sharing trust may depend on varying levels of perceived sensitivity in the domain or topic of CA interaction.
- *Reliance trust*: trust required for relying on chatbot recommendations or decision support, that is, trust impacting user beliefs or behaviour beyond the context of the CA interaction.
- *Long-term trust*: trust required for repeated / routine use. Long term trust corresponds to the notion of relational trust in Rousseau et al. (1998)

Extending the trust model of Rousseau et al. (1998), the four trust concepts for CAs may be mapped out on a timeline of the evolving relation between user and CA as follows:
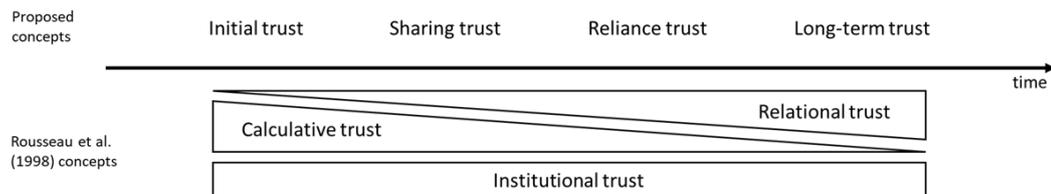
Figure 4 Extended trust model.

Balancing trust and trustworthiness. When considering trust, it is critical to distinguish between perceived trust and trustworthiness.

Perceived trust is held by the trustor, typically the user. Perceived trust and related trust beliefs may be measured through a range of self-report measurements, for example from information systems research (e.g. Lankton et al., 2015), social robotics (c.f. review in Hancock et al., 2020). Perceived trust may be impacted by the trustworthiness of the trustor. However, as information on this may not be available, other characteristics may impact trust. For CAs, anthropomorphism may be such a characteristic, as it may impact trust though not be correlated with trustworthiness.

Trustworthiness is a characteristic of the trustee, typically the service provider. Trustworthiness may depend on factors such as transparency, reliability, consistency, sincerity, honesty, integrity, benevolence, competence, and cooperation. These factors, though not necessarily static, may be considered observable characteristics in a trustee.

There is a need to study trustworthiness and perceived trust in parallel – to address potential overtrust (low trustworthiness and high perceived trust) and undertrust (high trustworthiness and low perceived trust). There is a lack of approaches or measurements for the integrated study of trustworthiness and trust.

Measuring trust by integrating self report measures and behavioural measures. In existing scales and measurements, trust is typically construed as personal, mainly available to researchers through self report measurements. Nevertheless, trust can be interpreted as reflected in and through people's behaviour, rather than merely a stance prior to the use of some device or system. Trust as reflected in trusting behaviour may enables trust to be measured also on the basis of user behaviour. There seem however to be a lack of distinct behaviour scales for trust assessment.

Possibly, trust may be measured by having a CA asking about sensitive information and monitor users' disclosing behaviour. Specifically, a tiered approach may be useful, based on asking questions of personal information of increasing level of sensitivity to infer a person's

level or trust. However, the choice of behavioural measures of trust may depend on the context of the CA.

An integrated approach, combining self-report measures and tiered behavioural measures seems a promising approach for future research.

A proposed integrating framework for measuring trust and trusting behaviour. Following from the above, instruments and data sources for measuring trust may be divided into two broad groups: Subjective and objective measures:

- *Subjective measures* concern the measurement of trust determinants / trusting beliefs or behavioural intent (e.g., Lankton et al., 2015; Yagoda & Gillan, 2012). As a subjective measurement, perceptions of trust are expected to be explicit from the subject's report, corresponding to the subject's trust beliefs. Nevertheless, these might not be consistent with the subject's trusting behaviour due to personal perceptions and attitudes of the subject regarding the conversational AI system – e.g., due to scepticism of AI (Araujo et al., 2020).
- *Objective measures* include measures of physiological states, speech / voice, interaction with agent (e.g., sharing behaviour), changes in beliefs due to agent, behaviour in the world due to agent. Accordingly, the subject's behaviour would implicitly indicate higher or lower levels of trust. THe association between trusting behaviour and trust should be studied individually, depending on context, settings and task. Within the scope of conversational AI, behaviour such as self-disclosure (e.g., Laban et al., 2021a), reciprocity (e.g., Zonca et al., 2021), and changes in disclosure and expression over time (e.g., Laban et al., 2021b) could implicitly indicate changes in trust. These behaviours, however, might not be consistent with one's trust beliefs due to, for example, habits and needs (e.g., having the need to share, or being an impulsive individual) or affect-based factors of trust like the system's heuristics and demonstrated social cues (e.g., one might be more likely to share information with a more persuasive system despite not trusting it; e.g., Ghazali et al., 2019).

Subjective and objective measures may be included in a framework of trusting beliefs and trusting behaviour as follows:

4.6.3   Future Research

The following questions require further research effort to address:
- Developing a comprehensive framework to capture how trust evolves across long-term use.
- Refining the framework for trusting beliefs and trusting behaviour.
- Developing integrated approaches and measures for studying users perceived trust and the trustworthiness of service providers, to mitigate overtrust and undertrust.
- Developing integrated measures of trust and trusting behaviour, combining self report measures and tiered behavioural measures to support standardised measure for trust in conversational agents, and incorporating this in conversational systems.
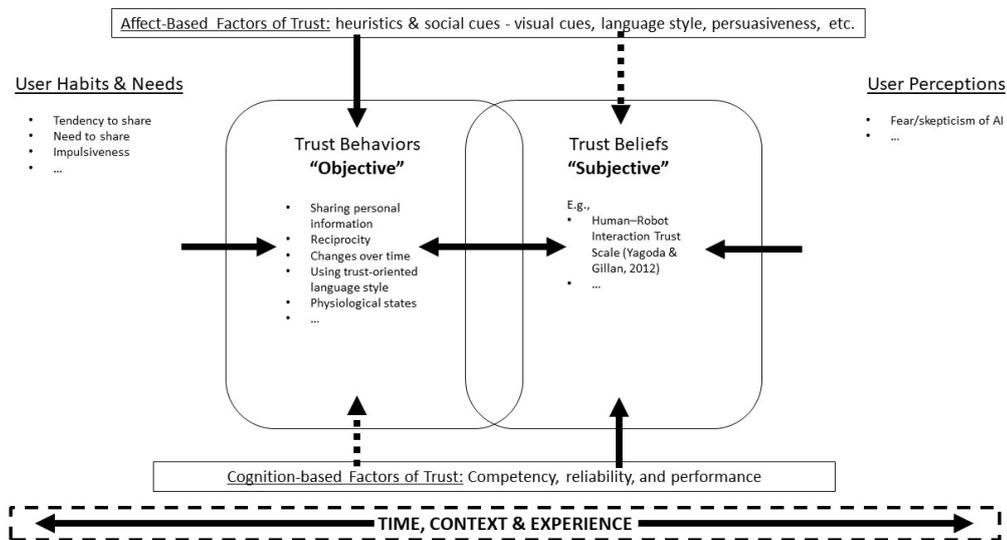
Figure 5 Framework of trusting beliefs and trusting behaviour.

References

1   Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI and Society, 35(3), 611 – 623.

2   Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. International Journal of Human-Computer Studies, 58(6), 737-758.

3   Ghazali, A. S., Ham, J., Barakova, E., & Markopoulos, P. (2019). Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance. Advanced Robotics, 33(7 − 8), 325 − 337.

4   Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2020). Evolving trust in robots: specification through sequential and comparative meta-analyses. Human factors, 0018720820922080.

5   Laban, G., George, J.-N., Morrison, V., & Cross, E. S. (2021a). Tell me more! Assessing interactions with social robots from speech. Paladyn, Journal of Behavioral Robotics, 12(1), 136 − 159.

6   Laban, G., Kappas, A., Morrison, V., & Cross, E. S. (2021b). Protocol for a Mediated Long-Term Experiment with a Social Robot. PsyArXiv.

7   Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. Journal of the Association for Information Systems, 16(10).

8   Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.

9   Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In Foundations of trusted autonomy (pp. 135-159). Springer, Cham.

10 Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In Trust in Human-Robot Interaction (pp. 3-25). Academic Press.

11 Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of management review, 20(3), 709-734.

12 McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. ACM Transactions on management information systems (TMIS), 2(2), 1-25.

13 Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23(3), 393-404.

14 Söllner, M., Hoffmann, A., & Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. European Journal of Information Systems, 25(3), 274-287.

15 Yagoda, R. E., & Gillan, D. J. (2012). You Want Me to Trust a ROBOT? The Development of a Human – Robot Interaction Trust Scale. International Journal of Social Robotics, 4(3), 235 – 248.

16 Zonca, J., Folsø, A., Sciutti, A. (2021). The role of reciprocity in human-robot social influence. iScience, 24(12), 103424.

4.7 Breakout Group 6: Interaction Design

Group 6

Contributors:
Group 6a (in person): Oliver Bendel, Sebastian Hobert, Ryan Schuetzler, Elisabeth André, Leigh Clark, Clayton Lewis, Stefan Schaffer, Eren Yildiz, Effie Law
Group 6b (online): Stefan Morana, Heloisa Candello, Christine Liebrecht, Zhou Yu, Dakuo Wang, Michelle Zhou, Ana Paula Chaves, Cosmin Munteanu, Soomin Kim

4.7.1 Goal and key questions

Goal: Identify interaction designs to strengthen trust in CA.
Key questions: How to design trusted conversational user interfaces?
Relevant aspects: UX – human-AI sociability; Human-in-the-loop; Evaluation – reliability/acceptance; Group interaction.

4.7.2 Key insights

4.7.2.1 Group 6a

The group started with reflecting on the following aspects of trust:

- Brand and UX: The producer of a chatbot affects our perception of trust; we trust certain products because we trust certain brands; the implication of UX-trust relation
- Group effect: If we trust people, and those people trust a chatbot, then we are more likely to trust it as well.
- Domain-dependency: Certain domains are more sensitive to trust fluctuation
- Modality: Intricate relations between modality, risk and trust

Next, the group focused on some specific aspects of conversational interactions. The multifaceted nature of trust and the numerous factors of people's interactions with CAs that can impact on perceptions and behaviours, complicating our understanding of how, why and when to design for trust and subsequently evaluate it. We present a discussion of some critical aspects of CA interactions and highlight the need for a holistic approach to creating trustworthy CAs.

- **Multimodality** A critical question is whether multimodality can increase or decrease trust. On the other hand, multimodality might help to increase accuracy (i.e. automatic emotion recognition) which might help to increase trust in CAs. On the other hand, multimodality could decrease trust (privacy issues due to permission requests to webcam or other sensors). It depends on whether requesting permissions to access multiple sensors, e.g., on a phone, might test the trust of users, especially in the case of iPhone where each sensor is requested in sequence might annoy users. Nonetheless, it could creep-out users if the multimodal information is used in inappropriate or clumsy ways, especially when the verbal information conflicts with the nonverbal (e.g., "You say that you are happy, but your voice sounds sad. Are you depressed?"). Here below we discuss several aspects of multimodality:
  - **Preferences**: Cultural aspects can lead to different preferences among users. Some users may not like to use voice but text input only (if this is possible). An adaptation to such individual preferences should be considered during CA design. According to individual differences of users, a customization would be desirable as some people might prefer different modalities for interaction. Preferences may include choosing the voice, the tone of the voice, or the formal/informal style.
  - **Input**: An important question for future research is how multimodal sensory perception can be used to enhance CA effectiveness/accuracy. Depending on the use case, different multimodal sensors could be used to improve the interaction, including keyboard, camera, microphone, as well as accelerometer, thermal sensor, GSR and others. A fusion of the information coming from modality specific modules should generate a more reliable intention detection.
  - **Output generation**: The answers from the system have to be output using the appropriate modality. Usually a symmetry between the input and the output modality is expected by the user. The output generation module has to prepare the system feedback for the necessary modalities. This can include text generation, speech synthesis or graphic generation. When generating output text, the CA often has to integrate database answers into output text. Thereby errors can occur while producing the correct form of the word(s), e.g. if it's singular or plural, and the correct cases (Genitive, Dative or Accusative), or verb form. Today mostly templates are used to generate output prompts. Neural methods taking into account the integration of such database results are not yet mature.
- **Transparency** Trust might be fostered if the CA provides explanations about what it is capable of doing or understanding. The relation between trust and transparency: Is it reasonable to assume the more transparency we have, the more trust we get? Feedback from the chatbot should be personalized. If I want shorter feedback, the chatbot should do it so. Furthermore, other aspects of multimodality have to be considered:
  - Explaining why certain permissions might be needed: Do we trust the explanation?
  - Do explanations matter? Are too many permissions/explanations detrimental to trust or acceptance?
  - Baseline level based on your general preferences
- **Voice and Language** There are numerous features of CA speech that can impact on people's perceptions and behaviours. Features of voice quality, "those characteristics which are present more or less all the time that a person is talking' (Abercrombie, 1967, p. 91 in Laver, 1980, p. 1)," include an agent's perceived accent, gender, age, prosody and human-likeness. In addition to voice quality, the linguistic content delivered by a

CA can have similar impacts. Examples include language, dialect, register and style (e.g. Bendel 2018).

- **Context** Using context information (e.g., in learning contexts) might help to provide more accurate answers. If context information is missing, it might be annoying for users.
  - Application context (e.g., health, mental health, customer service)
  - Environmental context (e.g., room, building)
  - Social context

- **Evaluation**
  - A user-centered design process is important. Co-design or participatory design or human-centred approaches will help.
  - Questionnaire including trust related scales: e.g., https://ueqplus.ueq-research.org
  - How to measure trust using questionnaires and without questionnaires? Is it possible?

- **Should a CA Be Humanlike?** What is humanlikeness? Is it the ability of the bot to sound human, talk like a human? Or the ability to do what a human would do? The humanlikeness of the CA, at least insofar as it does not enter the uncanny valley, is likely to increase trust as long as the bot is upfront about its botness. Alexa's voice and capabilities could improve to the point of being completely humanlike. This may be related to the notion of partner models, which "refer...to a person's internal representation of an interlocutor's (human or machine) dialogic competence" (Doyle et al., 2021). Some studies have pointed out situations in which a more humanlike agent underperforms compared to a less humanlike agent with respect to a desired outcome (e.g., Schuetzler et al. 2018). These findings at least suggest that humanlikeness and its consequences are not universally desirable.

4.7.2.2 Group 6b

The main points of the discussion on the key question "How to design trusted conversational user interface" are categorised summarised in the following:

- *Domain*: design of CA is domain-dependent, as shown in examples: tourism, education for early childhood, financial, healthcare, informal public spaces such as museum
- *Transparency*: Explainable AI; Personal identifiable data storage (what do you know about me); Split the content in small chunks/topics; Strategies to show many options – personalised recommendations tailored by interest
- *Chatbot language design*: Humanlike design increases frustration; Register theory (age, location, language style); Infrastructure behind the chatbot
- *Accuracy*: answer as expected
- *Relationship*: engagement, satisfaction
- *Voice and text interfaces*: Speech interfaces can have higher cognitive load than text ones, depending on the task; Text interface – privacy information
- *Conversation flow*: Proactive vs. reactive bot; Decision-making system vs. informational bots; Disambiguation; Repair strategies; Multi-bot vs. single-bot
- *Settings*: privacy and public settings
- *Development*: technical devices usability and bugs

4.7.3   Future Research

The following research topics on CA interaction design can be explored as future work:

- Individual differences – configurable preferences are one way to adapt an agent to individual differences, but we must remember that trends/correlations/construct relationships are typically studied in aggregate, but individuals vary significantly from the mean.
- Identify which research findings that are generalizable across a variety of contexts and which are limited to within some specific context
- Potential limitations and ethical considerations of imitating human-likeness in CA design
- Resolving conflicts from multimodal sensors
- Impact of different styles/levels of embodiment (e.g. robotics, virtual avatars) on trust
- How best to appropriately evaluate the impact of interaction design choices on trust

References
1   Abercrombie, D. (1967). Elements of general phonetics (Vol. 203). Edinburgh: Edinburgh University Press.
2   Bendel, Oliver. (2018). From GOODBOT to BESTBOT. In: The 2018 AAAI Spring Symposium Series. AAAI Press, Palo Alto 2018. pp. 2-9.
3   Doyle, P. R., Clark, L., & Cowan, B. R. (2021, May). What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-14).
4   Laver, J. (1980). The phonetic description of voice quality: Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.
5   Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker, J. F. (2018). The Influence of Conversational Agent Embodiment and Conversational Relevance on Socially Desirable Responding. Decision Support Systems, 114(October), 94 – 102.

5   Open problems

5.1   Trust-CA: Conclusion and Suggested Readings

Trust-CA All

5.1.1   Conclusion

Through the twenty talks, six breakout groups and informal discussions, the Seminar's attendees explored the topic of Trust-CA widely as well as deeply. As the field is emerging, there are still many questions to answer, as shown in the report of each of the breakout groups. Among them, ethics of CA is a key concern. In fact, in the pre-Seminar survey, many of the respondents mentioned different aspects of ethics pertaining to CA and other AI-infused autonomous systems. Ethical considerations are highly relevant to the three main challenges for the Seminar (see Executive Summary). While the outcomes of the Seminar can provide insights to resolving these challenges, more research efforts are required. Encouraging dialogues and collaborations among different research communities working on conversational agents is essential for advancing this field. The Seminar Trust-CA has made a critical step along this direction.

5.1.2   Suggested Readings

In moving forward, it is necessary to review what has achieved in the past through reading the related publications. Prior to the seminar, the organizers asked the attendees to list their recommended readings of relevance to trust in conversational agents. The following readings were suggested.

References

1   Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Society.

2   Batliner, A., Hantke, S., & Schuller, B. W. (2020). Ethics and good practice in computational paralinguistics. IEEE Transactions on Affective Computing.

3   Bendel, O. (2018) From GOODBOT to BESTBOT. In The 2018 AAAI Spring Symposium Series (pp. 2-9). Palo Alto, CA: AAAI Press.

4   Bendel, O. (2020). The morality menu project. In M. Nørskov, J. Seibt, O. S. Quick, (Eds.). Culturally Sustainable Social Robotics – Challenges, Methods and Solutions: Proceedings of Robophilosophy 2020. (pp. 257-268). Amsterdam, Netherlands: IOS Press

5   Bendel, O., Schwegler, K., Richards, B., (2017). Towards Kant machines. In The 2017 AAAI Spring Symposium Series (pp. 7-11). Palo Alto, CA: AAAI Press.

6   Bhaskara, A., Skinner, M., & Loft, S. (2020). Agent transparency: A review of current theory and evidence. IEEE Transactions on Human-Machine Systems, 50(3), 215 – 224.

7   Biancardi, B., Dermouche, S., & Pelachaud, C. (2021). Adaptation mechanisms in human-agent interaction: Effects on User's Impressions and Engagement. Frontiers in Computer Science.

8   Bodó, B. (2020). Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators. New Media & Society.

9   Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., and Cowan, B. R. (2019). What makes a good conversation? challenges in designing truly conversational agents. In Proceedings of CHI'19. New York, NY: ACM.

10 Cranshaw, J., Elwany, E., Newman, T., Kocielnik, R., Yu, B., Soni, S., ... & Monroy-Hernández, A. (2017). Calendar. help: Designing a workflow-based scheduling agent with humans in the loop. In Proceedings of CHI'17. (pp. 2382-2393). New York, NY, ACM.

11 Cross, E. S., & Ramsey, R. (2021). Mind Meets Machine: Towards a Cognitive Science of Human-Machine Interactions. Trends in Cognitive Sciences, 25(3), 200 – 212.

12 De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. Journal of Experimental Psychology: Applied, 22(3), 331 – 349.

13 De Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': the importance of trust repair in human – machine interaction. Ergonomics, 61(10), 1409 – 1427.

14 Devillers, L., Kawahara, T., Moore, R. K., & Scheutz, M. (Eds.) (2020). Spoken language interaction with virtual agents and robots (SLIVAR): Towards effective and ethical interaction (Dagstuhl Seminar 20021). Dagstuhl Reports (Vol. 10). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum fur Informatik.

15 Falcone, R., & Castelfranchi, C. (2004). A belief-based model of trust. In Trust in Knowledge Management and Systems in Organizations (pp. 306-343). IGI Global.

16 Fan, X., Chao, D., Zhang, Z., Wang, D., Li, X., Tian, F. (2021) Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case study. Journal of Medical Internet Research, 23(1).

17   Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. International Journal of Human-Computer Studies, 132, 138 – 161.

18   Følstad A., Nordheim C.B., & Bjørkli C.A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. In Proceedings of the International Conference on Internet Science – INSCI 2018. LNCS, vol 11193. Cham, Switzerland: Springer.

19   Følstad, A., Araujo, T., Papadopoulos, S., Law, E. L. C., Luger, E., Goodwin, M., & Brandtzaeg, P. B. (Eds.). (2021). Chatbot Research and Design: 4th International Workshop, CONVERSATIONS 2020, Virtual Event, November 23-24, 2020, Revised Selected Papers (Vol. 12604). Springer Nature.

20   Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In International Symposium on Collaborative Technologies and Systems (pp. 106-114)

21   Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 145-152). New York, NY: ACM.

22   Garfinkel, H. (1963). A conception of and experiments with "trust" as a condition of stable concerted actions. In O. J. Harvey (Ed.), Motivation and Social Interaction: Cognitive Determinants (pp. 187-238). Ronald Press.

23   Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. Omega, 32(6), 407 – 424.

24   Grudin, J., & Jacques, R. (2019). Chatbots, humbots, and the quest for artificial general intelligence. In Proceedings of CHI'19. New York, NY: ACM.

25   Han, X., Zhou, M. X., Turner, M., & Yeh, T. (2021). Designing effective interview chatbots: Automatic chatbot profiling and design suggestion generation for chatbot debugging. In Proceedings of CHI'21. New York, NY: ACM.

26   Hepenstal, S., Kodagoda, N., Zhang, L., Paudyal, P., & Wong, B. L. (2019). Algorithmic transparency of conversational agents. In Joint Proceedings of the ACM IUI 2019 Workshops. CEUR Workshop Proceedings.

27   Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407 – 434.

28   Jaisie S., Munteanu, C., Ramanand, N., & Tan, Y. R. (2021). VUI influencers: How the media portrays voice user interfaces for older adults. In CUI 2021 – 3rd Conference on Conversational User Interfaces (Article 8), New York, NY: ACM,

29   Khadpe, P., Krishna, R., Fei-Fei, L., Hancock, J. T., & Bernstein, M. S. (2020). Conceptual metaphors impact perceptions of human-AI collaboration. In Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2). New York, NY: ACM.

30   Komiak, S. Y., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Quarterly, 941 – 960.

31   Kulms, P., & Kopp, S. (2019). More human-likeness, more trust? The effect of anthropo-morphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. In Proceedings of Mensch und Computer 2019 (MuC'19) (pp. 31-42). New York, NY: ACM.

32   Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. Journal of the Association for Information Systems, 16(10), 880 – 918.

33   Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50-80.

34   Lee, Y. C., Yamashita, N., & Huang, Y. (2020). Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. In Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1). New York, NY: ACM.

35   Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In Foundations of trusted autonomy (pp. 135-159). Cham, Switzerland: Springer.

36   Lucas, G., Stratou, G., Lieblich, S., & Gratch, J. (2016). Trust me: multimodal signals of trustworthiness. In ICMI '16: Proceedings of the 18th ACM International Conference on Multimodal Interaction (pp. 5-12). New York, NY: ACM.

37   Marche, S. (2021). The Chatbot Problem. The New Yorker, July 23, 2021.

38   Marge, M., Espy-Wilson, C., Ward, N. G., Alwan, A., Artzi, Y., Bansal, M., ... & Yu, Z. (2021). Spoken language interaction with robots: Recommendations for future research. Computer Speech & Language.

39   Meng, L. & Schaffer, S. (2020). A reporting assistant for railway security staff. In Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20) (article 31). New York, NY: ACM.

40   Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. Journal of Consumer Research, 26(4), 323 – 339.

41   Moore, R. K. (2017). Appropriate voices for artefacts: some key insights. In 1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR-2017) (pp. 7-11). Skovde, Sweden: VIHAR.

42   Moradinezhad, R., Solovey, E.T. (2021). Investigating trust in interaction with inconsistent embodied virtual agents. International Journal of Social Robotics.

43   Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. International Journal of Man-Machine Studies, 27(5-6), 527 – 539.

44   Munteanu, C.; Benett, A.; Rafih, H., Liaqat, A., & Aly, Y. (2018) Designing for older adults: Overcoming barriers to a supportive, safe, and healthy retirement. Wharton Pension Research Council Working Papers.

45   Nickel, Philip J. (2013). Trust in technological systems. In: M. J. de Vries, S. O. Hansson, & A. W. Meijers (eds.): Norms in Technology (pp. 223-237). Dordrecht, Netherlands: Springer

46   Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230 – 253.

47   Parviainen, E., & Søndergaard, M. L. J. (2020). Experiential qualities of whispering with voice assistants. In Proceedings of CHI'20. New York, NY: ACM.

48   Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018, April). Voice interfaces in everyday life. In Proceedings of CHI'18 (paper 640). New York, NY: ACM.

49   Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: trust-building factors and implications for conversational agent design. International Journal of Human-Computer Interaction, 37(1), 81 – 96.

50   Schaefer, K. E. (2016). Measuring trust in human robot interactions: Development of the trust perception scale-HRI. In Robust Intelligence and Trust in Autonomous Systems (pp. 191-218). Springer, Boston, MA.

51   Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human Factors, 58(3), 377 – 400.

52   Shneiderman, B. (2020). Human-centered artificial intelligence: Trusted, reliable & safe. International Journal of Human – Computer Interaction, 36(6), 495 – 504.

53   Sin, J., & Munteanu, C. (2020). An empirically grounded sociotechnical perspective on designing virtual agents for older adults. Human – Computer Interaction, 35(5-6), 481 – 510.

54   Skjuve, M., & Brandtzæg, P. B. (2018). Chatbots as a new user interface for providing health information to young people. In Youth and News in a Digital Media Environment – Nordic-Baltic Perspectives (pp. 59-66). Gothenburg, Sweden: Nordicom.

55   Torre, I., Goslin, J., White, L., & Zanatto, D. (2018). Trust in artificial voices: A "congruency effect" of first impressions and behavioural experience. In Proceedings of the Technology, Mind, and Society. New York, NY: ACM.

56   Van Pinxteren, M. M. E., Pluymaekers, M., & Lemmink, J. G. A. M. (2020). Human-like communication in conversational agents: a literature review and research agenda. Journal of Service Management, 31(2), 203 – 235.

57   Wang, L., Wang, D., Tian, F., Peng, Z., Fan, X., Zhang, Z., Yu, M., Ma, X., & Wang, H. (2021) CASS: Towards building a social-support chatbot for online health community. In Proceedings of the ACM on Human-Computer Interaction 5(CSCW1) (article 9). New York, NY: ACM.

58   Xiao, Z., Zhou, M. X., Chen, W., Yang, H. & Chi, C. (2020). If I hear you correctly: Building and evaluating interview chatbots with active listening skills. In Proceedings of CHI'20. New York, NY: ACM.

59   Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. Computers & Education, 161.

60   Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting virtual agents: The effect of personality. ACM Transactions on Interactive Intelligent Systems (TiiS), 9(2-3).

Participants

- Elisabeth André
Universität Augsburg, DE
- Oliver Bendel
FH Nordwestschweiz –
Windisch, CH
- Leigh Clark
Swansea University, GB
- Asbjørn Følstad
SINTEF – Oslo, NO
- Frode Guribye
University of Bergen, NO
- Sebastian Hobert
Georg August Universität –
Göttingen, DE
- Andreas Kilian
Universität des Saarlandes –
Saarbrücken, DE

- Dimosthenis Kontogiorgos
KTH Royal Institute of
Technology – Stockholm, SE
- Matthias Kraus
Universität Ulm, DE
- Guy Laban
University of Glasgow, GB
- Effie Lai-Chong Law
Durham University, GB
- Minha Lee
TU Eindhoven, NL
- Clayton Lewis
University of Colorado –
Boulder, US
- Birthe Nesset
Heriot-Watt University –
Edinburgh, GB

- Catherine Pelachaud
Sorbonne University – Paris, FR
- Martin Porcheron
Swansea University, GB
- Stefan Schaffer
DFKI – Berlin, DE
- Ryan Schuetzler
Brigham Young University –
Provo, US
- Björn Schuller
Universität Augsburg, DE
- Eren Yildiz
University of Umeå, SE

Remote Participants

- Theo Araujo
University of Amsterdam, NL
- Susan Brennan
Stony Brook University, US
- Heloisa Candello
IBM Research – Sao Paulo, BR

- Ana Paula Chaves
Federal University of Technology
– Paraná, BR
- Cristina Conati
University of British Columbia –
Vancouver, CA
- Benjamin Cowan
University College – Dublin, IE

- Laurence Devillers
CNRS – Orsay, FR & Sorbonne
University – Paris, FR
- Jasper Feine
KIT – Karlsruher Institut für
Technologie, DE
- Jonathan Grudin
Microsoft – Redmond, US

Evelien Heyselaar
Radboud University
Nijmegen, NL

Soomin Kim
Seoul National University, KR

Stefan Kopp
Universität Bielefeld, DE

Yi-Chieh Lee
NTT – Kyoto, JP

Oliver Lemon
Heriot-Watt University –
Edinburgh, GB

Q. Vera Liao
IBM TJ Watson Research Center
– White Plains, US

Christine Liebrecht
Tilburg University, DE

Roger K. Moore
University of Sheffield, GB

Stefan Morana
Universität des Saarlandes –
Saarbrücken, DE

Cosmin Munteanu
University of Toronto
Mississauga, CA

Ana Paiva
INESC-ID – Porto Salvo, PT

Symeon Papadopoulos
CERTH – Thessaloniki, GR

Caroline Peters
Universität des Saarlandes –
Saarbrücken, DE

Rolf Pfister
Cognostics – Pullach, DE

Olivier Pietquin
Google – Paris, FR

Aleksandra Przegalinska
Kozminski University, PL

Elayne Ruane
University College Dublin, IE

Marita Skjuve
SINTEF – Oslo, NO

Cameron Taylor
Google – London, GB

Ricardo Usbeck
Universität Hamburg, DE

Margot van der Goot
University of Amsterdam, NL

Dakuo Wang
IBM T.J. Watson Research
Center – Yorktown Heights, US

Saskia Wita
Universität des Saarlandes –
Saarbrücken, DE

Levi Witbaard
OBI4wan – Zaandam, NL

Zhou Yu
Columbia University –
New York, US

Michelle X. Zhou
Juji Inc. – Saratoga, US

Sparsity in Algorithms, Combinatorics and Logic

Edited by

Daniel Král'[1], Michał Pilipczuk[2], Sebastian Siebertz[3], and Blair D. Sullivan[4]

1   Masaryk University – Brno, CZ, dkral@fi.muni.cz
2   University of Warsaw, PL, michal.pilipczuk@mimuw.edu.pl
3   University of Bremen, DE, siebertz@uni-bremen.de
4   University of Utah – Salt Lake City, US, sullivan@cs.utah.edu

───── **Abstract** ─────

This report documents the program and the outcomes of Dagstuhl Seminar 21391 "Sparsity in Algorithms, Combinatorics and Logic". The seminar took place in a hybrid format from September 26 – October 1, 2021 and brought together 61 researchers. This report includes a discussion of the motivation of the seminar, presentation of the overall organization, abstracts of talks, and a report from each of the working groups.

1 Executive Summary

Daniel Král' (Masaryk University – Brno, CZ, dkral@fi.muni.cz)
Michał Pilipczuk (University of Warsaw, PL, michal.pilipczuk@mimuw.edu.pl)
Sebastian Siebertz (University of Bremen, DE, siebertz@uni-bremen.de)
Blair D. Sullivan (University of Utah – Salt Lake City, US, sullivan@cs.utah.edu)

Motivation

It was realized already in the early days of computer science that structures (networks, databases, etc.) that are sparse appear ubiquitously in applications. The sparsity of input can be used in a variety of ways, e.g. to design efficient algorithms. This motivates a theoretical study of the abilities and limitations of sparsity-based methods. However, a priori it is not clear how to even define sparsity formally. Multiple sparsity-oriented paradigms have been studied in the literature, e.g. bounded maximum or average degree models, topologically constrained classes of graphs or graphs with bounded width parameters. However, many of those paradigms suffer from being either too restrictive to model real-life applications, or too general to yield strong tractability results.

Sparsity in Algorithms, Combinatorics and Logic, *Dagstuhl Reports*, Vol. 11, Issue 08, pp. 115–128
Editors: Daniel Král', Michał Pilipczuk, Sebastian Siebertz, and Blair D. Sullivan
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In the late 2000's, Nešetřil and Ossona de Mendez proposed a new framework of uniform structural sparsity for classes of graphs that generalized existing definitions and initiated the development of a toolbox of sparsity-based methods for analyzing graphs. The central notions of their framework are bounded expansion and nowhere dense classes. It quickly turned out that the proposed notions can be used to build a mathematical theory of sparse graphs that offers a wealth of tools, leading to new techniques and powerful results. This theory has been extensively developed in the recent years.

It is particularly remarkable that the concepts of classes of bounded expansion and nowhere dense classes can be connected to fundamental ideas from multiple other fields of computer science, often in a surprising way, providing several complementary viewpoints on the subject. On one hand, foundations of the area are grounded in structural graph theory, which aims at describing structure in graphs through various decompositions and auxiliary parameters. On the other hand, nowhere denseness seems to delimit the border of algorithmic tractability of first-order logic, providing a link to finite model theory and its computational aspects. Finally, there is a fruitful transfer of ideas to and from the field of algorithm design: sparsity-based methods can be used to design new, efficient algorithms, especially in the paradigms of parameterized complexity, approximation algorithms, and distributed computing. Further, classic techniques for designing algorithms on sparse inputs inspire new combinatorial results on sparse graphs.

The aim of this Dagstuhl Seminar is to bring together researchers working on various aspects of sparsity in their own fields, in order to facilitate the exchange of ideas, methods and questions between different communities. So far, the synergy effect between graph theory, logic, and algorithm design has led to fundamental developments in the theory of sparse graph classes. Our goal is to inspire a new wave of developments by "stirring in the pot" of researchers working on different facets of sparsity. An important part of the seminar will be the discussion of the (still fledgling) area of real-life applications of sparsity-based methods, where theory and practice could meet.

Seminar organization

Due to the on-going COVID pandemic, the seminar was held in hybrid format. In total the seminar was attended by 61 participants around the world (from North America to Europe to Asia). 32 of the participants were on-site and 29 remote. To make the hybrid format a success and in particular allow all members to participate in talks and working groups, the following measures were taken.

1. To accommodate both on-site and remote participants, a mix of on-site and remote talks were scheduled. The talks were scheduled in the early to late afternoon (MEZ local time), allowing remote audience members from all parts of the world to attend.

2. Both on-site and remote talks were streamed via zoom. The zoom session was projected onto a whiteboard in the seminar room. The remote participants could see and hear the on-site whiteboard and slide presentations. They could interrupt and ask questions or ask questions in the chat, which were then read by the organizers. This turned out to be a quite successful setup in which all participants could discuss in real-time.

3. On the first day of the seminar we had a short introduction of all participants, one invited tutorial lecture, one contributed talk and the open problem session. In total, we had 5 tutorial lectures and 12 contributed talks spread over the week. The topics and speakers were chosen to create a joint understanding of the state of the art in the fields that were brought together in the seminar.

4. The remaining program put a strong emphasis on open time for ad-hoc discussions and working in groups. After the open problem session on Monday, several groups of on-site and remote participants were formed, who approached the posed problems.
5. A discord server was set up to coordinate further communication and to keep track of the progress of the working groups.
6. A social event was organized online on Tuesday evening.

Work on open problems

Following the open problem session on the first day of the workshop, spontaneous groups working on selected open problems emerged. These typically included a mix of on-site and online participants, working in either synchronous or asynchronous manner using the Discord platform as a mean of communication. Below we list a selection of directions that were pursued during the seminar.

Model-checking on interpretations of locally well-behaved structures.   It is known that model-checking First Order logic (FO) can be done in fixed-parameter time on classes of graphs that are locally well-behaved, for instance have locally bounded treewidth. However, the question is whether this is still true if the input graph is "logically disguised", or more precisely, has been additionally mapped through some FO transduction. The aim of this research group was to provide an affirmative answer by proving the following theorem: For every class of graphs $\mathcal{C}$ that is stable and can be transduced from a class of locally bounded cliquewidth, the model-checking problems for FO is fixed-parameter tractable on $\mathcal{C}$. This would generalize several known results on efficient FO model-checking on classes of dense graphs, e.g. map graphs or interpretations of classes of bounded degree, as well as provide multiple new results.

Transducing paths from classes of unbounded shrubdepth.   The emerging logically-motivated structure theory for graphs uses First-Order transductions as the main notion of embedding. It is important to understand possible duality theorems for this notion, of the following form: If a class of graphs $\mathcal{C}$ does not admit a decomposition of some form, then $\mathcal{C}$ transduces a class of specific obstacles witnessing this conclusion. The aim of this research group was to prove the most basic conjecture following this pattern: If a class of graphs $\mathcal{C}$ has unbounded shrubdepth, then $\mathcal{C}$ transduces the class of all paths.

Treedepth vs pathwidth.   It is known that every graph of pathwidth $\Omega(ab)$ has treewidth at least a or contains a binary tree of depth b as a minor. It is also known that every graph of treedepth $\Omega(abc)$ has treewidth at least a, or contains a binary tree of depth b as a minor, or contains a simple path of length 2^c. This suggests the following conjecture: every graph of treedepth $\Omega(bc)$ has either pathwidth at least b or contains a simple path of length 2^c. The aim of this research group was to resolve this conjecture.

Treewidth-twin-width.   The definition of the recently introduced graph parameter twin-width revolves around the mechanism of contraction sequences: simplification operations using which one can "fold" the whole graph into a single vertex. The main complexity measure of a contraction sequence is the maximum number of error edges that are adjacent to any vertex at any time. The goal of this group was to investigate the combinatorial properties of a graph parameter dubbed *treewidth-twin-width* obtained by additionally requiring that at all times, the graph composed of the error edges has bounded treewidth. Of particular interest is whether various classes known to have bounded twin-width actually have bounded treewidth-twin-width.

Integer programs equivalent to ones with bounded primal treedepth.   Integer programming is known to be efficiently solvable for instances with small primal or dual treedepth. While we have a relatively good understanding of conditions when the instance can be transformed to an instance with small dual treedepth, less is known in the case of primal treedepth. The aim of this research group was to relate, for a given instance of integer programming, the smallest possible primal treedepth of an equivalent instance to invariant properties of the instance itself, in particular, to the structure of the column matroid of the constraint matrix. The ultimate goal would be to design algorithms for constructing an equivalent instance with small primal treedepth while avoiding a blow up in the entry complexity (such a blow up would prevent using the existing IP algorithms to solve the constructed instance).

Acknowledgements

2 Table of Contents

3   Overview of Talks

3.1   Twin-width and Sparsity

Édouard Bonnet (ENS – Lyon, FR)

This is a tutorial on twin-width putting the focus on sparse classes. We sketch the characterization of bounded twin-width by so-called mixed minors. This is instrumental in bounding the twin-width of proper minor-closed classes; either directly, or via the functionally equivalent oriented twin-width. We see that all notions of sparsity collapse on hereditary classes of bounded twin-width. We showcase the algorithmic usefulness of a contraction sequence (witnessing low twin-width) by presenting an FPT algorithm for k-Independent Set on bounded twin-width graphs. We scale down the contraction sequences so that they exactly capture bounded rank-width and bounded linear rank-width, or in the sparse setting, bounded treewidth and bounded pathwidth. This way, the same algorithmic approach (the one presented for k-Independent Set) tackles FO model checking on a wide variety of classes, and reproves Courcelle's theorems on MSO model checking.

3.2   Lacon- and Shrub-Decompositions: Characterizing Transductions of Bounded Expansion Classes

Jan Dreier (TU Wien, AT)

This talk introduces lacon- and shrub-decompositions of graphs. We show that a graph class is a transduction of a class with bounded expansion iff it admits lacon- or shrub-decompositions with bounded expansion. This shows that on sparse graph classes, transductions are no more expressive than boolean combinations of purely existential transductions.

3.3   Combinatorial toolbox of sparsity and approximation algorithms

Zdenek Dvorák (Charles University – Prague, CZ)

We give an overview of several basic tools from the sparsity theory and their applications in the design of approximation algorithms:
- The bounds on the expansion of shallow covers and the link between polynomial expansion and sublinear separators, used to analyse local search algorithms in graph classes with sublinear separators.

- The representations of short paths, applied to the approximation of distance domination number.
- Low-treewidth covers applied to design of constant-factor approximation algorithms and PTASes for problems expressible in the first-order logic.

3.4   Logics with Invariantly Used Relations

Kord Eickmeyer (TU Darmstadt, DE)

We study the expressive power and the complexity of model-checking for logics that are enriched by invariantly used relations: formulae in these logics may speak about e.g. a linear order on the set of elements of a structure, provided that the truth value of the formula be independent of the particular choice of a linear order. Invariant access to a linear order or a successor relation strictly increases the expressive power of first-order logic, but all known separating examples are structurally very complex. We review results showing a collapse in expressive power on certain trees and structures of bounded tree-depth. As for model-checking, we show how to interpret a successor relation in a structure with a k-walk or, alternatively, a spanning tree of bounded degree. This can be used to obtain fixed-parameter tractable model-checking algorithms for successor invariant first-order logic on classes of bounded expansion.

3.5   Stable graphs of bounded twin-width

Jakub Gajarský (University of Warsaw, PL), Michal Pilipczuk (University of Warsaw, PL), and Szymon Torunczyk (University of Warsaw, PL)

We prove that every class of graphs $\mathcal{C}$ that is monadically stable and has bounded twin-width can be transduced from some class with bounded sparse twin-width. This generalizes analogous results for classes of bounded linear cliquewidth [1] and of bounded cliquewidth [2]. It also implies that monadically stable classes of bounded twin-width are linearly χ-bounded.

References

1 Jaroslav Nešetřil, Patrice Ossona de Mendez, Roman Rabinovich, and Sebastian Siebertz. *Classes of graphs with low complexity: The case of classes with bounded linear rankwidth.* Eur. J. Comb., 91:103223, 2021.
2 Jaroslav Nešetřil, Patrice Ossona de Mendez, Michał Pilipczuk, Roman Rabinovich, and Sebastian Siebertz. *Rankwidth meets stability.* In SODA 2021, pages 2014–2033. SIAM, 2021.

3.6   FO model checking of intersection graphs and twin-width

Petr Hlinený (Masaryk University – Brno, CZ) and Filip Pokrývka (Masaryk University – Brno, CZ)

We survey past research on the complexity of FO model checking on classes of intersection graphs of geometric sets (e.g., interval graphs, permutation and circle graphs, map graphs), and show how this research direction has been affected by the new notion of twin-width. In particular, Bonnet et al have shown that FO model checking of a hereditary class of permutation graphs is in FPT if and only if the class excludes some permutation graph (assuming FPT$\neq$W[1]). Inspired by that, we prove that a hereditary class of circle graphs (i.e., of intersection graphs of chords of a circle) has the FO model checking in FPT if and only if (again) the class excludes some permutation graph. We also shortly comment on a recent conjecture of Rose McCarty about FO model checking of bounded perturbations of circle graphs.

3.7   Product structure of planar graphs

Gwenaël Joret (UL – Brussels, BE)

The product structure theorem for planar graphs states that every planar graph is a subgraph of the strong product of a bounded treewidth graph (treewidth at most 8) and a path. In this tutorial, I will first sketch the proof of this theorem. Then I will give an overview of its recent applications, and I will mention some recent generalizations of the theorem to other classes of graphs. I will conclude with a number of open questions.

3.8   Treedepth and Integer Programming

Martin Koutecký (Charles University – Prague, CZ)

A breakthrough result from 2018 by Koutecký, Levin, and Onn states that integer programming can be solved in time f(a,d) poly(L), where L is the input length, a is the "numeric measure" (largest coefficient of the constraint matrix) and d is the smaller of primal/dual treedepth of the constraint matrix. We give a high-level overview of this algorithm, survey

the progress since then (strongly polynomial and near-linear algorithms, mixed integer programming, and matroid parameters), and then highlight a refinement of treedepth, so-called d-fold treedepth, which emerges in this context.

3.9 Obstructions for bounded shrub-depth and rank-depth

O-joung Kwon (Incheon National University, KR)

Shrub-depth and rank-depth are dense analogues of the tree-depth of a graph. It is well known that a graph has large tree-depth if and only if it has a long path as a subgraph. We prove an analogous statement for shrub-depth and rank-depth, which was conjectured by Hlinený, Kwon, Obdrzalek, and Ordyniak [Tree-depth and vertex-minors, European J. Combin. 2016]. Namely, we prove that a graph has large rank-depth if and only if it has a vertex-minor isomorphic to a long path. This implies that for every integer t, the class of graphs with no vertex-minor isomorphic to the path on t vertices has bounded shrub- depth.

3.10 Graph Modification in Practice

Brian Lavallee (University of Utah – Salt Lake City, US)

We describe a new framework for generalizing approximation algorithms for structural graph classes so that they apply to graphs "close" to a class (a scenario we expect is common when working with real-world networks) while still guaranteeing approximation ratios. The so-called structural rounding framework edits graphs into a nearby algorithmically tractable class, applies a class-specific approximation algorithm, then lifts the partial solution to the original graph. We give a general characterization of when an optimization problem is amenable to this approach, and show that it includes many well-studied graph problems, such as Independent Set, Vertex Cover, Minimum Maximal Matching, Chromatic Number, r-Dominating Set, and Connected Dominating Set. Guided by an experimental evaluation of the framework's efficacy for Vertex Cover on near-bipartite graphs, we propose studying bicriteria approximation hardness for editing to parameterized graph classes.

3.11 Conjectures on vertex-minors

Rose McCarty (University of Waterloo, CA)

We survey various conjectures about proper vertex-minor-closed classes. The focus is on a conjecture of Geelen which would describe the structure of graphs in a proper vertex-minor-closed class. This conjecture is analogous to the Graph Minors Structure Theorem of Robertson and Seymour, but for vertex-minors instead of minors. We discuss the main pieces of this conjecture and its relationship with other areas of sparsity.

3.12 Posets, planarity, and sparsity

Piotr Micek (Jagiellonian University – Kraków, PL)

This talk was an overview on connections between poset's dimension and the combinatorial side of the sparsity world. Already Dilworth has proved that for a poset to have large dimension, the poset must be wide. It does not have to be tall though. Indeed, so-called standard examples have height 2 and unbounded dimension. A remarkable feature of planar posets is that if they have large dimension then they are also tall. In other words, we can bound dimension of planar posets from above by a function of the height. This result by Strein and Trotter, proved in 2014, was a spark that kicked of a new line of research. We discussed the latest results in this line including how to bound the dimension of a poset in terms of the (3h)-th weak coloring number of its cover graph. This implies bounds for the planar case and far beyond in the hierarchy of sparse classes of graphs. We also discussed the current state-of-art and pointed the key open problems in the area: including a thrity year old problem by Nešetřil and Pudlák (with essentially no progress over the years) on Boolean dimension of planar posets. Another exciting line of research is dim-boundedness which is an analogue of chi-boundedness for graphs. Here we still do not know if the class of planar posets is dim-bounded.

3.13 Empirical Evaluation of Approximation Algorithms for Generalized Graph Coloring and Uniform Quasi-Wideness

Wojciech Nadara (University of Warsaw, PL)

The notions of bounded expansion and nowhere denseness not only offer robust and general definitions of uniform sparseness of graphs, they also describe the tractability boundary for several important algorithmic questions. In this paper we study two structural properties of

these graph classes that are of particular importance in this context, namely the property of having bounded generalized coloring numbers and the property of being uniformly quasi-wide. We provide experimental evaluations of several algorithms that approximate these parameters on real-world graphs. On the theoretical side, we provide a new algorithm for uniform quasi-wideness with polynomial size guarantees in graph classes of bounded expansion and show a lower bound indicating that the guarantees of this algorithm are close to optimal in graph classes with fixed excluded minor.

3.14   Sparsity in Practice – a bit of introspection

Felix Reidl (Birkbeck, University of London, GB)

After several years of working towards making sparsity-based graph algorithms feasible in practice, I present my thoughts on why certain projects seem to run on their own while others stagnate. And the reasons do not seem to lie in the algorithms themselves, rather, it is a matter of whether we develop a solver (a general-purpose machinery aimed at a broad an fuzzy problem landscape) or a solution (a software that solves a specific, tangible problem for a collaborator). In either case, however, there are many small lessons to be learned about the engineering process of sparseness-based algorithms.

3.15   Improved Bounds for Centered Coloring

Felix Schröder (TU Berlin, DE)

A vertex coloring ϕ of a graph G is p-centered if for every connected subgraph H of G either ϕ uses more than p colors on H or there is a color that appears exactly once on H. Centered colorings form one of the families of parameters that allow to capture notions of sparsity of graphs: A class of graphs has bounded expansion if and only if there is a function f such that for every $p \geq 1$, every graph in the class admits a p-centered coloring using at most $f(p)$ colors.

In the talk, we surveyed recent upper bounds for the maximum number of colors needed in a p-centered coloring of graphs from several widely studied graph classes:
1. planar graphs admit p-centered colorings with $O(p^3 \log p)$ colors;
2. bounded degree graphs admit p-centered colorings with $O(p)$ colors while it was conjectured that they may require exponential number of colors in p;
3. graphs avoiding a fixed graph as a topological minor admit p-centered colorings with a polynomial in p number of colors. All these upper bounds imply polynomial algorithms for computing the colorings. We also discussed some non-trivial lower bounds:

4. there are graphs of treewidth t that require $\binom{p+t}{t}$ colors in any p-centered coloring and this bound matches the upper bound;

5. there are planar graphs that require $\Omega(p^2 \log p)$ colors in any p-centered coloring. We briefly talked about the proof methods for the other results before diving a little deeper into how to obtain result (2) with the entropy compression method.

Participants

- Marthe Bonamy
University of Bordeaux, FR
- Édouard Bonnet
ENS – Lyon, FR
- Marcin Brianski
Jagiellonian University –
Kraków, PL
- Jan Dreier
TU Wien, AT
- Zdenek Dvorák
Charles University – Prague, CZ
- Kord Eickmeyer
TU Darmstadt, DE
- Fedor V. Fomin
University of Bergen, NO
- Jakub Gajarský
University of Warsaw, PL
- Martin Grohe
RWTH Aachen University, DE
- Meike Hatzel
TU Berlin, DE
- Petr Hliněný
Masaryk University – Brno, CZ

- Gwenaël Joret
UL – Brussels, BE
- Noleen Koehler
University of Leeds, GB
- Martin Koutecký
Charles University – Prague, CZ
- Daniel Král'
Masaryk University – Brno, CZ
- Stephan Kreutzer
TU Berlin, DE
- Nikolas Mählmann
Universität Bremen, DE
- Piotr Micek
Jagiellonian University –
Kraków, PL
- Wojciech Nadara
University of Warsaw, PL
- Jaroslav Nešetril
Charles University – Prague, CZ
- Jan Obdržálek
Masaryk University – Brno, CZ
- Patrice Ossona de Mendez
EHESS – Paris, FR

- Kristýna Pekárková
Masaryk University – Brno, CZ
- Marcin Pilipczuk
University of Warsaw, PL
- Michal Pilipczuk
University of Warsaw, PL
- Felix Schröder
TU Berlin, DE
- Sebastian Siebertz
Universität Bremen, DE
- Dimitrios M. Thilikos
University of Montpellier &
CNRS, FR
- Szymon Torunczyk
University of Warsaw, PL
- Torsten Ueckerdt
KIT – Karlsruher Institut für
Technologie, DE
- Alexandre Vigny
Universität Bremen, DE
- Bartosz Walczak
Jagiellonian University –
Krakow, PL

Remote Participants

- Saeed Amiri
Universität Köln, DE
- Daniel Cranston
Virginia Commonwealth
University – Richmond, US
- Anuj Dawar
University of Cambridge, GB
- Peter Gartland
University of California –
Santa Barbara, US
- Archontia C. Giannopoulou
University of Athens, GR
- Mario Grobler
Universität Bremen, DE
- Shweta Jain
University of Utah –
Salt Lake City, US
- Ken-ichi Kawarabayashi
National Institute of Informatics –
Tokyo, JP
- O-joung Kwon
Incheon National University, KR
- Aurelie Lagoutte
Université Clermont Auvergne –
Aubiere, FR

- Brian Lavallee
University of Utah –
Salt Lake City, US
- Rose McCarty
University of Waterloo, CA
- Yosuke Mizutani
University of Utah –
Salt Lake City, US
- Sang-il Oum
IBS – Daejeon, KR
- Filip Pokrývka
Masaryk University – Brno, CZ
- Evangelos Protopapas
University of Montpellier &
CNRS, FR
- Wojciech Przybyszewski
University of Warsaw, PL
- Daniel Quiroz
Universidad de Valparaiso, CL
- Felix Reidl
Birkbeck, University of
London, GB
- Benjamin Rossman
Duke University – Durham, US

- Saket Saurabh
The Institute of Mathematical
Sciences – Chennai, IN
- Nicole Schirrmacher
Universität Bremen, DE
- Nicole Schweikardt
HU Berlin, DE
- Luc Segoufin
INRIA & ENS Paris, FR
- Marek Sokolowski
University of Warsaw, PL
- Giannos Stamoulis
University of Montpellier 2, FR
- Blair D. Sullivan
University of Utah –
Salt Lake City, US
- Nate Veldt
Texas A&M University – College
Station, US
- David Wood
Monash University –
Clayton, AU