DAGSTUHL
REPORTS

Volume 9, Issue 10, October 2019

Social Agents for Teamwork and Group Interactions (Dagstuhl Seminar 19411)
Elisabeth André, Ana Paiva, Julie Shah, and Selma Šabanovic 1

Quantum Cryptanalysis (Dagstuhl Seminar 19421)
Michele Mosca, Maria Naya-Plasencia, and Rainer Steinwandt 47

Theory of Randomized Optimization Heuristics (Dagstuhl Seminar 19431)
Carola Doerr, Carlos M. Fonseca, Tobias Friedrich, and Xin Yao 61

Analysis of Autonomous Mobile Collectives in Complex Physical Environments (Dagstuhl Seminar 19432)
Mario Gleirscher, Anne E. Haxthausen, Martin Leucker, and Sven Linker 95

Programming Languages for Distributed Systems and Distributed Data Management (Dagstuhl Seminar 19442)
Carla Ferreira, Philipp Haller, and Guido Salvaneschi 117

Algorithms and Complexity in Phylogenetics (Dagstuhl Seminar 19443)
Magnus Bordewich, Britta Dorn, Simone Linz, and Rolf Niedermeier 134

Dagstuhl Reports, Vol. 9, Issue 10 ISSN 2192-5283

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
http://www.dagstuhl.de/dagpub/2192-5283

Publication date
March, 2020

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at `http://dnb.d-nb.de`.

Digital Object Identifier: 10.4230/DagRep.9.10.i

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.

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

Report from Dagstuhl Seminar 19411

Social Agents for Teamwork and Group Interactions

Edited by

Elisabeth André[1], Ana Paiva[2], Julie Shah[3], and Selma Šabanovic[4]

1 **Universität Augsburg, DE,** andre@informatik.uni-augsburg.de
2 **INESC-ID – Porto Salvo, PT,** ana.paiva@inesc-id.pt
3 **MIT – Cambridge, US,** julie_a_shah@csail.mit.edu
4 **Indiana University, US,** selmas@indiana.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19411 "Social Agents for Teamwork and Group Interactions". It summarises the three talks that were held during the seminar on three different perspectives: the impact of robots in human teamwork, mechanisms to support group interactions in virtual settings, and affect analysis in human-robot group settings. It also details the considerations of six working groups covering the following topics: datasets, design, team dynamics, social cognition, scenarios, and social behaviours.

Seminar October 6–11, 2019 – http://www.dagstuhl.de/19411

2012 ACM Subject Classification Human-centered computing → Collaborative interaction, Human-centered computing → Interaction design, Human-centered computing → User studies, Computing methodologies → Cognitive robotics

Keywords and phrases Social Agents, Social Robotics, Multi-Agent Systems, Human-Agent Interaction, Groups and Teams

Digital Object Identifier 10.4230/DagRep.9.10.1

Edited in cooperation with Filipa Correia and Sarah Strohkorb Sebo

1 Executive Summary

Elisabeth André (Universität Augsburg, DE, andre@informatik.uni-augsburg.de)
Ana Paiva (INESC-ID & University of Lisbon, PT, ana.paiva@inesc-id.pt)
Julie Shah (Massachusetts Institute of Technology, US, julie_a_shah@csail.mit.edu)
Selma Šabanovic (Indiana University, US, selmas@indiana.edu)

As artificial agents and social robots become more prominent in our lives, they will also increasingly become parts of the groups and teams in which people spend much of their time. The objective of this Dagstuhl Seminar was to explore and discuss theories, methods, and techniques for building embodied social agents (including robots) that can operate in groups as members of a mixed team consisting of humans and agents. Recent advances in AI, and particularly in conversational agents, are likely to lead to an increased placement of agents in groups, covering a variety of application scenarios including healthcare, education, the workplace, and the home. Platforms such as Amazon Echo, Google Home, and new social robots such as Nao, Pepper, and Aibo facilitate such placement. Studies with robots in open-ended environments, including homes and public spaces, also suggest that people often engage with robots in such contexts in groups, rather than just individually. Yet, existing research on human-agent interaction and human-robot interaction so far focuses mostly on one-on-one interactions between a human and a social agent. To stimulate growing research in settings where one or more humans interact with multiple agents or robots, this seminar focused on human-agent communication, interaction, and teamwork in groups. As such, we discussed how agents shape the dynamics of groups, how agents and robots are able to perceive other members of a group and how they relate to each other, and how to move from one-to-one interactions to multi-party interactions of agents and humans in groups and teams. By bringing together researchers from different communities, such as human-robot interaction, multi-agent systems, social psychology, and organizational studies, we aim to generate common ground and new approaches in this interdisciplinary area. While this new domain of inquiry relies on existing research at the intersection between AI, robotics, and the social sciences, our aim is to highlight open questions that current work has not sufficiently addressed.

2 Table of Contents

Executive Summary
Elisabeth André, Ana Paiva, Julie Shah, and Selma Šabanovic 1

Overview of Talks

Teamwork with Robots
Malte Jung 4

Supporting Interactions in Online Groups
Kobi Gal 5

Affect and Personality Analysis in Human-Human-Robot Interaction Settings
Hatice Gunes 5

Working Groups

Working Group on Datasets with Humans and Agents in Groups
Patrícia Alves-Oliveira, Elisabeth André, Tony Belpaeme, Filipa Correia, Kobi Gal, James Kennedy, Iolanda Leite, Catharine Oertel, André Pereira 7

Working Group on Design
Cindy L. Bethel, Kobi Gal, Michio Okada, André Pereira, Samuel Mascarenhas . . 14

Working Group on Human-Agent Team Dynamics
Filipa Correia, Hatice Gunes, Stefan Kopp, Iolanda Leite, Mark Neerincx, Catharine Oertel, Rui Prada, Marynel Vazquez 18

Working Group on Social Cognition for Robots and Virtual Agents
Cindy L. Bethel, Merijn Bruijnes, Malte Jung, Samuel Mascarenhas, Christoforos Mavrogiannis, Simon Parsons, Catherine Pelachaud, Rui Prada, Laurel Riek, Sarah Strohkorb Sebo, Julie Shah, Elaine Short, Marynel Vazquez 21

Working Group on Scenarios for Human-Agent and Human-Robot Groups
Patrícia Alves-Oliveira, Mary Ellen Foster, Hatice Gunes, Sviatlana Höhn, Franziska Kirstein, Andrea Marrella, Mark Neerincx, Michio Okada, Ana Paiva, Elisabeth Sklar, Hannes Högni Vilhjálmsson 37

Working Group on Social Behaviours for Group Interactions between Humans and Social Agents and Robots
Elisabeth André, Dirk Heylen, James Kennedy, Ana Paiva, Catherine Pelachaud, Hannes Högni Vilhjálmsson 42

Participants 46

3 Overview of Talks

3.1 Teamwork with Robots

Malte Jung (Cornell University, US, mfj28@cornell.edu)

Research on human-robot interaction (HRI) to date has largely focused on examining a single human interacting with a single robot. This work has led to advances in fundamental understanding about the psychology of HRI (e.g. how specific design choices affect interactions with and attitudes towards robots) and about the effective design of HRI (e,g. how novel mechanisms or computational tools can be used to improve HRI). However, the single-robot-single-human focus of this growing body of work stands in stark contrast to the complex social contexts in which robots are increasingly placed. While robots increasingly support teamwork across a wide range of settings covering search and rescue missions, minimally invasive surgeries, space exploration missions, or manufacturing, we have limited understanding of how groups people will interact with robots and how robots will affect how people interact with each other in groups and teams. In this talk I present empirical findings from several studies that show how robots can shape in direct, but also subtle ways how people interact and collaborate with each other in teams.

References

1 Malte F Jung. Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(3):18, 2016.
2 Malte F Jung. Affective grounding in human-robot interaction. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 263–273. IEEE, 2017.
3 Malte F Jung, Dominic DiFranzo, Brett Stoll, Solace Shen, Austin Lawrence, and Houston Claure. Robot assisted tower construction-a resource distribution task to study human-robot collaboration and interaction with groups of people. *arXiv preprint arXiv:1812.09548*, 2018.
4 Nikolas Martelaro, Malte Jung, and Pamela Hinds. Using robots to moderate team conflict: The case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 271–271. ACM, 2015.
5 Solace Shen, Petr Slovak, and Malte F Jung. Stop. i see a conflict happening.: A robot mediator for young children's interpersonal conflict resolution. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 69–77. ACM, 2018.
6 Hamish Tennent, Solace Shen, and Malte Jung. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 133–142. IEEE, 2019.

3.2 Supporting Interactions in Online Groups

Kobi Gal (Ben-Gurion University, IL, kobig@bgu.ac.il)

Advances in network technologies and interface design are enabling group activities of varying complexities to be carried out, in whole or in part, over the internet (e.g., citizen science, Massive Online Open Courses (MOOC) and questions-and-answers sites). The need to support these highly diverse interactions brings new and significant challenges to AI; how to design efficient representations for describing online group interactions; how to provide incentives that keep participants motivated and productive; and how to provide useful, non-intrusive information to system designers to help them decide whether and how to intervene with the group's work. I describe two ongoing projects that address these challenges in the wild, and discuss the potential impact of this work to environment design.

References

1 Avi Segal, Kobi Gal, Ece Kamar, Eric Horvitz, and Grant Miller. Optimizing interventions via offline policy evaluation: Studies in citizen science. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
2 Avi Segal, Kobi Gal, Ece Kamar, Eric Horvitz, Alex Bower, and Grant Miller. Intervention Strategies for Increasing Engagement in Volunteer-Based Crowdsourcing. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2016.
3 Avi Segal, Ya'akov Kobi Gal, Robert J Simpson, Victoria Victoria Homsy, Mark Hartswood, Kevin R Page, and Marina Jirotka. Improving productivity in citizen science through controlled intervention. In *Proceedings of the 24th International Conference on World Wide Web*, pages 331–337. ACM, 2015.

3.3 Affect and Personality Analysis in Human-Human-Robot Interaction Settings

Hatice Gunes (University of Cambridge, GB, hatice.gunes@cl.cam.ac.uk)

Designing intelligent systems and interfaces with socio-emotional skills is a challenging task. Past works have mainly focussed on automatically analysing expressions, affect and personality of people in individual settings. However, when we move from single user settings to multi-user and group ones, the process of affect analysis calls for new definitions, new datasets with meaningful annotations, and appropriate feature extraction and classification mechanisms in space and time. This talk questions some of the initial assumptions made in this area, and presents an overview of the works we have undertaken in recent years in human-human and human-human-robot interaction settings.

Firstly, the talk presents a set of experiments for affect analysis of subjects in group settings. These individuals were recorded watching videos alone and watching videos as part of a group [3, 4]. Our results show that: 1) facial appearance representation (i.e., the proposed Volume Quantized Local Zernike Moments Fisher Vector) outperforms other unimodal features in affect analysis in both settings; 2) temporal learning models perform better than the static learning models; 3) it is possible to predict the context, i.e., whether a person is alone or in-a-group, using their non-verbal behavioural cues; 4) people in the same group share similarities in facial behaviours which contributes to automatic affect prediction; and 5) when the expressive behaviour of one subject in a group setting is not available, behaviours expressed by the other subject(s) can be used for affect prediction [4].

Secondly, the talk introduces a novel dataset we have collected, the Multimodal Human-Human-Robot-Interactions (MHHRI) dataset [2], acquired with the aim of studying personality simultaneously in human-human interactions (HHI) and human-human-robot interactions (HRI) and its relationship with engagement. Multimodal data was collected during a controlled interaction study where dyadic interactions between two human participants and triadic interactions between two human participants and a robot took place with interactants asking/answering a set of personal questions. Interactions were recorded using two static and two dynamic cameras as well as two biosensors, and meta-data was collected by having participants to fill in two types of questionnaires, for assessing their own personality traits and their perceived engagement with their partners (self-annotated labels) and for assessing personality traits of the other participants partaking in the study (acquaintance labels).

Thirdly, using the MHHRI dataset, the talk introduces a number of experiments we have conducted for automatic prediction of personality and engagement. We analyse interactions with the robot from the viewpoint of human participants through an ego-centric camera placed on their forehead [1]. We focus on human participants' and robot's personalities and their impact on the human-robot interactions. We automatically extract nonverbal cues (e.g., head movement) from first-person perspective and explore the relationship of nonverbal cues with participants' self-reported personality and their interaction experience. We generate two types of behaviours for the robot (i.e., extroverted vs. introverted) and examine how robot's personality and behaviour affect the findings. Significant correlations are obtained between the extroversion and agreeable-ness traits of the participants and the perceived enjoyment with the extroverted robot. Plausible relationships are also found between the measures of interaction experience and personality and the first-person vision features [1].

Finally, using the MHHRI dataset, the talk introduces work that focuses on the automatic analysis and classification of engagement based on humans' and robot's personality profiles in the triadic human-human-robot interaction setting [5]. More explicitly, the study investigates how participants' personalities can be used together with the robot's personality to predict the engagement state of each participant as well as the engagement of the overall group. The fully automatic system is first trained to predict the Big Five personality traits of each participant by extracting individual and interpersonal features from their nonverbal behavioural cues. Then the output of the personality prediction system is used as an input to the engagement classification system. Third, we focus on the concept of "group engagement", which we define as the collective engagement of the participants with the robot, and analyse the impact of similar and dissimilar personalities on the engagement classification. Our experimental results show that: 1) using the automatically predicted personality labels for engagement classification yields an F-measure on par with using the manually annotated personality labels, demonstrating the effectiveness of the automatic personality prediction module proposed; 2) using the individual and interpersonal features without utilizing personality information is not sufficient for engagement classification, instead incorporating the participants and robots personalities with individual/interpersonal features increases engagement classification performance; and 3) the best classification performance is achieved when the participants and robot are extroverted, while the worst results are obtained when all are introverted [5].

References

1 Oya Celiktutan and Hatice Gunes. Computational analysis of human-robot interactions through first-person vision: Personality and interaction experience. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 815–820. IEEE, 2015.

2 Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing*, 2017.

3 Wenxuan Mou, Hatice Gunes, and Ioannis Patras. Alone versus in-a-group: A multi-modal framework for automatic affect recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(2):47, 2019.

4 Wenxuan Mou, Hatice Gunes, and Ioannis Patras. Your fellows matter: Affect analysis across subjects in group videos. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–5. IEEE, 2019.

5 Hanan Salam, Oya Celiktutan, Isabelle Hupont, Hatice Gunes, and Mohamed Chetouani. Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access*, 5:705–721, 2016.

4 Working Groups

4.1 Working Group on Datasets with Humans and Agents in Groups

Patrícia Alves-Oliveira (ISCTE-IUL & INESC-ID, PT & Cornell University, US, patricia_alves_oliveira@iscte-iul.pt)
Elisabeth André (Universität Augsburg, DE, andre@informatik.uni-augsburg.de)
Tony Belpaeme (Ghent University, BE & University of Plymouth, UK, tony.belpaeme@ugent.be)
Filipa Correia (INESC-ID & University of Lisbon, PT, filipacorreia@tecnico.ulisboa.pt)
Kobi Gal (Ben Gurion University, IL, kobig@bgu.ac.il)
James Kennedy (Disney Research, US, james.r.kennedy@disney.com)
Iolanda Leite (KTH Royal Institute of Technology, SE, iolanda@kth.se)
Catharine Oertel (TU Delft, NL, C.R.M.M.Oertel@tudelft.nl)
André Pereira (KTH Royal Institute of Technology, SE, atap@kth.se)

Data is, to put it simply, another word for information. Recent years have seen a growth in datasets capturing data relevant for group interactions. While data was predominantly collected to form a quantitative understanding of social interactions, the recent attention for machine learning has resulted in datasets that contain features and annotations which are aimed at classifying or predicting social behavior. Due to the data-hungry nature of machine learning and the low cost of storage, the size of typical datasets is now several orders of magnitude larger than a decade ago.

We describe what constitutes a good dataset, and provide a number of resources related to datasets with a specific focus on group interactions, either between human-human, or human-agent groups. We provide dimensions upon which datasets for group interactions involving social agents could be characterized and classify existing datasets accordingly. Such a resource would be helpful for the research community to quickly identify data to work with, or identify gaps in available data. To conclude, we discuss several challenges in producing and annotating datasets, with a focus specifically on aspects pertinent for groups.

4.1.1 Where to find and deposit datasets

While datasets used to be stored on personal repositories, there is now an opportunity to host data on more persistent and curated sites. There are many online data repositories and as a researcher, care needs to be taken in choosing which repository offers the best conditions for your data. The most popular repositories are listed below, in order of relevance to the study of group interactions.

- **Zenodo** (https://zenodo.org/) is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN, using the same systems to store data from the Large Hadron Collider. Datasets are given a digital object identifier (DOI), making it easily citable, and works together with GitHub allowing code and data to coexist.
- **Open Science Framework** (https://osf.io/) is run by the US non-profit organisation Center for Open Science, which facilitates open collaboration in science research. It was created in response to the replication crisis in psychology. Data and code can be stored in OSF Storage, the OSF public repository.
- **FigShare** (https://figshare.com/) started out as an online statistical figure repository, allowing authors to share high-resolution versions of printed material which can be interactively explored. Today it is an open access repository for a range of research outputs, including figures, reports, datasets, images, and videos. All deposited material is given a DOI. It is owned by Digital Science, a technology company based in the UK.
- **Github** (https://github.com/) started out as a versioning control system for computer code, but has increasingly been used to store datasets. Often datasets are stored in combination with code to analyse or report on the data.
- **Linguistic Data Consortium** (https://www.kaggle.com/) is an open consortium of universities, companies and government research laboratories. It stores data related to language, such as audio recordings of speech and text corpora, and is often used for Natural Language Processing and machine learning. Dryad Digital Repository (datadryad.org) is a repository with a focus on medical sciences. It charges a fee for submitting data, but guarantees free access for academic purposes.
- **Kaggle** (https://www.kaggle.com/) stores a wide range of datasets aimed predominantly at data science and machine learning. Many datasets serve machine learning competitions or challenges, and have been released by commercial organisation looking for new ways to capitalise on data they hold. The website allows online data science exploration, offers cloud computing, serves as a code repository and learning centre for data scence and machine learning. Kaggle is owned by Google.
- **IDIAP Data Distribution Portal** (https://www.idiap.ch/dataset) is hosted by the IDIAP Research Institute in Switzerland. It holds a small collection of data, mainly aimed at machine learning. It is unclear if and how new data can be added.

A record of all data repositories for scientific research is maintained at Re3Data (https://www.re3data.org/), which lists and tracks anything from single datasets to large data repositories. Data On The Mind (http://www.dataonthemind.org/) holds a modest list of datasets spread over the web. A promising development, with the potential to grow into a powerful tool for researchers, is Google's Dataset Search (https://toolbox.google.com/datasetsearch). It aggregates data collections from across the web, but at the moment does not have a function to return only results useful for particular purposes, such as machine learning or scientific analysis.

4.1.2 Available datasets for group interactions

Human-human datasets are important for the development of social agents in groups and teams. They can provide insights for understanding human behavior, be used to derive agent behavior (e.g., through machine learning techniques), or be used in evaluating algorithms. Many datasets that contain human-human interactions are publicly available. There is also good variability in terms of the activity and settings in which they have been collected.

The availability of human group interaction datasets partially arises from multiple communities that have a focused problem that they wish to solve. For example, the Emotion Recognition in the Wild Challenge 2019, which has a focus on group cohesion prediction (https://sites.google.com/view/emotiw2019). Additionally, datasets receive more attention in some communities, to the extent of having dedicated conferences (e.g., the International Conference on Language Resources and Evaluation – LREC).

In contrast, there are few datasets currently available that contain group interactions with artificial agents. This is likely in part due to the greater quantity of work that has been conducted in dyadic scenarios with artificial agents, but could also be due to the greater diversity of goals for research involving agents in social groups. Creating challenges for the community could provide impetus for collecting and sharing datasets for specific problems. There would subsequently be the possibility of reusing these datasets for solving other issues.

Furthermore, while many research institutes run empirical studies involving agents and humans in groups, few of these studies are captured and shared as datasets. This additional step poses many significant challenges (see 4.1.6), but could also provide a great source of data for the community.

Group datasets with one agent or robot

This section provides all of the publicly available group datasets including at least one agent that we are aware of:

- The Vernissage Dataset (http://vernissage.humavips.eu/). This dataset includes human-robot interactions with multiple participants and the commonly used robot platform NAO. It was collected using a Wizard-of-Oz protocol and has multiple camera views, audio streams, robot behavior logs, and some annotations.
- UE-HRI dataset (https://www.tsi.telecom-paristech.fr/aao/en/2017/05/18/ue-hri-dataset). The UE-HRI dataset consists of recordings of humans interacting with the social robot Pepper. It has spontaneous interactions in a naturalistic setting. There are a mixture of dyadic and triadic interactions. A variety of sensor data is available including video, audio, depth, sonar, laser and user touch inputs.

Group datasets with only humans

This section provides an overview of some human group datasets. The list here is by no means comprehensive, but is used to give some insight into the types of datasets that are publicly available and how they have been used in research:

- Elea Dataset (https://www.idiap.ch/dataset/elea).
  The aim of the Elea corpus was to create a resource to study group interaction. It is a multi modal corpus featuring both audio and video data. Annotations of both gaze and

voice activity among others are available. The corpus is particularly suited for studying group dynamics in terms of group performance measures.

- AMI (http://groups.inf.ed.ac.uk/ami/download/). The AMI corpus is one of the largest openly available corpora. It is a multi modal meeting corpus. Both audio and video data is available as well as some annotations in gaze and voice activity. The corpus has been used amongst others to study dominance.
- WOLF (https://www.idiap.ch/dataset/wolf). The WOLF corpus is based on a role-playing game where some people take on the roles of werewolves or villagers. The game is designed around deception, which has also been the research question most commonly addressed with this corpus. It is a multi modal corpus including both audio and video data.

4.1.3 Dataset dimensions

Datasets should clearly state the characteristics and procedures that were utilized during the data collection. The following list tries to specify several dimensions to describe the content of the dataset. Examples of existing datasets using these dimensions are present in Table 1.

1. **Resource name**: State the full name for newly created resources, followed by the acronym, if any.
2. **Size of the dataset**: Put the size of your resource on the basis of the group and data amount (interactions, time length, group sizes, and make-up).
3. **Demographics**: Describe the participants present in your dataset (e.g. “Newborns”, “Children”, “Teenagers”, “Adults”, “Elderly”, “Mixed”).
4. **Activity**: Explain what people are doing (e.g. playing a game, watching a movie together, ...) and the type of agent that is being used.
5. **Languages**: State the language spoken and/or read in the dataset (e.g., English, German, Japanese, none).
6. **Modalities**: Choose an appropriate label or combination of labels for describing the recorded data: “Visual”, “Audio”, “Physiological”, ...
7. **Annotation**: Describe how and if data was labeled by using: “Human Labelled”, “Automatic Labelled”, “Not Labelled”.
8. **Availability**: State the availability of the resource for the community. If the dataset is available on the web, at least for research (“Open”); if data is associated to an institution (“From Data Centers”); dataset distributed directly by the owner, usually associated with informed consent restrictions (“From Owner”); or other (“Other”).
9. **Setting**: Describe where the dataset was collected (e.g. in the lab, public open space, school, etc.).
10. **Resource production status**: State is the resource already existed or if it was newly created. For newly created resources, describe if the production is completed (“Complete”) or if work is still in progress (“Work-in-Progress”). In the case of an existing resource, describe whether it has been simply used (“Existing-used”) or you have updated or modified it (“Existing-updated”).
11. **URL/DOI/Publication** (if available): Indicate the URL of the resource/tool/guidelines described, if it exists, including the URL of the resource documentation if available.

4.1.4 Annotation tools

The following list contains examples of tools that can be used to perform the annotation and the statistical analysis, in some cases, of the dataset.

Table 1 Dataset dimensions for Humans and Agents in Groups.

Resource Name	Size	Demographics	Activity	Languages	Modalities	Annotations	Availability	Setting	Status
Vernissage	13 sessions (about 11 minutes each) of NAO interacting with two persons	Adults	Quiz activity about art and culture	English	Audio Video Mocap Robot logs	Speech transcriptions and several nonverbal cues such as 2D head-location, nodding, visual focus of attention (VFOA) and addressees.	Open	Lab	Complete
UE-HRI	54 interactions Aprox. 9 hours 1 agent, 1 or more humans	Adults	Social chit-chat	French?	Audio Video Depth Sonar Laser UI input Robot logs	Engagement	Open	Public space – university hallway	Partially annotated

- ELAN https://tla.mpi.nl/tools/tla-tools/elan/
- Noldus ObserverXT https://www.noldus.com
- NOVA https://github.com/hcmlab/nova – Nonverbal Analyzer is a tool for annotating and analyzing behaviours in social interactions. It supports Annotators using Machine Learning already during the coding process. Further it features both, discrete labels and continuous scores and a visualization of streams recorded with the SSI Framework

4.1.5 Feature extraction tools / feature set

The following list contains examples of tools that automatically detect some of the features commonly used in the annotation of human behaviour.

- OpenSmile (audio features) https://www.audeering.com/opensmile/
- EmoVoice (audio features) https://github.com/hcmlab/emovoice
- GeMAPS: Geneva Minimalist Acoustic Parameter Set [5]
- OpenFace https://github.com/TadasBaltrusaitis/OpenFace
- OpenPose https://github.com/CMU-Perceptual-Computing-Lab/openpose

4.1.6 Open Challenges

Datasets gain value from being shared, but data on social interactions will always contain data collected by recording people. There are formidable challenges facing dataset collection and dissemination:

- **Logistics:** Collecting data with groups is challenging for practical reasons. If our aim is to have a fixed group size, recruiting participants and ensuring that the same group size number in all sessions can be difficult. This is especially challenging when considering repeated interactions between the agent an the same group of users.
- **Annotation:** Group interactions are inherently more complex than dyadic interactions which means that the collected data and consequent annotations can become more noisy. For example, many tools for automatic feature extraction/annotation do not necessarily support multiple users (or only consider a limited set of users), and even human-annotated labels can become more subjective (e.g. coder agreement might decrease as the number of participants increase).
- **Gold-Standard:** The interpretation of nonverbal data is highly subjective. Thus the question arises of how to get a golden standard. Should it rely on the assessment of

(multiple) observers/annotators or on the assessment of the human group members involved in the interaction with the robot(s)? In general, obtaining gold standard data requires significant human effort.

- **Generalizability:** Datasets are collected in heterogeneous settings (see above). Can we transfer findings from one setting to the other? For example, can we transfer findings from a setting with two robots and one human to a setting with three robots and two humans? Can we generalize data from agents with different embodiments (e.g., robots vs virtual agents)?
- **Privacy:** There is an increasing understanding of the importance to share data to facilitate a comparison of algorithms and approaches. On the other hand, there are also increasing awareness to privacy, which constraints the use of data. Given recent initiatives to protect citizens, such as the European General Data Protection Regulation (GDPR)[1] or the Children's Online Privacy Protection Act (COPPA)[2] in the USA, the storage and use of data are now heavily regulated.
- **Persistence/Access:** Terms and conditions of the data collection require a legal support and ethical revision. Regarding the data collection of human behavior in group settings, one of the main challenges is to grant access to any participant without exposing the data of another participant. A possible solution is to apply anonymization tools (e.g. [1]) that would allow: (1) to grant participants their right to access their own data without revealing the identity of other participants; (2) the extraction and labeling of behaviours without revealing the identity of participants. Another challenge is the actual persistence and storage of the data and the fact that the withdrawal of one participation may compromise the data of other participant(s).
  In addition to existing previously mentioned challenges, terms and conditions and written consent are often worded to support the immediate goal of the study or technology for which the data was collected, but are not necessarily drawn up to support future use by others. For example, a consent form can stipulate that "data can be used by the research team for future analyses", but this wording limits future use only to the research team affiliated with the institution who collected the data and does not allow non-affiliates to use the data. Also, the insistence that data can only be used for academic purposes often sits in the way of current developments in the industry, where commercial software, often hosted on cloud services, is used to analyze data, thereby blurring the distinction between academic and commercial access to the data. Another issue is that it may be technically difficult to collect consent from every individual that generates data.

4.1.7 What makes a good dataset?

There is a huge proliferation of available datasets for designing and testing machine learning algorithms. Most of these datasets are collected and designed to solve a small defined machine learning problem and too specific to extract principles for humans and agents in groups. However, some recommendations for designing good datasets in the machine learning community can be useful as a basis for datasets that have been collected for studying human/agent group interactions.

Datasets should contain examples of research questions that can be answered or studied by using the data. Ideally, it should enable the community to work on their own research

[1] https://en.wikipedia.org/wiki/General_Data_Protection_Regulation
[2] https://en.wikipedia.org/wiki/Children%27s_Online_Privacy_Protection_Act

questions or models by using the same data. Most datasets extracted from real-life applications contain noise, missing values or irrelevant data. A good dataset should be cleaned from these problems before it is shared. This makes the dataset easy to work with and new contributors or users do not have to repeat this process again. Data pre-processing should also be performed to optimize the data for visualization (e.g., a txt file for Elan) or machine learning (e.g., data normalization) purposes.

The annotation schemes and features contained in the dataset should be clearly labelled and motivated by research in human-agent group interactions. All behaviors that contribute to understanding group behavior in human-agent group interactions should be annotated. This includes both extracting important features from the humans in the interaction but also logging all of the important timings and behaviors from the intelligent agents. Finally, to increase the usefulness and reusability of these datasets, they should be collected in settings that promote authentic and natural group interactions.

4.1.8 Conclusions and future challenges

Datasets are one of the most cited research outcomes, demonstrating the potential of sharing data with the community (e.g., [4, 8, 6]). Several fields of research have already acknowledged this impact and are actively contributing to data sharing. For example, the International Conference on Language Resources and Evaluation encourages and publishes datasets as a central part of conference contributions; MediaEval[3], a benchmarking initiative dedicated to evaluating new algorithms for multimedia access and retrieval, makes their datasets publicly available after work is concluded; the AVEC conference (Audio-Visual Emotion Challenge) and the Interspeech Computational Paralinguistics Challenges (ComPaRe) has challenges that release the associated datasets in open-access.

Despite the obvious benefits of dataset sharing, this culture have not yet been fully adopted by the field of Humans and Agents in Groups. Many factors are associated with the lack of dataset sharing culture among this field of research. Firstly, there is no venue that specifically values dataset sharing and publication. Secondly, researchers face many challenges when facing the option to share data, mainly because this field collects data from human participants which are associated with with legal, ethical, and privacy policies and restrictions. However, there are several benefits associated with sharing data. Specifically, sharing data advances the pace of research. Usually, in the Humans and Agents in Groups field, researchers collect new data for every new study performed. This requires extra research time due to the recruitment of participants and data collection. By adopting a dataset sharing culture, studies can be performed by re-using already existing materials and resources and analysis can be performed with already collected data. Additionally, research quality and transparency increases since analysis can be performed with already existing data, supporting reproducibility in this domain of research [7].

We present in this document the first attempt to share datasets within the field of Humans and Agents in Groups. We have defined best policies for datasets by defining what a good dataset is, what are the challenges associated with dataset sharing, and proposed interesting solutions to attend those. A major contribution of this work was to state the characteristics of the datasets since researchers can use the defined dimensions to share their datasets and also will ease information retrieval. Future directions to stimulate dataset sharing among this research field includes the organization of a workshop on datasharing (possible venues: HRI, CHI, CSCW), and the creation of challenges or competitions (similarly to DARPA or a simple competition in Kaggle).

[3] MediEval dataset: http://www.multimediaeval.org/datasets/

References

1 Pascal Birnstill, Daoyuan Ren, and Jürgen Beyerer. A user study on anonymization techniques for smart video surveillance. In *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2015.

2 Rupert Brown and Samuel Pehrson. *Group processes: Dynamics within and between groups.* John Wiley & Sons, 2019.

3 Ewart J. de Visser, Marieke M. M. Peeters, Malte F. Jung, Spencer Kohn, Tyler H. Shaw, Richard Pak, and Mark A. Neerincx. Towards a theory of longitudinal trust calibration in human-robot teams. *Int. Journal of Social Robotics*, in press.

4 Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

5 Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2):190–202, 2015.

6 Nils Petter Gleditsch, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. Armed conflict 1946-2001: A new dataset. *Journal of peace research*, 39(5):615–637, 2002.

7 Steven N Goodman, Daniele Fanelli, and John PA Ioannidis. What does research reproducibility mean? *Science translational medicine*, 8(341):341ps12–341ps12, 2016.

8 Brian McFee, Thierry Bertin-Mahieux, Daniel PW Ellis, and Gert RG Lanckriet. The million song dataset challenge. In *Proceedings of the 21st International Conference on World Wide Web*, pages 909–916. ACM, 2012.

9 Jurriaan Van Diggelen, Mark Neerincx, Marieke Peeters, and Jan Maarten Schraagen. Developing effective and resilient human-agent teamwork using team design patterns. *IEEE intelligent systems*, 34(2):15–24, 2018.

4.2 Working Group on Design

Cindy L. Bethel (Mississippi State University, US, cbethel@cse.msstate.edu)
Kobi Gal (Ben-Gurion University, IL, kobig@bgu.ac.il)
Michio Okada (Toyohashi University of Technology, JP, okada@tut.jp)
André Pereira (KTH Royal Institute of Technology, SE, atap@kth.se)
Samuel Mascarenhas (INESC-ID & University of Lisbon, PT, samuel.mascarenhas@gaips.inesc-id.pt)

4.2.1 Introduction

Design is a cycle based on the target: redesign and evolution. It can be applied in several stages of the development of human-agent teamwork. We address different perspectives where design considerations are relevant to create of multi-party settings for humans and agents.

4.2.2 Interactive Design Ideas

We identified three different categories where design principles can guide decisions to successfully implement human-agent teamwork: the target, the levers, and the application domain.

Target

One category is the target which reflects the focus on the goals and the metrics to evaluate and optimize the system. Examples of targets include:

- Ethical Interactions (e.g. fairness, interpretability), which constrains the design space for each of the other areas considered
- Partnerships/Teammates (e.g. best representations, composition)
- Goals (e.g. education, help, companionship, entertainment, entertainment, work collaboration, persuasion)

Levers

Another category is the levers, which are the factors that can be controlled. Examples of levers include:
- Form Factor/Embodiment
- Actions/Behaviors
- Environment
  - Incentives
  - Visability
  - Communication protocols
  - Social Setting (e.g., collaborative, competitive)

Application Domain

Another category is the application domain, which includes the aspects related to the task and physical space where the agent will operate. The application domain both informs the physical environment design (e.g., Amazon Warehouse) and also informs the social environment of who they are interacting with. Examples of application domains include:
- Manufacturing
- Entertainment
- Education
- Assistance (e.g., physical or social)
- Healthcare
- First Responder (e.g., law enforcement, fire, search & rescue)

4.2.3 Design principles for robots and agents to interact in groups and teams

From the website 'The Undercover Recruiter'[4], design criteria for successful team interactions include:
- They communicate well with each other.
- They focus on goals and results.
- Everyone contributes their fair share.
- They offer each other support.
- Team members are diverse.
- Good leadership.
- They are organised.
- They have fun.

[4] https://theundercoverrecruiter.com/qualities-successful-work-team/

Inspired by other design principles (e.g., Dieter Rams[5] and Yves Béhar[6]), we propose the following design principles for robots and agents to interact in groups and teams:

1. Good design for agents in teams should support fairness and ethical interactions.
2. Good design for agents in teams should consider the emotions and reactions of its members.
3. Good design for agents in teams should optimize:
   - Performance
   - Communication (e.g. support, feedback)
   - Affective signals (e.g.,team motivation, enjoyment)
   - Interpretability and simplicity of interactions
   - Organization and structure
   - Role allocation and dynamic changes to roles
   - Diversity (e.g., cultures, gender)
   - Support and feedback
   - Fairness and responsibilities
   - Aesthetics and affordances appropriate to accomplish the goal
   - Simplicity of interactions
   - Usability
   - User experience

4.2.4 Open Problems

In our opinion, the largest open problem is that the scientific study of designing a heterogeneous systems is difficult because these three above design principles all inherently interact.

Other open problems include:

- How to consider adaptation in team dynamics? How do you adapt each team member to have positive results for the overall team goals and subgoals?
- How to design incentives for team participants that maximize team performance and also caring about the motivation of team participants? How do you incentivize the right people to do the right tasks?
- How to construct a team to address the joint goal in the best way to maximize results?
- When to intervene in a team interaction to help out maximizing for optimal team performance?
- How to delegate roles within a team for maximum benefit?
- How do you decide in mixed teams when to transfer control for decision problems? How do you decide when to hand off a problem to someone else?
- How to infer participants' plans and goals?
- How to design behaviors for long-term interactions with agents that are involved in team dynamics?
- How do you evaluate the policy online?

4.2.5 Design cycle for agents to interact in groups and teams

The design cycle for creating an agent to interact in groups and teams involves:

[5] https://hackernoon.com/dieter-rams-10-principles-of-good-design-e7790cc983e9
[6] https://www.fastcompany.com/3067632/10-principles-for-design-in-the-age-of-ai

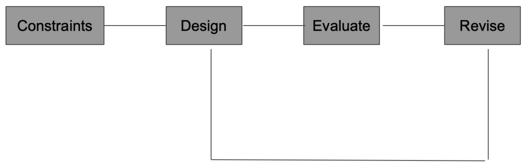

Figure 1 Design cycle.

1. **Elicitation** – Ask stakeholders about their needs, capabilities, and their issues. Consider the environment. How many agents? For how long? What exactly is the communication? How are you representing the information gathered (aggregate or personal data)?
2. **Representation** – Find the best representation that you can use to satisfy the targets/criteria that you have. (e.g. how are you giving badges in stack overflow?)
3. **Optimization** – Optimize either the environment or the policies of the agents.
4. **Evaluation** – Run randomized controlled studies to validate or determine what needs to be modified, then return to step 1 and repeat until optimal.

4.2.6 Promising Ideas

- Develop participatory design techniques to develop the best possible solutions over time in iterative design in partnership with the stakeholders. This allows for deployable systems in real-world applications and domains.
- Run randomized control studies on collaborative systems (e.g. stack overflow, reddit) that have large amounts of available data.
- Evaluate with actual team members in real-world scenarios and take it out of the lab.

4.2.7 Conclusions

In order to successfully create groups of humans and agents, the development of such agents and their tasks must follow certain criteria and guidelines. The discussion of this breakout group made a step forward towards the definition of those guidelines by creating a set of design principles for robots and agents to interact in group settings. Moreover, we propose a design cycle to address the stages and process of designing those interactions and we identified open problems and promising ideas within this topic.

References

1 Nihan Karatas, Soshi Yoshikawa, and Michio Okada. Namida: Sociable driving agents with multiparty conversation. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, pages 35–42. ACM, 2016.
2 Ryosuke Mayumi, Naoki Ohshima, and Michio Okada. Pocketable-bones: A portable robot sharing interests with user in the breast pocket. In *Proceedings of the 7th International Conference on Human-Agent Interaction*, pages 211–213. ACM, 2019.

4.3 Working Group on Human-Agent Team Dynamics

Filipa Correia (INESC-ID & University of Lisbon, PT, filipacorreia@tecnico.ulisboa.pt)
Hatice Gunes (University of Cambridge, GB, hatice.gunes@cl.cam.ac.uk)
Stefan Kopp (Universität Bielefeld, DE, skopp@techfak.uni-bielefeld.de)
Iolanda Leite (KTH Royal Institute of Technology, SE, iolanda@kth.se)
Mark Neerincx (TU Delft & TNO, NL, mark.neerincx@tno.nl)
Catharine Oertel (TU Delft, NL, C.R.M.M.Oertel@tudelft.nl)
Rui Prada (INESC-ID & University of Lisbon, PT, rui.prada@tecnico.ulisboa.pt)
Marynel Vázquez (Yale University, US, marynel.vazquez@yale.edu)

4.3.1 Introduction

The Human-Agent Team (HAT) working group focused on the identification of key HAT-concepts and challenges. Working *definitions* of the key concepts were formulated, the corresponding *dimensions of change* and characteristics of *team processes and patterns* were worked out, example *scenarios* were proposed to exemplify the HAT dynamics in research and development, and *challenges* were derived from this exploration.

4.3.2 Definitions

The first discussions centred on the core definitions and the particular properties that make a collection of people a group or a team. The three following concepts have an inheritance-type of relationship: "collection of agents", "group" and "team".

Collection of agents
- Multiple agents with individual goals, abilities, skill, expertise;
- Having a minimal degree of autonomy (ability to decide on their actions);
- Not necessarily co-located;
- But with ways to interact with each other, closely or loosely coupled.

Group
- Agents are individually aware of having a shared identity (group, commonality);
- Awareness of in-group/out-group agents.

Team
- Agents have a joint goal or task they are working on;
- Agents are aware of their working on it together;
- Agents are committed to it and mutually support each other;
- (not necessarily interdependent).

4.3.3 Dimensions of change

Given the definition of teams in section 4.3.2, we identified the following dimension of change and some examples of each:
- **Team organisation**: structure, roles, and norms
- **Team members**: number, capabilities, and autonomy
- **Relationships**: trust, liking, intimacy level, and power

- **Group attitude**: commitment, (group) identity, and trust
- **Shared cognition**: experiences, knowledge, skills, and awareness
- **Group properties**: cohesion, interdependencies, performance (change), resilience, and distribution of participation
- **Context**: environment, task, resources, and tools

4.3.4 Team Processes & Patterns

The members of a human-agent team adapt their behavior to each other and the dynamic environment in which the team operates. Constructive and destructive behaviour patterns can be implicitly or explicitly brought forward [3], e.g. based on experience or just "emerging".

The following team processes can be distinguished [1]:

Forming the group. This first stage includes the tentative communication, uncertainty and exchange of personal information of the group members. It leads to the sense of belonging and the shared identity of the group.

Establishing norms and common ground. At this stage, the group develops standards and/or agrees on the procedures to operate.

Assigning roles with responsibilities. The roles attribute a certain structure to the group and usually are intended to improve the communication among members.

Planning, executing, monitoring and repairing the (group) tasks. This stage hold the actual performance of the group on the defined task.

Managing relationships and conflicts. Trust is an important construct, characterizing relationships in a team. Trust develops over time and adequate trust calibration is crucial for collaboration. When the benefits and costs of the relationship outcomes are harmonized for the concerning team members (i.e., there is a balanced *relationship equity*), these team members will collaborate well [2].

4.3.5 Envisioned Scenarios

Goal: autonomous agent (virtual/robot) that can perform in a team such that the team performs better. Possible scenarios (5 to 10 year vision):

- Health-care in a hospital setting by human-robot teams (healthcare professionals, nurses, social robots, ...)
- Collaborative assembly
- Search and rescue, disaster response, ...
- Entertainment, education, citizen-science...

4.3.6 Open Challenges

Open challenges from an individual level to a group level at the same time as from a low-level sensing to a cognitive decision-making level (see Figure 2):

- What sensing/perception and acting/expressive capabilities do robots need to have in teams?
- How can it maintain the necessary level of social awareness and task-related awareness?
- How to accommodate for social norms, culture, individuality of team-members, etc.?
- How can robots engage in constructive team coordination patterns and processes?
- How can robots/agents learn what role is best suited for that particular team?

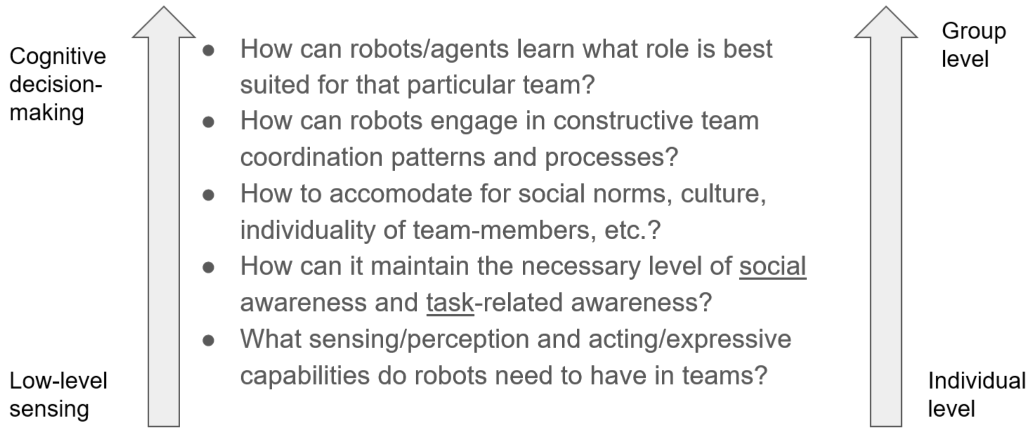

Figure 2 Open challenges for research & development of team agents, from low-level sensing to joint decision-making and from individual to team level.

4.3.7 Conclusions

The presented definitions, dimensions and team processes were discussed in the light of interpersonal group interactions and the some of the well-established dynamics of those groups [1, 4]. Therefore, this discussion highlighted the importance of redefining this concepts for groups (and teams) in which both humans and agents are part of. Nevertheless, some of the proposed challenges and the envisioned scenarios identify and guide future approaches and avenues to explore this topic.

References

1. Rupert Brown and Samuel Pehrson. *Group processes: Dynamics within and between groups*. John Wiley & Sons, 2019.
2. Ewart J. de Visser, Marieke M. M. Peeters, Malte F. Jung, Spencer Kohn, Tyler H. Shaw, Richard Pak, and Mark A. Neerincx. Towards a theory of longitudinal trust calibration in human-robot teams. *Int. Journal of Social Robotics*, in press.
3. Jurriaan Van Diggelen, Mark Neerincx, Marieke Peeters, and Jan Maarten Schraagen. Developing effective and resilient human-agent teamwork using team design patterns. *IEEE intelligent systems*, 34(2):15–24, 2018.
4. Donelson R Forsyth. *Group dynamics*. Cengage Learning, 2018.

4.4 Working Group on Social Cognition for Robots and Virtual Agents

Cindy L. Bethel (Mississippi State University, US, cbethel@cse.msstate.edu)
Merijn Bruijnes (University of Twente, NL, Merijn Bruijnes)
Malte Jung (Cornell University, US, mfj28@cornell.edu)
Samuel Mascarenhas (INESC-ID & University of Lisbon, PT, samuel.mascarenhas@gaips.inesc-id.pt)
Christoforos Mavrogiannis (University of Washington, US, cmavro@cs.uw.edu)
Simon Parsons (University of Lincoln, GB, simon.parsons@kcl.ac.uk)
Catherine Pelachaud (Sorbonne University, FR, catherine.pelachaud@upmc.fr)
Rui Prada (INESC-ID & University of Lisbon, PT, rui.prada@tecnico.ulisboa.pt)
Laurel Riek (University of California, US, lriek@eng.ucsd.edu)
Sarah Strohkorb Sebo (Yale University, US, sarah.sebo@yale.edu)
Julie Shah (Massachusetts Institute of Technology, US, julie_a_shah@csail.mit.edu)
Elaine Short (Tufts University, US, elaine.short@tufts.edu)
Marynel Vázquez (Yale University, US, marynel.vazquez@yale.edu)

4.4.1 Introduction

This section discusses the cognition required by physically embodied social agents that interact in a group context. We consider a *group* to be a set of agents (e.g., humans, robots, and virtual agents) that share certain characteristics or relationships. For example, group members might be connected by identity, location, or the beliefs that they hold. Consistent with existing work (e.g. [7, 30]) we consider teams as a specific type of group whose members "are interdependent in their tasks" [7], interact socially, and share a common goal.

While the term "social agents" covers a broad spectrum, ranging from chatbots to virtual characters to physical robots, in this article we focus specifically on *physically embodied agents*. This includes both physical robots able to enact physical change in the world and embodied virtual agents, who may appear within physical objects capable of engaging in groups with humans. We will thus refer to these as embodied social agents (ESA) throughout this work.

Social Cognition

Social cognition refers to how people process, store, and apply information about other people and social situations [45]. While social cognition is integral to how people perform in both work and non-work related interactions, research on computational models of social cognition largely focus on purely non-work related social interactions. There is evidence that this delineation undermines the task effectiveness and acceptance of robots and virtual agents. For example, medical professionals were found to actively sabotage a robot that was navigating within a hospital environment without consideration of the social norms that protect medical professionals from interruption during high workload situations [37]. Considering social cognition for groups and teams brings with it novel challenges that are yet to be addressed.

Social Cognition in Groups

To distinguish social cognition in groups from that in individuals, we draw from Brown and Pehrson's [5] work that highlights three concepts that are important when characterizing how people behave as group members versus individuals. The first concept, *social identity*, refers to how people define themselves in terms of a group and how they attach emotion and value to these self-categorizations. The second concept is *social context*, where people's social identities are dependent on the context in which they find themselves and the groups with which they are part of. This can have both macro-level societal structure as well as the micro-context of specific social environments. The third concept, *social actions*, refers to the affordances that groups offer in helping people enact change in the world. It also reflects the notion of collective action to achieve social change.

These distinctions are useful in framing how ESAs team in groups, where an ESA will have its own social identity of capabilities and functionalities that it can employ in group settings, as well a set of roles [53, 18], which may be known *a priori* or may be learned during interactions. For example, a robot tour guide may take the role of a mentor, teaching a group of students about an exhibit. After the museum closes it may adopt the role of a mechanic, helping workers take down an exhibit. In these different roles, the robot may exhibit very different capabilities and functionalities.

The social context can also be informative for an ESA to support its cognition, particularly with regards to supporting constraints on its behavior, establishing common ground, and informing its decision making [38, 41, 60]. Additionally, using context can help ESAs to reduce the problem space and bootstrap problem satisfaction [48]. Finally, social action can provide ESAs affordances for actions within groups. For example, a human group may be engaged in a shared workspace activity, and a robot can observe human-human interaction as a model for how it should behave within it [20, 22, 65].

4.4.2 Characterizing ESA-Group Interactions

We distinguish between different levels of cognition depending on how an ESA's actions take groups into consideration (Figure 3). First and foremost, ESAs need to be able to act with *groups in mind.* Next, ESAs can *act on groups* as outsiders, influencing or shaping aspects of group behavior. Finally, they can *act in groups* as another group member. At this level, they not only influence the group but also need to take part directly in the task and inherent dynamics of the interaction.

Figure 3 captures the idea that ESAs who *act in groups* are a subset of those who *act on groups* or can influence the outcomes of a group. Both of these types of ESA would be subsets of those who can *act with groups in mind.* This structure emerges from the fact that acting on groups requires all the abilities of acting with groups in mind plus some additional abilities. Similarly, acting in groups requires all the abilities of acting on groups but also requires some additional abilities.

Acting with Groups in Mind

This is the most basic level in which group cognition can be considered. At this level, ESAs account for different aspects of the group with its cognition irrespective of the number of people/ESAs who are part of the current interaction.

We consider any social interaction with two or more ESAs as group-situated to some degree. This position has implications for human-robot interaction (HRI) research as groups can be taken into account independent of whether the interaction that takes place is dyadic

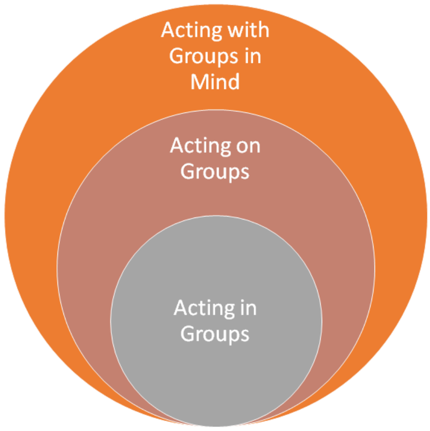

Figure 3 A framework for considering group cognition.

or more complex. For example, consider a robot that enters a room and hands over an object to a single person who is alone in the room. Here the robot is in a dyadic interaction, but can (and arguably should) take group membership (e.g. culture) into account when reasoning about how it completes its task. Of course, if the person is not alone, then there will be other aspects of group membership that will need to be taken into account, such as not interrupting a speaker.

Acting on Groups

Acting on a group refers to the level of group cognition that is involved when the purpose of an ESA's action is to influence or shape a group's structure or behavior without the ESA being a member of the group. Such influence is analogous to a dog herding a group of sheep. The dog shapes the behavior of the group of sheep without being a member of that group. An example of a robot designed to purely act on a group without being part of it is the Micbot [62]. The robot influences a group's participation dynamics by signaling engagement and by nudging group members to participate, though it is not directly a part of the group it influences. A key, implicit, idea in acting on a group is that the robot may work in a pro-social way to improve the functioning of the group. Doing this in a way that does not hijack the task that the group is working on will require the robot to provide subtle interventions, such as the Micbot probing members who have spent less time speaking. Being able to do this in a general, flexible, way, needs the robot to be equipped with a deep knowledge of group behaviour, and the ability to reason about this (in particular the ability to predict the consequences of specific interventions).

For ESAs to be able to effect groups, it is helpful to consider the kinds of specific affordances that groups offer for interaction and influence. Affordances are "properties of the world that are compatible with and relevant for people's interactions" [13]. Gaver's notion of affordances is different than Gibson's [15] original concept of affordances or Norman's [40] concept of perceived affordances. For Gibson, affordances naturally exist as properties of the world. Norman's concept of perceived affordances highlights that interactions are driven less by the actual properties of an object but rather by how it is perceived. For example a door might have the affordance to be pulled (based on its mechanical design) but the particular door handle design leads to a perceived affordance as "pushable." Since a robot is not dependent on human-like perception of the world, we adopt Gaver's definition of affordances here. Drawing from Gaver [13], we define group affordances as properties of *groups that are compatible with and relevant for interactions.*

Groups offer unique affordances dependent on characteristics such as structural composition (e.g. a leaderless group affords different interactions than a team with a pronounced leadership structure) or size (e.g. a small three-person group affords different interactions than a stadium full of people). ESAs can leverage such group affordances to influence their behavior. For example, Kwon and colleagues [31] showed that a robot can leverage a group's unique hierarchical structure to shape its behavior in systematic ways. While to our understanding there are no current approaches to reason about group affordances specifically, existing work has developed approaches for reasoning about affordances more generally. For example, Sarathy and Scheutz proposed a computational framework for inferring affordances [51]. Similarly, Shu, Ryu, and Zhoo [58] introduced an approach for learning "social affordances."

Acting in Groups

Acting in groups refers to a level of cognition that is required for robots to serve as members of a group and not just as tools used by the group. This involves abilities to influence groups and a general consideration of cognition as group-situated. The robot can still influence the actions of other group members as in the previous level but it also needs to have cognitive abilities that specifically pertain to the task or activity that the group is performing. As such, this level offers more opportunities for a robot to have more impact on the group than what would be possible for an outsider. It may also be more difficult to generalize its impact across different group tasks or activities. Groom and Nass argued that it is impossible for robots to become members of work groups or teams as they lack the basic abilities to build and maintain trust [19].

For an ESA to become a member of a group, it is not necessarily enough that it is able to identify whether a group exists or not. Additionally, the robot needs to be able to make sense of how the other group members perceive its membership status as well as what roles are expected to be allocated to the robot within the group. While not so important in highly controlled settings, these capabilities become quite relevant in scenarios where group formation and role allocation occurs in a highly dynamic manner.

4.4.3 Situation Awareness – Decision-making – Action: Group Interaction Throughout the Sense-Think-Act Cycle

Parasuraman, et al. [44] proposed a taxonomy that guides functional automation design. They identified four stages of human information processing that may be supported by automation: information acquisition (Stage 1), information assessment (Stage 2), decision and action selection (Stage 3), and action implementation (Stage 4). We explore how the same framework can guide the design of social interactions with agents. Many aspects of social interaction (depicted in Fig. 4) involve these four stages or processes of human cognition.

In our analogy, situational assessment is composed of processes for: *perception*, to perceive the state of other agents, the group or team, environment; *comprehension*, to assess the relationships among entities in the environment, including interactions between humans and robots; and *projection*, to infer the intent of other agents and predict plans, behaviors, trajectories. Under the Endsley model [11] for human situational awareness, this subsystem represents the comprehension and projection elements of situation assessment. Decision-making may be a skill-, rule-, or knowledge-based [46]. Action can include taking physical action, or implicit/explicit communication and signaling.

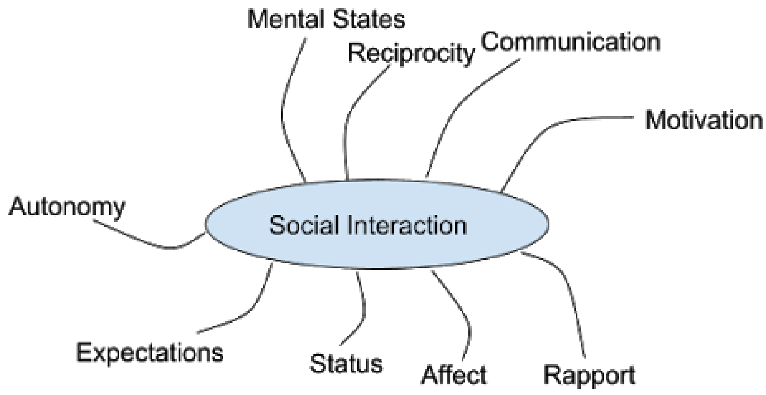

Figure 4 Dimensions of social interaction.

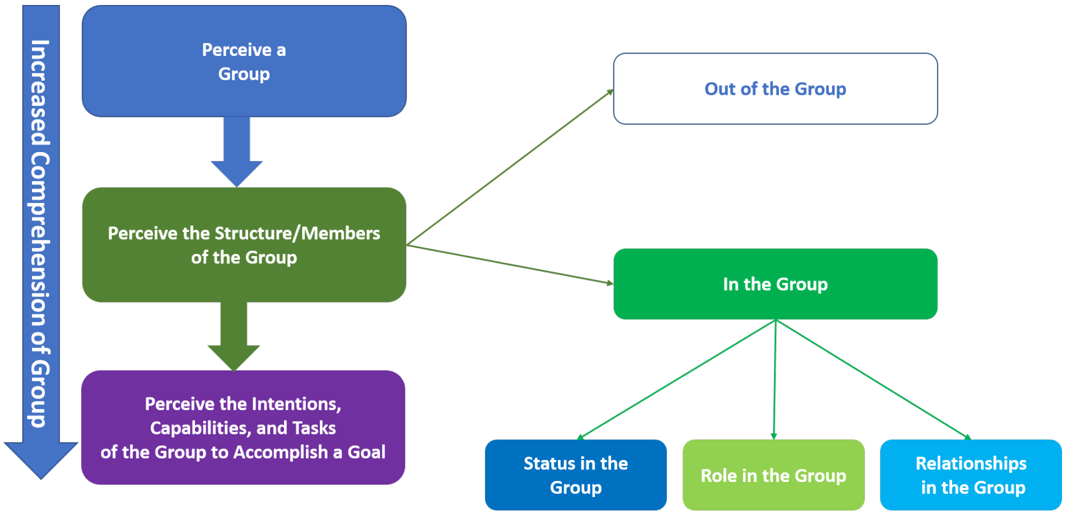

Figure 5 Overview of increased comprehension of group understanding.

As depicted in Figure 5, when agents (including humans and/or robots), encounter one another they first perceive the group. Once the group is perceived or identified, the next phase is to perceive the structure or members of the group. As part of that process, a determination needs to be made of whether individual agents are members in the group or if they are outside of the perceived group. For those agents perceived as being in the group, members will have roles, statuses, and there are relationships between those members that must be recognized and understood. The next phase is to then perceive the actual intentions, capabilities, and tasks that the group and its members need to consider to accomplish the goals or intentions of the group. From the moment of perception of a group there are increasing levels of comprehension or understanding of the dynamics of the group and its members.

In the case of new members joining a group, the members first tend to imitate the actions and behaviors of existing members of the group (refer to Figure 6). They follow the examples of those established members until they obtain more knowledge. As learning occurs over time, a certain level of knowledge is achieved and the member more fully integrates into the group. Imitation becomes knowledge through the process of learning about the intentions, capabilities, and tasks of the group, which allows for members to better understand why they are performing certain behaviors to be more fully integrated as part of the group.

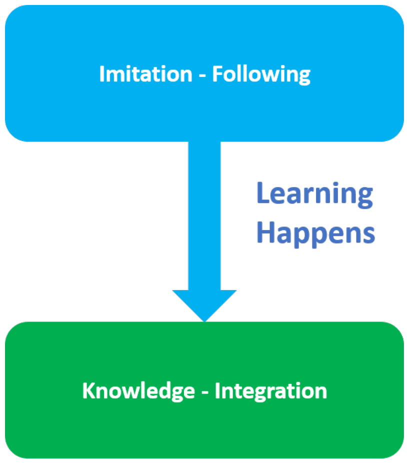

Figure 6 Members of groups transform from imitation to integration in their behaviors.

To further expand on the transformation from imitation to integration, group dynamics will often begin with a system of reactive behaviors. This can be especially true when new members are involved in the group. During these types of group interactions, the behaviors tend to be more reactive in nature. A situation in sensed or perceived and the group responds or reacts. Over time, as learning occurs, the dynamics begin to transform and the behaviors become a hybrid of deliberative decisions and reactive responses [1]. Members of the group have a better understanding based on experiences and knowledge of how to decide to respond to the situation encountered. During these types of interactions, a situation or encounter is perceived, followed by comprehension of the potential impacts of the situation, and then the agents are better able to decide the best response to perform. Additionally, there may be an element of planning or deliberative thoughts that occur once comprehension of the situation happens, planning can then occur for how to react, and then the appropriate response is performed to accomplish the goals or tasks for the group.

We are interested in the unique challenges posed by social cognition in groups, where assessing relationships among agents and projecting intent and behavior must be performed across varying time-scales and levels of information abstraction.

For example, an agent addressing specific people in the group can affect not only the behavior of the target user, but also the cohesion of the group as a whole, in both the short- and long-term [57]. Furthermore, the effect of agents on groups includes short-term phenomena, such as rapport or engagement, that lead to long-term effects such as trust or friendship. This means that agents need to detect, reason about, and take actions that will affect the group dynamics on multiple time scales, from seconds to years. Another challenge is the difficulty of computationally representing these social phenomena. Ideas such as engagement, trust, intent, and discomfort need to be translated into reward functions, actions, preconditions, and so on. In the current state-of-the-art that accounts for social considerations, it is often heuristic, because it is hard to measure social features and outcomes in a computational way (and to enumerate all of the possible failure modes). Assigning credit for social interactions and reasoning about the consequences of actions needs to be done over time and among group members. Finally, acting in groups magnifies the effects of actions: for example, humans can ally with each other with or against the agent, forming subgroups. Effects of agent behavior are also magnified in multi-agent scenarios because it increases the influence of the agent(s) as a group.

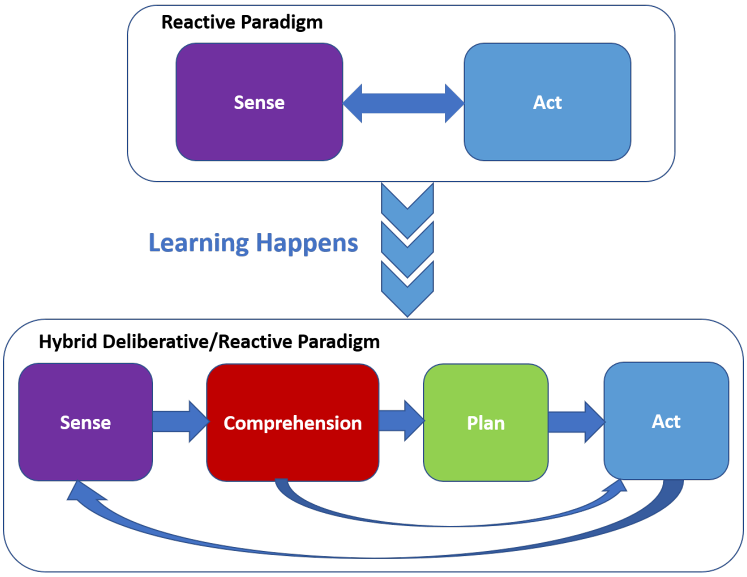

Figure 7 Reactive and Hybrid Deliberative/Reactive Paradigms for interactions in groups.

4.4.4 Group-Specific Considerations

Clearly it is harder for an ESA to deal with a group rather than an individual. We distinguish two aspects to this. First, it is often the case that group settings lead to phenomena not typically observed in one-to-one interactions. Second, groups give rise to elements that are not present in one-to-one interactions. Among the group-specific phenomena, we can identify the following:

Diffusion of responsibility. This has two effects: first, humans might be less willing to just take care of things, because there are more agents, and second, each ESA has less responsibility, even if the group as a whole has more effect.

Conforming. ESAs may conform to people within a group and people also will likely conform to agents within a group [50]. It is important to know when to conform and how to wield their influence when trying to have people conform to ESAs.

F-Formations. Another type of group phenomena that can emerge are *F-Formations*: spatial formations typical of situated group conversations which result from the need of interactants to share information and perceive one another during the interaction. While dyadic formations can be observed in one-to-one interactions, groups may lead to other types of configurations worth considering in HRI. For instance, detecting F-Formations can aid in identifying social interactions in human environments and, ultimately, help enable appropriate ESA behavior in human environments [64]. The disposition in space of the group should dynamically evolve as a new member join or leave the group. F-Formation should be updated on the fly to reflect these changes.

Modeling social influence. In groups, the ability to reason about social influence is critical to being able to determine the effects of one's actions. For example, the ESA may directly influence (nodes connect within a graph) some people and indirectly influence other people (as a result of the direct connections). Feedback will likely also be helpful in refining these models after actions are taken.

Turning to elements that are not present in one-to-one interactions, we have:

Group size. The size of the group in which interaction is taking place affects that interaction [66]. Large groups behave differently than small groups [63]: there may be local patterns, and an ESA may not need to know everything else. There is also a need to consider sub-groups, which may be dynamic, so relations may differ between members of the overall group, the members of a subgroup, and relations between subgroups.

Context. Context helps to constrain what is important in the interaction. This can happen through constraints on behavior (norms), common ground, and what actions other members of the group are likely to take. For example, in a negotiation (or other adversarial interaction), an individual would want to track how open the other group members seem to their proposal, and the individual might want to be less open about emotional states, and would expect other members of the group to engage in negotiation behavior.

Social status. Successful interactions must involve an awareness of the social status of group members. While issues of status are most salient when working in groups, even when interacting with a single person, status is still relevant because a person or ESA's status is a result of relationship with groups of people.

Finally, group interactions involve unique **conversational decision making skills** that ESAs must develop in order to verbally communicate. Examples of these conversational decision making skills include selecting whom to address and choosing whether or not to interrupt someone.

4.4.5 Social goals of ESAs within groups and teams

There are both long and short term social goals that aid an ESA's interactions on a group, in a group, and keeping groups in mind. These social goals may apply to dyadic interactions, however, become exponentially harder in groups due to the increased complexity of relationships and interactions.

Long-term group dynamics

ESAs must build and maintain relationships with people in their group or team, as well as the group itself. Stable relationships can be important for resilience to agent failure. Maintaining relationships involves, for example, taking into account the affective state of the group and its members. But, specifically in groups, the concept of in-group / out-group should be taken into account [3, 67] as the inter-group relations may influence the interpersonal ones. Properly building and maintaining relationships is important to manage group cohesion, stability, sense of shared identity and common purpose. Complying to politeness theory [4], for example, could be a means to achieve such goals. Some dimensions to be considered are:

Managing affect. At a minimum, in order to have an effective role within a group or team, relationships with the people in the group must remain positive and the ESA must avoid hindering the goals and work of the people with whom it is interacting [25, 37].

Building and maintaining interpersonal trust. Interpersonal trust is distinct from task-related trust. Interpersonal trust is related to long-term rapport and more of a function of the relationship between ESAs, independent of context. McKnight and Chervany [36] identify trust as conceptual categories competence, benevolence, integrity, predictability, and other characteristics such as open, careful, safe, shared understanding, and being personally attractive.

Managing social influence. ESAs should identify and manage different sources of influence among group members, for example, social power, friendship, expertise and status. They should find a balance between the leadership dynamics supported by these sources.

Building and maintaining group commitment. ESAs can be committed to be part of a group and/or to actively achieve the group's shared goal. Ensuring an ESA's commitment, as well as other members' commitment, over time is the result of a process that involves coordinating the ESA's and other members' participation in the group, working toward and maintaining the group stability, task progress and social cohesion.

Building and maintaining shared cognition. Through the group interactions the members work towards establishing common ground, shared knowledge and representations, shared experiences and mutual understanding. Moreover, each member of the group builds a cognitive model of self and all other members of the group and updates this as the interaction unfolds. This model may include individual mental and affective state, level of commitment, task responsibility and expertise, and social norms.

Short-term interaction in groups

There are shorter-term goals that will build into the long-term ESA-group relationship. Many of these phenomena have been studied in one-on-one HRI, but relatively fewer in group interactions. For example:

Avoiding discomfort. ESA behaviors may often cause human discomfort. The mechanisms underlying the emergence of discomfort as a result of agent behaviors are not sufficiently understood. This results in ESA behavior design resulting in undesired phenomena, such as the reciprocal dance in navigation [12].

Adhering to social norms. In order to maintain their membership in groups as well as relationship with group members, an ESA must adhere to social norms [33]. This may include being polite, being respectful toward others, and knowing when it is appropriate to be more direct [16].

Expressing intent. In order to ensure effective social interactions with groups of humans, it is crucial for an ESA to be able to succinctly express its own intentions. Ongoing work looks at encoding intentions into ESA actions (e.g., collaborative manipulation [10], navigation [35]). However, depending on the task and the context, different mechanisms, modalities, and strategies of conveying intent may be more applicable or effective. Research on identifying and understanding these mechanisms would enable ESAs to adapt to the social context more naturally.

Inferring human intent. Social interaction with people requires an understanding of human intentions. By leveraging signals encoded in human behavior, ESAs may infer latent human desires, preferences and goals. Past work has focused on interpreting task-specific human intentions to achieve desired performance in shared autonomy applications (e.g., [2, 24]). Future work could expand this work into the group domain to enable seamless interaction and coordination.

Repairing trust and expectations after errors. The ESA must assume that it will make errors and will need to detect and repair those issues [54]. This is important so that when a large mistake comes along, the consequences are not as extreme.

Setting expectations. ESA expectation setting can help manage trust and prevent large loss of trust if expectations are too high. Members of a group share responsibility for the social affairs in the group and compromising the initial expectations can have a detrimental effect. In human-human interactions such violations often trigger repairs. Compare the door in the face (DITF) and the foot in the door (FITD) techniques [9], where asking for an initial favor influences the acceptance of a request for a further favor.

Monitoring group state and status. A probe could be used to affirm or check one's status within a group, or to get the "vibe" of a group. Repairs can also act as probes to ensure that relationships are maintained.

Adhering to social roles expectations. When taking a specific role in the group, the agent must comply with the expected responsibilities and affordable behaviours.

Choosing the right member to address. In the case of groups, carefully considering the choice of the interaction partner in a particular moment is very important as it signals all participating members. Addressing the group as a whole is an option as well. This choice highly affects the group dynamics and the development of interpersonal relations, in particular, the attitudes towards the ESA.

Recognizing and expressing group-based emotions. Group-based emotions are particular emotions that come into play when considering group interactions. These are emotions that result from an appraisal process in which group identity is higher in salience than self identity [17]. ESAs that are able to express appropriate group-based emotions can promote a stronger sense of group identification from all the members [8].

Expressing affect. Expressing affect (both positive and negative) plays a crucial role in social control between group members [52, 26]. For example, expressing negative affect may be important for some interventions in the group dynamics, such as, to signal members to change their current behaviours. On the other hand, expressing positive can encourage behaviors of others. An ESA can apply such expressions strategically related to the task (e.g. social skills training [6, 61]).

4.4.6 First Steps to Advance the State-of-the-Art

In this section, we propose some promising areas for development of new research, building on the state-of-the-art. We divide the areas into three categories: designing and studying ESA behaviors in groups; creating theories and models to inform ESA decision making within groups; and developing control and learning algorithms for ESAs to use in group interactions.

Designing and studying agent behaviors in groups

There are many human behaviors that ESAs can emulate to further both long and short term group interaction goals (see Section 4.4.5). These behaviors are likely best investigated through design research and user studies. We believe that the following examples will lead to fruitful advances in our understanding of ESA cognition:

- **Using repair behaviors.** Despite growing body of work within HRI on repair (e.g. dyadic human-robot trust repair [32, 54] and robot conflict repair between human group members [27, 56]), there are still many aspects of repair worth exploring. For example, how does the repair of a dyadic relationship may influence other members in a group? Does it improve overall group cohesiveness? Or, when is a repair necessary to recover the reputation of the team member who the ESA maligned?

- **Setting expectations.** How can an ESA appropriately set expectations of its capabilities in group contexts that may vary in group membership as well as the knowledge and background of individual members? Expectation setting is critical to perceptions of robots and the ensuing interaction [43]. Besides the initial introduction, this can also be a recurring process during the interaction. For example, when the goal of the group changes different abilities might be needed and a new negotiation on group roles can ensue (and thus the expectations of each group member). Another example, in an education setting a teachable agent (an agent that is introduced as a peer that understands less than the student) is shown to make students put more effort in learning [59].
- **Rapport.** What factors influence the rapport within a human-ESA group? How does rapport evolve over time in a group interaction? Are there differences in rapport between human-human and human-ESA interactions? How can we expand prior work on one-on-one rapport [47] to group interactions?
- **Expressing intent.** How can an ESA express its intent in a group where the members have a diverse set of backgrounds, viewpoints, and knowledge bases? Additionally, how can an ESA express its intent exclusively to a subset of a group and not the other members?
- **Inferring and respecting social norms.** Human interactions are governed by norms [14]. ESAs that participate in human group activities will need to be able to conform to those norms [33] and, ultimately, to be able to infer those norms from observing the interactions between group members. How can they effectively correct for wrong or incomplete models of social norms? Furthermore, how should ESAs deal with situations in which their group members conform to different norms?
- **Expressing and detecting discomfort.** What features of group human-ESA interaction are indicative of discomfort with the ESA? How can an ESA express discomfort with human behavior in order to better meet its needs without damaging the group interaction?
- **Tracking long-term relationships.** How do group human-ESA relationships change over long-term interactions? What features of the interaction are important to understanding these relationships?

Creating theories and models to inform ESA decision making within groups

In order to understand a group, ESAs will need to use appropriate representations and abstractions of the group and its functions. We highlight several challenges that require the development of new representations, abstractions, and models.

- **Representing state.** How can we effectively represent states? There are cases in which the information that needs to be represented in different ways as a function of varying groups or contexts. There is also the problem of the curse of dimensionality.
- **Modeling Influence.** How can we represent influence within a group? And which representations (e.g. graphs with nodes and weighted ties) would be helpful to show the dispersion of influence? How can we dynamically alter this representation?
- **Abstracting behavior.** How can we abstract/represent behaviors over space and time in a computationally efficient fashion? What is the right type and level of abstraction to capture salient yet computationally practical aspects of group behaviors? And how could an ESA decide on where to focus its attention within a long-term memory of observations? Existing work has looked at abstracting collective group behavior into symbolic structures (e.g., in navigation [34]). This allows for the use of classical planning and inference methods. However, it is still unclear how to formalize the abstraction across different tasks and contexts.

- **Modifying representations over time.** If an ESA receives feedback signals from people around it that are not congruent with its current representations, how will we allow the ESA to change its representations? For example, if a pizza delivery robot makes a delivery within a library and communicates at a normal volume with the librarian who ordered the pizza, those in the library may provide the robot with negative feedback signals. How can the robot modify its representations of delivery locations and appropriate behavior within them (e.g. speaking volume), so that it behaves appropriately in all of its delivery locations?
- **Modeling Adaptation/Consensus.** What is the right direction towards modeling adaptation and consensus within groups of humans and ESAs? How can we leverage multi-modal signals inherent in ESAs' behaviors to model the state of consensus within a group? And what is an effective way of incorporating such a model into the ESA's decision making? Existing literature has proposed decision-theoretic [39] and information-theoretic approaches [29]. Can we unify such models towards formalizing a general theory of adaptation for human-ESA interaction within groups? Some work on modeling human-multi-ESA interactions has been done [28].

Developing control and learning algorithms for ESAs to use in group interactions

In this section, we propose new areas for technical development of intelligent behavior for ESAs, including promising new areas for developing computational models and methods that will enable ESAs to successfully accomplish the goals discussed in the previous sections.

- **Probing for interaction and group state.** Active information gathering can enable ESAs to collect specific information faster than with passive observation [55]. It can help robots understand the mental states of other ESAs [49]. In the context of group interactions, it would be interesting to explore what types of behavior allow an ESA to most effectively obtain information about a group? How do groups respond to information-seeking and group-evaluating ESAs? How can active information gathering actions help repair interactions in groups? Additionally, what new methods are needed to take advantage of the agent's embodiment and role in the group?
- **Sensing and assessing groups.** How can we measure key components of group interactions using a virtual agent or robot group member? How can we build on current work, for example on expertise assessment [20, 21, 22, 23]? How can we infer intentions, dominance relationships, and group affect? Another challenge is to deal with groups of multiple independent ESAs and one or more humans [42]. Especially when groups of multiple agents form 'in the wild', our models might not be able to recognise social behavior or other ESAs.
- **Simultaneous planning for task and social goals.** How do current planning methods need to be modified to allow planning for at least one social goal at the same time as task goals?
- **Using transfer learning.** How can we develop knowledge over time about a single group? How can we transfer modeled social dynamics from one task to another?
- **Building rapport** How can we use control theory or other continuous representations to build rapport between an ESA and a group? How can we make social agents in groups more responsive to human behavior? How can we make the ESA's responses more intelligent and relevant?
- **Detecting anomalies.** What anomalies can occur in social signals? What new methods might allow us to detect these anomalies and reason about their implications?

4.4.7 Conclusion

This report outlines the social cognitive abilities required for ESAs to perform effectively in a group context. The report characterizes types of ESA-group interactions, introduces a cognitive behavioral architecture for robots to act in groups, and highlights group dimensions relevant for the successful integration of ESAs into contexts. Based on the assertion that any social interaction is a group social interaction to some degree, we claim that our recommendations have relevance not only for the design of virtual agents and robots that are deployed in immediate group and team contexts but for HRI in general.

References

1 Ronald C Arkin. *Behavior-based robotics.* MIT Press, 1998.
2 Reuben M. Aronson and Henny Admoni. Semantic gaze labeling for human-robot shared manipulation. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, pages 2:1–2:9. ACM, 2019.
3 Marilynn B Brewer. The psychology of prejudice: Ingroup love and outgroup hate? *Journal of social issues*, 55(3):429–444, 1999.
4 P. Brown and S. Levinson. *Politeness. Some Universals in Language Usage.* Cambridge University Press, Cambridge, 1987.
5 Rupert Brown and Samuel Pehrson. *Group processes: Dynamics within and between groups.* John Wiley & Sons, 2019.
6 Merijn Bruijnes, Rieks op den Akker, Arno Hartholt, and Dirk Heylen. Virtual suspect william. In *International Conference on Intelligent Virtual Agents*, pages 67–76. Springer, 2015.
7 Susan G Cohen and Diane E Bailey. What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23(3):239–290, 1997.
8 Filipa Correia, Samuel Mascarenhas, Rui Prada, Francisco S Melo, and Ana Paiva. Group-based emotions in teams of humans and robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 261–269. ACM, 2018.
9 James P Dillard, John E Hunter, and Michael Burgoon. Sequential-request persuasive strategies: Meta-analysis of foot-in-the-door and door-in-the-face. *Human Communication Research*, 10(4):461–488, 1984.
10 Anca Dragan and Siddhartha Srinivasa. Integrating human observer inferences into robot motion planning. *Autonomous Robots*, 37(4):351–368, 2014.
11 Mica R Endsley and Daniel J Garland. *Situation awareness analysis and measurement.* CRC Press, 2000.
12 Franck Feurtey. Simulating the collision avoidance behavior of pedestrians. Master's thesis, University of Tokyo, Tokyo, Japan, 2000.
13 William W Gaver. Technology affordances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 79–84. ACM, 1991.
14 Michele J Gelfand, Jesse R Harrington, and Joshua Conrad Jackson. The strength of social norms across human groups. *Perspectives on Psychological Science*, 12(5):800–809, 2017.
15 James J Gibson. The theory of affordances. *Hilldale, USA*, 1(2), 1977.
16 Erving Goffman. *Interaction ritual: Essays on face-to-face interaction.* Aldine, 1967.
17 Amit Goldenberg, Eran Halperin, Martijn van Zomeren, and James J Gross. The process model of group-based emotion: Integrating intergroup emotion and emotion regulation perspectives. *Personality and Social Psychology Review*, 20(2):118–141, 2016.
18 Michael A Goodrich and Alan C Schultz. Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction*, 1(3):203–275, 2007.
19 Victoria Groom and Clifford Nass. Can robots be teammates?: Benchmarks in human-robot teams. *Interaction Studies*, 8(3):483–500, 2007.

20 Tariq Iqbal, Samantha Rack, and Laurel D Riek. Movement coordination in human-robot teams: a dynamical systems approach. *IEEE Transactions on Robotics*, 32(4):909–919, 2016.

21 Tariq Iqbal and Laurel D Riek. A method for automatic detection of psychomotor entrainment. *IEEE Transactions on Affective Computing*, 7(1):3–16, 2015.

22 Tariq Iqbal and Laurel D Riek. Coordination dynamics in multihuman multirobot teams. *IEEE Robotics and Automation Letters*, 2(3):1712–1717, 2017.

23 Tariq Iqbal and Laurel D Riek. Human-robot teaming: Approaches from joint action and dynamical systems. *Humanoid Robotics: A Reference*, pages 2293–2312, 2019.

24 Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. *The International Journal of Robotics Research*, 37(7):717–742, 2018.

25 Malte F Jung. Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction*, 23(3):18, 2016.

26 Malte F Jung. Affective grounding in human-robot interaction. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction*, pages 263–273. IEEE, 2017.

27 Malte F Jung, Nikolas Martelaro, and Pamela J Hinds. Using robots to moderate team conflict: the case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 229–236. ACM, 2015.

28 Reshmashree B Kantharaju, Alison Pease, Dennis Reidsma, Catherine Pelachaud, Mark Snaith, Merijn Bruijnes, Randy Klaassen, Tessa Beinema, Gerwin Huizing, Donatella Simonetti, et al. Integrating argumentation with social conversation between multiple virtual coaches. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 203–205. ACM, 2019.

29 Ross A. Knepper, Christoforos Mavrogiannis, Julia Proft, and Claire Liang. Implicit communication in a joint action. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 283–292, 2017.

30 Steve WJ Kozlowski and Bradford S Bell. Work groups and teams in organizations. *Handbook of Psychology*, pages 333–375, 2003.

31 Minae Kwon, Mengxi Li, Alexandre Bucquet, and Dorsa Sadigh. Influencing leading and following in human-robot teams. *Proceedings of Robotics: Science and Systems, FreiburgimBreisgau, Germany*, 2019.

32 Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. Gracefully mitigating breakdowns in robotic services. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction*, pages 203–210. IEEE, 2010.

33 Bertram F Malle, Matthias Scheutz, and Joseph L Austerweil. Networks of social and moral norms in human and robot agents. In *A World with Robots*, pages 3–17. Springer, 2017.

34 Christoforos I Mavrogiannis and Ross A Knepper. Multi-agent path topology in support of socially competent navigation planning. *The International Journal of Robotics Research*, 38(2-3):338–356, 2019.

35 Christoforos I. Mavrogiannis, Wil B. Thomason, and Ross A. Knepper. Social momentum: A framework for legible navigation in dynamic multi-agent environments. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 361–369. ACM, 2018.

36 D. Harrison McKnight and Norman L. Chervany. What is trust? A conceptual analysis and an interdisciplinary model. *AMCIS 2000 Proceedings*, page 382, 2000.

37 Bilge Mutlu and Jodi Forlizzi. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, pages 287–294. ACM, 2008.

38 Aastha Nigam and Laurel D Riek. Social context perception for mobile robots. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3621–3627. IEEE, 2015.

39 Stefanos Nikolaidis, David Hsu, and Siddhartha S. Srinivasa. Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research*, 36(5-7):618–634, 2017.

40 Donald A Norman. Affordance, conventions, and design. *interactions*, 6(3):38–43, 1999.

41 Maria Francesca O'Connor and Laurel D Riek. Detecting social context: A method for social event classification using naturalistic multimodal data. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, volume 3, pages 1–7. IEEE, 2015.

42 Harm op den Akker, Rieks op den Akker, Tessa Beinema, Oresti Banos, Dirk Heylen, Björn Bedsted, Alison Pease, Catherine Pelachaud, Vicente Traver Salcedo, Sofoklis Kyriazakos, et al. Council of coaches a novel holistic behavior change coaching approach. In *4th International Conference on Information and Communication Technologies for Ageing Well and e-Health, ICT4AWE 2018*, pages 219–226. SciTePress, 2018.

43 Steffi Paepcke and Leila Takayama. Judging a bot by its cover: an experiment on expectation setting for personal robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction*, pages 45–52. IEEE, 2010.

44 Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 30(3):286–297, 2000.

45 Mina Park, Jae-Jin Song, Seo Jin Oh, Min-Sup Shin, Jun Ho Lee, and Seung Ha Oh. The relation between nonverbal IQ and postoperative CI outcomes in cochlear implant users: Preliminary result. *BioMed Research International*, 2015, 2015.

46 Jens Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, (3):257–266, 1983.

47 Laurel D Riek, Philip C Paul, and Peter Robinson. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, 3(1-2):99–108, 2010.

48 L.D. Riek. The Social Co-Robotics Problem Space: Six Key Challenges. *In Proceedings of Robotics: Science, and Systems, Robotics Challenges and Visions*, 2013.

49 Dorsa Sadigh, S Shankar Sastry, Sanjit A Seshia, and Anca Dragan. Information gathering actions over human internal state. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 66–73. IEEE, 2016.

50 Nicole Salomons, Michael van der Linden, Sarah Strohkorb Sebo, and Brian Scassellati. Humans conform to robots: Disambiguating trust, truth, and conformity. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 187–195. ACM, 2018.

51 Vasanth Sarathy and Matthias Scheutz. A logic-based computational framework for inferring cognitive affordances. *IEEE Transactions on Cognitive and Developmental Systems*, 10(1):26–43, 2016.

52 Matthias Scheutz, Paul Schermerhorn, James Kramer, and Christopher Middendorff. The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, volume 2, pages 226–233, 2006.

53 J Scholtz. Theory and evaluation of human robot interactions. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, page 10 pp. IEEE, 2003.

54 Sarah Strohkorb Sebo, Priyanka Krishnamurthi, and Brian Scassellati. "I don't believe you" investigating the effects of robot trust violation and repair. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction*, pages 57–65. IEEE, 2019.

55 Burr Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.

56 Solace Shen, Petr Slovak, and Malte F Jung. Stop. I see a conflict happening: A robot mediator for young children's interpersonal conflict resolution. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 69–77. ACM, 2018.

57 Elaine Short and Maja J Mataric. Robot moderation of a collaborative game: Towards socially assistive robotics in group interactions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication*, pages 385–390. IEEE, 2017.

58 Tianmin Shu, Michael S Ryoo, and Song-Chun Zhu. Learning social affordance for human-robot interaction. *arXiv preprint arXiv:1604.03692*, 2016.

59 Annika Silvervarg and Kristian Månsson. How do you introduce an agent? The effect of introduction type on how a teachable agent is experienced by students. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 29–34. ACM, 2018.

60 Kristen Stubbs, Pamela J Hinds, and David Wettergreen. Autonomy and common ground in human-robot interaction: A field study. *IEEE Intelligent Systems*, 22(2):42–50, 2007.

61 Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PloS one*, 12(8):e0182151, 2017.

62 Hamish Tennent, Solace Shen, and Malte Jung. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction*, pages 133–142. IEEE, 2019.

63 Edwin J Thomas and Clinton F Fink. Effects of group size. *Psychological bulletin*, 60(4):371, 1963.

64 Marynel Vázquez, Elizabeth J Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E Hudson. Towards robot autonomy in group conversations: Understanding the effects of body orientation and gaze. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 42–52. ACM, 2017.

65 Marynel Vázquez, Alexander May, Aaron Steinfeld, and Wei-Hsuan Chen. A deceptive robot referee in a multiplayer gaming environment. In *2011 International Conference on Collaboration Technologies and Systems (CTS)*, pages 204–211. IEEE, 2011.

66 Susan A Wheelan. Group size, group development, and group productivity. *Small group research*, 40(2):247–262, 2009.

67 Michele Williams. In whom we trust: Group membership as an affective context for trust development. *Academy of Management Review*, 26(3):377–396, 2001.

4.5 Working Group on Scenarios for Human-Agent and Human-Robot Groups

Patrícia Alves-Oliveira (ISCTE-IUL & INESC-ID, PT & Cornell University, US, patricia_alves_oliveira@iscte-iul.pt)
Mary Ellen Foster (University of Glasgow, GB, MaryEllen.Foster@glasgow.ac.uk)
Hatice Gunes (University of Cambridge, GB, hatice.gunes@cl.cam.ac.uk)
Sviatlana Höhn (University of Luxembourg, LU, sviatlana.hoehn@uni.lu)
Franziska Kirstein (Blue Ocean Robotics, DK, fk@blue-ocean-robotics.com)
Andrea Marrella (Sapienza University of Rome, IT, marrella@diag.uniroma1.it)
Mark Neerincx (TU Delft & TNO, NL, mark.neerincx@tno.nl)
Michio Okada (Toyohashi University of Technology, JP, okada@tut.jp)
Ana Paiva (INESC-ID & University of Lisbon, PT, ana.paiva@inesc-id.pt)
Elizabeth Sklar (University of Lincoln, GB, elizabeth.sklar@kcl.ac.uk)
Hannes Högni Vilhjálmsson (Reykjavik University, IS, hannes@ru.is)

4.5.1 Introduction

As social agents and robots become more prominent in our lives, they also increasingly become members of groups and teams, and new scenarios emerge. This document reflects upon the different types of scenarios that are being built and investigated where humans and agents act in groups. It provides a framework for scenarios analysis looking at different dimensions that affect the way the scenarios are conceived, analysed and evaluated.

4.5.2 The ENACTED Framework for scenarios analysis

The *Environment Norms Autonomy Composition Task Embodiment Duration (ENACTED)* framework is designed to provide an organised structure around the definition of scenarios for human-agent and/or human-robot group interaction.

Environment and Domain

The environment conditions the way the groups interact and collaborate. Different environments with the same agents, robots and task may shape and condition the way the group dynamics unfolds. For example, the robots may share the same physical environment as the humans (i.e., *proximal* interaction, such as a robot collecting/delivering items from/to a human in an office, on a farm or in a warehouse); the interaction may involve physically touching each other (e.g., robot lifting a human in a healthcare or rescue operation); the actors may be physically separate from each other (i.e., *remote* interaction, such as a robot surveilling the structure of a nuclear power plant); the "robot" may not have a physical body (i.e., *chatbot*, such as a helper on a web site); the "robot" may be embodied in a virtual platform (i.e., *virtual agent*, such as an agent demonstrating how to conduct a physical task). In situations where the group members are not all physically co-located, there might be multiple settings defined which together comprise the scenario's environment. For example, in the search-and-rescue scenario, you might have a human first-responder who is located in a van outside of a collapsed building while you have a group of robots inside the building searching for victims and providing sensor data to the human team member.

Table 2 Versions of autonomy and delegation.

	Human decides	Robot decides
Human acts	100% human, no robots	Robot transfers a task to a human, e.g when the robot reaches its limits.
Robot acts	Human operator determines what the robot needs to do and when.	100% robot's autonomy, no human needed, but can be included as a controlling instance.

Norms, Social Interactions and Culture

Groups follow certain social norms that shape the way their members interact. In fact, particular groups can even create new norms as the group interactions unfold. Some questions to consider in the scenarios include:

- How are norms defined for behaviour?
- Are norms associated with particular roles/tasks/responsibilities?
- What is the alignment between normative behaviour and expected behaviour when agents and humans act in a group setting?
- Are there actions that are within the norms but are unexpected?

Often, it is considered that the norms for social interaction human groups are also conditioned by cultural learning[7]. What happens if the actions of the agents in a group fall outside what is acceptable?

Degree of Autonomy

There are scenarios in which a robot handles completely autonomously (e.g. killer drone), in others it accomplishes particular tasks that a human explicitly asked it to do (arrange a meeting with a project partner and send a calendar invite), or a robot decides to transfer a task to a human when it reaches its limits (transfer a call to an operator if a chat robot cannot satisfy the caller's request).

We can represent the dimension of autonomy as a combination of two features: autonomy of decisions and autonomy of actions. Table 2 summarises their combinations.

Group Composition

When considering group composition, we need to look at the concepts of outgroup and ingroup first. An outgroup is any group that the entity perceives as not belonging to, while an ingroup is a group that the entity psychologically identifies as being a member of.

Number of entities: The number of entities in the group directly affect the composition of the group. Differently from dyadic interactions, the number of entities in the group need to be considered in relation to other compositional aspects, as well as the other factors included in the ACTED Framework.

Questions that need to be addressed include the following:

- When going from three to four to many number of entities (n), at what point (n=?) the group is formed and the dynamics do not change if one entity leaves/changes or another entity joins the group?

[7] https://journals.sagepub.com/doi/full/10.5772/57260

- How do the dynamics of the group change when we have one vs many agents/robots w.r.t one vs many humans? E.g., One drone approaching a human vs many drones? How does the perception differ if they all look different vs. they all look the same?

Roles: The notion of roles relates to top-down vs. bottom-role assignment or acquisition of a role in a group or team setting. There will be notable differences in assigned team roles compared to acquired team roles. Assigned role is likely to start with a static role (e.g., in a group-based card game setting) and may or may not change (to a different role depending on the group interaction and the dynamics). Instead, acquired role is expected to be formed through the actual interaction taking place. Additionally, we need to consider the roles of the robot and the humans in terms of augmentation or replacement of human abilities.

Questions that need to be addressed here include the following:

- What are the responsibilities that the roles bring along?
- With dynamic roles, what is the means of change (changing roles)? Does change evolve/emerge as the group interacts? Or is there some external factor or norm that dictates the change?
- Does "change" refer to different actors (human, robot) switching between role assignments? Or could it also refer to how the norms/behaviours/responsibilities/actions of the role changes (i.e., the assignment of named role to actor does not change but the actions associated with that role change)?

Homogeneity: The homogeneity vs. heterogeneity of the entities is another factor that directly affects the composition of the group. This in itself is multi-dimensional and would consider various aspects that are used in defining ingroup versus outgroup membership. Examples include not only the commonly used gender (all-female groups, all-robot groups, etc.) and ethnicity (all-white groups vs. all-Asian groups) criteria but also other aspects such as colour (the colour of the robot, the colour of the shirt worn by the team members to indicate belonging), rhythm (how fast/slow one talks, walks, interacts etc.) or other categorisation aspects such as personality (a group consisting of extroverts only vs. a group consisting of both extroverts and introverts). Therefore, the perception of homogeneity or heterogeneity of a group or team can be easily modified using such criteria (e.g. same/similar colours, sounds, movements – fast/slow, rhythm). Stereotypes usually emerge from the tendency to see members of an outgroup as similar (outgroup homogeneity) and members of one's ingroup as different from each other (ingroup heterogeneity).

Questions that need to be addressed include the following:

- How is the composition of the team defined in terms of homogeneity vs. heterogeneity? Does it change? Is it flexible?
- Which factors define hetero/homogeneity beyond roles? I.e., category (robot vs. human), appearance (e.g., orange robot vs. blue robot), behaviour/personality, acceptance (making the human feel in-group or out-group).

Task

One of the dimensions of analysis is the type of task and the environment where the task is being carried out. The task, is related with the type of problem the group is solving: What is the problem that you are trying to solve? Are you addressing a human need? Are you answering fundamental question about human interaction? Are you addressing an identified business need? Are you pursuing a common goal or an individual goal? Are you pursuing a social or task-based goal or a mixture of both?

Types of tasks:

- Informative Tasks (e.g., robot as museum guide or shopping mall guide)
- Competitive Tasks (e.g., robot playing a game with the purpose of winning)
- Collaborative Tasks (e.g., working together to solve a problem or teach a language or complete an action in a factory)
- Action-oriented Tasks (e.g., robot as a butler, or robots in factories)
- Creative Tasks (e.g., playful interactions by acting and reacting)
- Planning Tasks (e.g.,)
- Mixed-motive Tasks (e.g.,)
- Judgemental Tasks (e.g., a jury making a judgement of good vs. bad without acting)

Embodiment and interaction Resources

A presence of a physical robotic body in a group does not automatically mean that this body uses its all capacities in interaction with other group members. In this section we describe the identified types of embodiment and the types of interaction resources that can be associated or made available with the given type of embodiment. For instance, users can talk to a Pepper robot, but it is also possible to use WhatsApp to interact with it.

From chatbots to humanoid robots: Chatbots (text or voice-based) usually do not have a special "body", their interface is usually tailored to process input/output signals. Therefore, for text-based chatbots usually "live" in messengers or webchat, and voice-based chatbots need a microphone for user input and a speaker for output. However, a humanoid robot such as NAO or Pepper can be also used as interfaces for other services, for which the full embodiments is actually not needed.

Interactional Resources: The choice of a particular type of embodiment does not necessarily prescribe to use a specific communication channel. Sometimes it is even desired for the purpose of a particular research work to change or to limit the use of particular communication channel. Therefore, we see it as a separate feature.

Relative position: The relative position of the robot(s) is determined by their size and task. We intuitively found the following four but other may exist:

1. Human is inside a robot (autonomous car, autonomous space shuttle, smart-home);
2. Robot shares space with a human (humanoid robot in a lab, robotic arms in production halls and similar);
3. Robot inside a human (smart medical devices, artificial eye etc.).

Duration of the interaction

We distinguish between long-term and short-term interaction. Categories such as episodic and continuous interaction are included into sort and long-term scenarios.

Short-term interaction: Short-term interaction may include one or multiple *independent* sessions to accomplish short, usually well-defined tasks. Example: robotic information desk in a shopping mall, robot in a hospital lifting a patient.

Even a large number of short subsequent interaction can is classified a short-term as long as these interactions do not form one process or are not parts of a more complex task. As soon as sequential dependencies between subsequent tasks occur, we speak about long-term interaction.

	1: MuMMER	2: Patient Lifting Robot	3: Strawberry Harvesting Robot	4: Promoting Creativity with Robots	5: PAL	6: Gaming / Entertainm. / Science Comm	7: Robots for Emergency Management	8: Social Training	9: Text a Robot
AUTONOMY	Full (respond to approach, initiates)	Semi (controlled by nurse)	Full (Collab. with "farmer-in-the-loop" to learn)	Full (collaborate with children)	Full	Full (Detect gestures and interpret who wins)	Full (Reactive to orders, proactive in task)	Full (react to turn-taking, initiate turns)	Full
COMPOSITION	One robot, many humans, dynamic	One robot, two humans (nurse, patient)	Multiple robots, one human (farmer)	One or more robots, one or more humans (children)	One robot + avatar, many humans (patients, parents, HCP)	One robot, one to five humans (children)	Many robots, many humans. Hetero-geneous	3-5 agents, one human	One robot, many humans
TASK	Guidance (small talk in shopping mall)	Patient-lifting (in hospital)	Strawberry identification	Story creation (creative, but goal-oriented task)	Child diabetes self-managem ent (learning and behavioral change)	Game (playing rock-paper scissor)	Operation in dangerous environments (evaluate damage,remo ve debris)	Improvement of social skills (turn-taking); (human joins group of agents)	Guidance of people, patrolling building
EMBODIMENT	Physical humanoid (Pepper)	Physical non-humanoid (custom robot platform)	Physical non-humanoid (robot platform in poly tunnel)	Physical non-humanoid (toys)	Physical and virtual humanoid (both Nao)	Physical humanoid (Nao)	Physical non-humanoid (tracked wheels and digger)	Virtual humanoid (human-sized)	Physical humanoid + text-based chat
DURATION	Long series, short inter. (<3 min)	Infinite series, short inter. (throughout day)	Long series, cont. inter. changes in composition	Short interaction	Long series (6 mon), short inter. daily (1 - 60 min)	On day series (7 hours), one off short inter. (5 mins)	Infinite series, short interaction	Course series (3 weeks), short inter. (<15 min)	Long series, short interactions (several mins)

Figure 8 Exercise in applying the a chosen set of 5 fairly well defined dimensions to 9 different social group scenarios.

Long-term interaction: Long-term interaction includes multiple sessions with sequential order and sequential dependencies. The order and dependencies can be defined through:

- Information gathered in earlier sessions that is used in later sessions (e.g. names, daily duties of team members, preferences of team members...);
- Level of social proximity (politeness, intimacy) expressed through use of particular interactional resources (invitations, thank-yous and effort for interaction management);
- Conflict and relationship management (e.g. the need to interact for many weeks and a conflict identified in week 1).

4.5.3 Scenarios Evaluation

The scenarios evaluation is available on Figure 8.

4.6 Working Group on Social Behaviours for Group Interactions between Humans and Social Agents and Robots

Elisabeth André (Universität Augsburg, DE, andre@informatik.uni-augsburg.de)
Dirk Heylen (University of Twente, NL, d.k.j.heylen@utwente.nl)
James Kennedy (Disney Research, US, james.r.kennedy@disney.com)
Ana Paiva (INESC-ID & University of Lisbon, PT, ana.paiva@inesc-id.pt)
Catherine Pelachaud (Sorbonne University, FR, catherine.pelachaud@upmc.fr)
Hannes Högni Vilhjálmsson (Reykjavik University, IS, hannes@ru.is)

4.6.1 Introduction and Definitions

When considering social agents, it is first important to have a collective understanding of what is being studied. Breazeal et al. [2] define a social robot as a "robot designed to interact with people in a natural, interpersonal manner". Accordingly, the following criteria for a social agent was established for this working group: "a social agent must have been designed as such". That is to say, the agent must have capabilities for social interaction. An object can be part of a social group, yet it may not be a social agent. Therefore, a vacuum cleaning robot would not be considered social as it has not been explicitly designed with social interaction in mind. However, if the same robot were equipped to understand human conversation and avoid vacuuming when a conversation is taking place, it would then be considered social.

Social agents must communicate using verbal and nonverbal social signals. Such signals include facial expressions, posture, vocal prosody, or spoken language. We then define social behaviours as a higher-level task, like handing over an object or acting towards a goal. A social behaviour can be produced from the combination of multiple social signals. When a group of agents starts to have the same goal, and agrees to this goal, they become a team. The outcome of the actions of the individuals affects all members of the team. In this context, there is a distinction between the social behaviours of the individuals and their behaviours toward the task goal.

4.6.2 Open Problems & Current Challenges

Many research papers have been published that explore social signal and social behaviour analysis between humans (e.g., [3]), and between humans and agents (e.g., [5, 7]). Figure 9, based on [6], provides an overview of the complexity that exists when considering behaviours in groups. Social behaviours form part of this, creating many challenges in considering these processes during design and evaluation of social agents in group settings.

From our understanding of the field gleaned through experience and published works, we identify the open problems and challenges listed below. In particular, we focus on issues for groups, where groups consist of more than two agents (whether human or artificial).

- **Awareness** of the group is important. Which cues signal awareness? Considering the group as an entity, do social behaviours depend on the group identity?
- **Timing** of the cues is extremely important in social behaviours. Sometimes it is not so much about the behaviour itself but *when* to do it. It does not matter what you do (to a certain extent!) provided that you do it at the right moment.

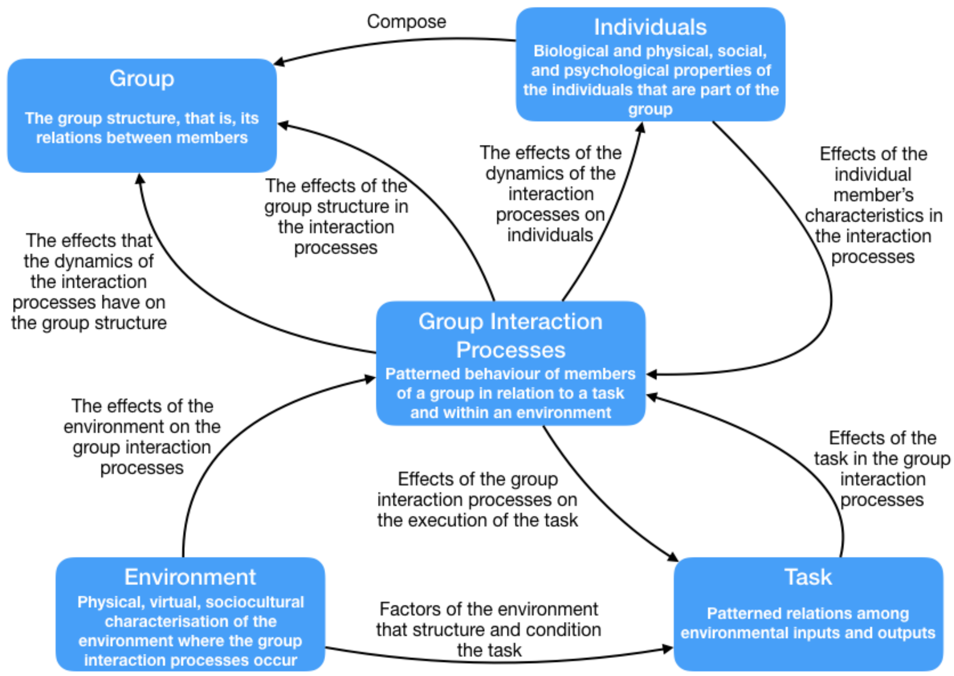

Figure 9 Group Processes based on [6].

- **Social Roles**. Multiple simultaneous roles are possible within a group. Social group hierarchy has an influence on the role.
- **Type of Task** such as creative, competitive, etc. Tasks can be broken into segments, and each has a 'Purpose'. Changing the 'purpose' might also change the social roles.
- **Types of Groups and Context** – Groups are placed in a formation in the world (physical or virtual), but members of the group may additional group structure and membership, like for example a family. A "family" is a group but may not be placed physically in the same F-formation in a given scenario. The influence of this group may still be present in an interaction; how do we model this?
- **Adaptation** – How social behaviours are used and people adapt the social behaviours. How should agents adapt in a group: to the group as a whole, or to individual members of the group? For example, if a robot is a speaker to a class, the members of the class can be considered as a whole. It is not the individual members that matter but rather the whole group. Understanding how to adapt in a way that the identity (personality, character, etc) of the agent is maintained is unclear.
- **Cohesion of a Group** – task or social? How to make a group cohesive? And what is a difference with a team?
- **Emotional Behaviours** affect the team cohesion and performance. Balancing both the positive and negative emotions of individuals is important.
- **Stereotypes** should the behaviours of agents and robots follow and promote certain stereotypes. How do we evaluate when social behaviours may be aligning to stereotypes, and how might we avoid this?
- **Transparency** of the social behaviours. If an agent is being used to influence people, should it explain the behaviours it is using to do so? If it did, how would this affect the influence it has? If the influence is reduced, this may be a negative consequence for agents that aim to positively change people's lives (e.g., a weight loss coach [4]).

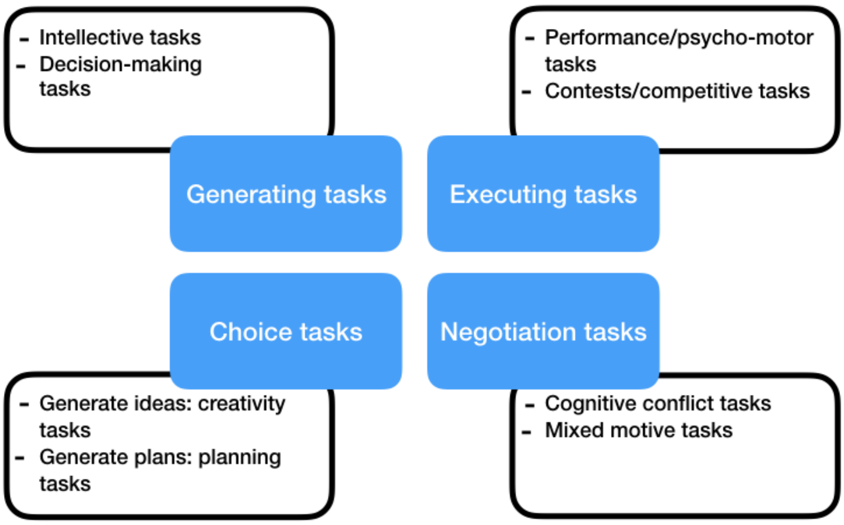

Figure 10 Types of tasks in group interactions based on [6].

4.6.3 Methodologies for Studying Groups

Social skills should enhance interactions with people, to improve the group. Finding the right set of social abilities should be tied to the function that the agent has in the group; we should not be designing just for the sake of being social. Group interactions can be classified differently according to task on demand, see for instance Figure 10. Such frameworks may guide the design of social behaviours for agents in group settings.

Many annotation schemes currently exist for dyadic interactions, i.e., interactions between two agents. These have been used extensively in research to understand social signals and social behaviours between people. This understanding is often transferred to our design of social agents, where the annotation schemes can subsequently be used to evaluate the behaviour of the agent in interactions. However, many complexities arise in applying these same schemes to groups, as any factors that encode behaviours in relation to other group members will increase exponentially with the addition of group members. This is problematic when annotating data manually due to the significant impact on the amount of resources required. Automated analysis as seen in [1] may provide part of the answer, but exploration of alternative schemes is also needed. Many existing schemes either do not capture concepts that arise when studying groups, or are limited in their coding scheme for labelling certain phenomena (for example, synchrony and mimicry in a group is difficult to capture outside of pairwise labelling between individuals). We see the development of annotation schemes and metrics specifically for groups as a key step toward greater understanding of group social behaviour, potentially leading to improved agent interaction capabilities.

References

1 Tobias Baur, Gregor Mehlmann, Ionut Damian, Florian Lingenfelser, Johannes Wagner, Birgit Lugrin, Elisabeth André, and Patrick Gebhard. Context-Aware Automated Analysis and Annotation of Social Human–Agent Interactions. ACM Transactions on Interactive Intelligent Systems (TiiS) 5(2):1–31. ACM, 2015.

2 Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. Social Robotics. In *Springer Handbook of Robotics*, pages 1935–1972. Springer, 2016.

3 Kevin El Haddad, Sandeep Nallan Chakravarthula, and James Kennedy. Smile and Laugh Dynamics in Naturalistic Dyadic Interactions: Intensity Levels, Sequences and Roles. In *Proceedings of the International Conference on Multimodal Interaction*. ACM, 2019.
4 Cory D. Kidd, and Cynthia Breazeal. A robotic weight loss coach. In *Proceedings of the National Conference on Artificial Intelligence*, 22(2):1985. AAAI Press, MIT Press, 2007.
5 Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: a survey. International Journal of Social Robotics 5(2):291–308. Springer, 2013.
6 Joseph Edward McGrath. Groups: Interaction and Performance. Vol. 14. Prentice-Hall, 1984.
7 Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schroeder. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. IEEE Transactions on Affective Computing 3(1):69–87. IEEE, 2011.

Participants

- Patrícia Alves-Oliveira
ISCTE-IUl, INESC-ID – Lisboa, PT & Cornell University – Ithaca, US
- Elisabeth André
Universität Augsburg, DE
- Tony Belpaeme
Ghent University, BE
- Cindy Bethel
Mississippi State University, US
- Merijn Bruijnes
Delft University of Technology, NL
- Filipa Correia
INESC-ID – Porto Salvo & University of Lisbon, PT
- Mary Ellen Foster
University of Glasgow, GB
- Kobi Gal
University of Edinburgh, GB
- Hatice Gunes
University of Cambridge, GB
- Dirk Heylen
University of Twente, NL
- Sviatlana Höhn
University of Luxembourg, LU
- Malte Jung
Cornell University – Ithaca, US
- James Kennedy
Disney Research – Glendale, US
- Franziska Kirstein
Blue Ocean Robotics – Odense, DK
- Stefan Kopp
Universität Bielefeld, DE
- Iolanda Leite
KTH Royal Institute of Technology – Stockholm, SE
- Andrea Marrella
Sapienza University of Rome, IT
- Samuel Mascarenhas
IST – Lisbon, PT
- Christoforos Mavrogiannis
University of Washington – Seattle, US
- Mark Neerincx
TU Delft, NL
- Catharine Oertel
TU Delft, NL
- Michio Okada
Toyohashi University of Technology, JP
- Ana Paiva
INESC-ID – Porto Salvo & University of Lisbon, PT
- Simon Parsons
University of Lincoln, GB
- Catherine Pelachaud
Sorbonne University – Paris, FR
- André Tiago Abelho Pereira
KTH Royal Institute of Technology – Stockholm, SE
- Rui Prada
INESC-ID – Porto Salvo & University of Lisbon, PT
- Laurel Riek
University of California – San Diego, US
- Sarah Sebo
Yale University – New Haven, US
- Julie Shah
MIT – Cambridge, US
- Elaine Short
Tufts University – Medford, US
- Elizabeth Sklar
University of Lincoln, GB
- Marynel Vázquez
Yale University – New Haven, US
- Hannes Högni Vilhjálmsson
Reykjavik University, IS

Report from Dagstuhl Seminar 19421

Quantum Cryptanalysis

Edited by

Michele Mosca[1], Maria Naya-Plasencia[2], and Rainer Steinwandt[3]

1 **University of Waterloo, CA,** michele.mosca@uwaterloo.ca
2 **INRIA – Paris, FR,** maria.naya_plasencia@inria.fr
3 **Florida Atlantic University – Boca Raton, US,** rsteinwa@fau.edu

Abstract

This seminar report documents the program and the outcomes of Dagstuhl Seminar 19421 *Quantum Cryptanalysis*, which took place in October 2019. After outlining the motivation and organizational aspects of this particular seminar, abstracts of presentations that were given by participants are provided.

Seminar October 13–18, 2019 – http://www.dagstuhl.de/19421
2012 ACM Subject Classification Hardware → Quantum technologies, Security and privacy → Cryptanalysis and other attacks, Theory of computation → Computational complexity and cryptography
Keywords and phrases computational algebra, post-quantum cryptography, quantum circuit complexity, quantum computing, standardization
Digital Object Identifier 10.4230/DagRep.9.10.47
Edited in cooperation with Shaun Miller

1 Executive Summary

Michele Mosca
María Naya-Plasencia
Rainer Steinwandt

Motivation and scope

This fifth installment of a Dagstuhl seminar on *Quantum Cryptanalysis* was heavily informed by NIST's ongoing standardization effort in post-quantum cryptography. Several NIST employees attended the seminar and lead a discussion session on the topic. As one would hope hoped for, many talks had an algorithmic focus. Two areas were of particular interest for this seminar:

Quantum cryptanalytic progress. Identifying new cryptanalytic improvements that make use of quantum algorithms and expanding the applicability of the best known cryptanalytic attacks by means of quantum technology. Different quantum attack models can be considered here, and attack models that are close to being realizable with today's technology are particularly relevant. We want to fully leverage quantum computing, including expected mid-term advancements.

Quantum resource estimation. Establishing reasonably precise quantum resource counts for cryptanalytic attacks against symmetric and asymmetric schemes, especially for problem instances and parameter choices that are actually deployed or considered for standardization for future deployment. In addition to logical resources, understanding the overhead caused by handling imperfections of quantum hardware is of interest.

In addition to original quantum cryptanalytic research, the program included presentations with a strong survey component, explaining key concepts of particular areas within post-quantum cryptography. Deviating from prior editions, this time we did not include a presentation to document the status of the development of quantum hardware. Such a talk could have been a welcome addition, but the seminar program was already packed with a substantial number of relevant cryptanalytic results, and it was important to leave sufficient time for discussions.

Organization

Following the organization of the prior quantum cryptanalysis seminars in Dagstuhl, for this fifth edition, again experts from academia, government, and industry came together. We re-invited a number of leading experts in the field from the prior quantum cryptanalysis seminar edition, and at the same time invited several new participants. This included in particular young scientists, who entered this exciting research area more recently. In total, we had with 46 participants a slightly larger number of participants than in the preceding meeting. In line with the Dagstuhl tradition and with prior quantum cryptanalysis seminars, for Wednesday afternoon we left the schedule open. Seminar participants could devote the afternoon to an excursion, to discussions, or to work on their research.

Results and next steps

At this point, communication and collaboration between the classical cryptographic and the quantum algorithmic research communities has become very fruitful, and it seems fair to say that this seminar is also of significant value in supporting ongoing standardization efforts in post-quantum cryptography. In addition to quantum cryptanalytic results on asymmetric cryptography, more results on symmetric cryptography are emerging. There is still substantial research potential – and research need – in quantifying security margins in the presence of quantum computing, and the field keeps moving fast. Improved software tools become available to analyze quantum resources and describe quantum algorithms, bringing research in quantum cryptanalysis closer together with areas in traditional computer science.

2 Table of Contents

Executive Summary
Michele Mosca, María Naya-Plasencia and Rainer Steinwandt 47

Overview of Talks

Challenges in evaluating costs of known lattice attacks
Daniel J. Bernstein 51

On quantum algorithms for isogenies
Jean-François Biasse 51

Modeling the Runtime of Cryptanalytic Algorithms
Christian Bischof 51

The offline Simon's algorithm
Xavier Bonnetain 52

On quantum versions of the Strong Exponential Time Hypothesis
Harry Buhrman 52

On factoring RSA integers and computing discrete logarithms on quantum computers
Martin Ekera 53

Recent results onrank-based cryptography
Philippe Gaborit 53

Some new distributional property testing results
András Gilyén 53

Finding Hash Collisions with Quantum Computers by Using Differential Trials with Smaller Probability than Birthday Bound
Akinori Hosoyamada 54

Using isogenies for post-quantum cryptography
David Jao 54

Improved quantum circuits for modular arithmetic and elliptic curve discrete log
Samuel E. Jaques 54

The Fermat-FHE system
Antoine Joux 55

Quantum speed-ups for sieving algorithms
Elena Kirshanova 55

Quantum circuits for the CSIDH: optimizing quantum evaluation of isogenies
Tanja Lange 56

Quantum Period Finding with a Single Output Qubit-Factoring n-bit RSA with $n/2$ Qubits?
Alexander May 56

Faster provable sieving algoriths for SVP and CVP in ℓ_p norm
Priyanka Mukhopadhyay 57

An attack on LEDAcrypt
Ray Perlner and Daniel C. Apon 57

On the condition number of Macaulay matrices
Rachel Player 57

Quantum speedups for lattice sieves are tenuous at best
John M. Schanck 58

Quantum Merging Algorithms
André Schrottenloher 58

Implementing Grover oracles for quantum key search on AES and LowMC
Fernando Virdia 59

Participants 60

3 Overview of Talks

3.1 Challenges in evaluating costs of known lattice attacks

Daniel J. Bernstein (University of Illinois – Chicago, US)

This talk is a survey of open questions regarding the performance of algorithms in the literature to attack lattice-based cryptosystems.

3.2 On quantum algorithms for isogenies

Jean-François Biasse (University of South Florida – Tampa, US)

In this presentation, we introduce two general frameworks to compute isogenies between elliptic curves using a quantum computer. First, in the more general case, we rely on the quantum search algorithm of Grover. We present recent results showing how to optimize this strategy by classically precomputing short isogeny paths and incorporating this information in the quantum circuit performing the Grover search. Second, we use a subexponential quantum algorithm which is applicable when the endomorphism ring of the elliptic curves involved in the instance of the problem is isomorphic to an imaginary quadratic order. This algorithm relies on the Sieve of Kuperberg which solves the Dihedral Coset Problem. We insist on a recent result showing that we can trade off quantum effort for classical one. This work in progress suggests that there might be hybrid classical/quantum attacks whose circuit size fit under both classical and quantum circuit size limits described by NIST in their standardization process.

3.3 Modeling the Runtime of Cryptanalytic Algorithms

Christian Bischof (TU Darmstadt, DE)

Joint work of Christian Bischof, Michael Burger, Giam Nam Ngyuen

To assess the hardness of lattice-based cryptography, we need to assess the time requitred to solve the shortest vector problem (SVP). Sieving algorithms exhibit exponential complexity in time and space, so actual computational experiments are inherently bounded as to what lattice dimensions can be solved. To extrapolate to larger lattices, we extended the extra-P modeling framework in joint work with Felix' Wolf Group at TU Darmstadt for exponential modeling. In addition to good predictions, it also enables us to investigate tradeoffs with respect to algorithmic parameter selection in sieving algorithms.

3.4 The offline Simon's algorithm

Xavier Bonnetain (INRIA – Paris, FR)

Joint work of Xavier Bonnetain, Akinori Hosoyamada, María Naya-Plasencia, Yu Sasaki, André Schrottenloher

In symmetric cryptanalysis, the model of superposition queries has lead to surprising results, with many constructions being broken in polynomial time thanks to Simon's period-finding algorithm. But the practical implications of these attacks remain blurry. In contrast, the results obtained so far for a quantum adversary making classical queries only are less impressive.

In this paper, we introduce a new quantum algorithm which uses Simon's subroutines in a novel way. We manage to leverage the algebraic structure of cryptosystems in the context of a quantum attacker limited to classical queries and offline quantum computations. We obtain improved quantum-time/classical-data tradeoffs with respect to the current literature, while using only as much hardware requirements (quantum and classical) as a standard exhaustive search using Grover's algorithm. In particular, we are able to break the Even-Mansour construction in quantum time $O(2^{n/3})$, with $O(2^{n/3})$ classical queries and $O(n^2)$ qubits only. In addition, we propose an algorithm that allows to improve some previous superposition attacks by reducing the data complexity from exponential to polynomial, with the same time complexity.

Our approach can be seen in two complementary ways: reusing superposition queries during the iteration of a search using Grover's algorithm, or alternatively, removing the memory requirement in some quantum attacks based on a collision search, thanks to their algebraic structure.

We provide a list of cryptographic applications, including the Even-Mansour construction, the FX construction, some Sponge authenticated modes of encryption, and many more.

3.5 On quantum versions of the Strong Exponential Time Hypothesis

Harry Buhrman (CWI – Amsterdam, NL)

Joint work of Harry Buhrman, Subhasree Patro, and Florian Speelman
Main reference Harry Buhrman, Subhasree Patro, Florian Speelman: "The Quantum Strong Exponential-Time Hypothesis", CoRR, Vol. abs/1911.05686, 2019.
URL https://arxiv.org/abs/1911.05686

The strong exponential-time hypothesis (SETH) is a commonly used conjecture in the field of complexity theory. It states that CNF formulas cannot be analyzed for satisfiability with a speedup over exhaustive search. This hypothesis and its variants gave rise to a fruitful field of research, fine-grained complexity, obtaining (mostly tight) lower bounds for many problems in P whose unconditional lower bounds are hard to find. In this work, we introduce a framework of Quantum Strong Exponential-Time Hypotheses, as quantum analogues to SETH.

Using the QSETH framework, we are able to translate quantum query lower bounds on black-box problems to conditional quantum time lower bounds for many problems in BQP. As an example, we illustrate the use of the QSETH by providing a conditional quantum time lower bound of $\Omega(n^{1.5})$ for the Edit Distance problem. We also show that the n^2 SETH-based lower bound for a recent scheme for Proofs of Useful Work, based on the Orthogonal Vectors problem, also holds for quantum computation assuming QSETH.

3.6 On factoring RSA integers and computing discrete logarithms on quantum computers

Martin Ekera (KTH Royal Institute of Technology – Stockholm, SE)

In this talk, we give an overview of the state of factoring integers and computing discrete logarithms on quantum computers, focusing on algorithms that have been demonstrated to be polynomial time. More specifically, we treat Shor's algorithms for the IFP, OFP and DLP, Seifert's algorithm for the IFP and OFP with tradeoffs, and our algorithms for the short DLP, DLP and RSA IFP, with or without tradeoffs.

We quantify the costs reductions we obtain, both in the logical circuit model, and in a full stack implementation modelled upon existing superconducting quantum computing architectures. The full stack cost estimates are a joint work with Craig Gidney.

We provide tight analyses of the success probabilities in the aforementioned algorithms, classical simulators for the algorithms, and efficient lattice-based post-processing. We show that our algorithms outperform Shor's and Seifert's algorithms for the short DLP and the RSA IFP.

3.7 Recent results onrank-based cryptography

Philippe Gaborit (University of Limoges, FR)

Rank-based cryptography was introduced by Gabidulin et al. in 1991, since then many systems have been proposed. Rank-based cryptography has the inherent good property that the complexity of best known attacks increases faster than for Hamming metric for a given size of key.In this talk we will review recent results on rank-based cryptography, in particular recent submissions to NIST, based on problems with no masking. We will consider the Ouroboros approach and some advanced encryption schemes.

3.8 Some new distributional property testing results

András Gilyén (Caltech – Pasadena, US)

A fundamental problem in statistics and learning theory is to test properties of distributions. We show that quantum computers offer speed-ups for such problems. We describe a natural query input model, that serves as the quantum analog of classical sampling. Then we describe a generic approach that leads to speed-ups for estimating the entropy of distributions, testing equality of two unknown distributions and other problems. Our approach is based on the results of Bravyi, Harrow, and Hassidim (2009), combined with the recent technique of Quantum Singular Value Transformation. The utilized general techniques also allow us to derive similar speed-ups for testing quantum distributions (i.e., density operators). We also show that the quantum speed-ups are at most cubic for classical distributions, as implied by a recent result of Chailloux (2018). Finally, we mention a new result of van Apeldoorn and Montanaro for quadratically speeding-up the estimation of an entire distribution using the quantum Fourier transform.

3.9 Finding Hash Collisions with Quantum Computers by Using Differential Trials with Smaller Probability than Birthday Bound

Akinori Hosoyamada (NTT – Tokyo, JP)

Joint work of Akinori Hosoyamada, Yu Sasaki

We give dedicated quantum attacks on concrete hash functions that exploit their internal structures, which has not received much attention so far. We show collision attacks on 7-round AES-MMO and 6-round Whirlpool, which are not broken (from the view point of collision-resistance) in the classical setting.

3.10 Using isogenies for post-quantum cryptography

David Jao (University of Waterloo, CA)

We give a survey of isogeny-based cryptography and related quantum cryptanalytic techniques.

3.11 Improved quantum circuits for modular arithmetic and elliptic curve discrete log

Samuel E. Jaques (University of Oxford, GB)

Joint work of Samuel E. Jaques, Thomas Naehrig, Michael Häner, Martin Roetteler, Mathias Soeken

Shor's algorithm will break elliptic curve cryptography, but it will require a large quantum computer. How large, exactly? Previous work attempting to quantify the cost has mostly tried to minimize the number of logical qubits; however, error correction overhead could mean that a circuit with fewer gates but more logical qubits would require fewer physical qubits. Thus, we revisit the previous circuits, improve on them, and explore costs besides the minimum qubit count. Our main improvements come from measurement-based "AND" gates, alternative addition circuits, and more efficient pebbling. Our most dramatic result is a circuit depth nearly 10,000 times shorter with only 22% more qubits.

3.12 The Fermat-FHE system

Antoine Joux (Sorbonne University – Paris, FR)

In this talk, we recast state-of-the-art constructions for fully homomorphic encryption in the simple language of arithmetic modulo large Fermat numbers. The techniques used to construct our scheme are quite standard in the realm of (R)LWE based cryptosystems. However, the use of arithmetic in such a simple ring allows to present scheme from elementary mathematical concepts and to implement it easily based on a large number library.

In terms of performance, our test implementation of the proposed scheme is slower than the current speed records but remains within a comparable range. We hope that the detailed study of our Fermat-based scheme by the community can make it even more competitive and provide new insights into FHE constructions at large.

3.13 Quantum speed-ups for sieving algorithms

Elena Kirshanova (Immanuel Kant Baltic Federal University – Kaliningrad, RU)

Joint work of Elena Kirshanova, Erik Mårtensson, Eamonn W. Postlethwaite, Subhayan Roy Moulik
Main reference Elena Kirshanova, Erik Mårtensson, Eamonn W. Postlethwaite, Subhayan Roy Moulik: "Quantum Algorithms for the Approximate k-List Problem and their Application to Lattice Sieving", IACR Cryptology ePrint Archive, Vol. 2019, p. 1016, 2019.
URL https://eprint.iacr.org/2019/1016

The Shortest Vector Problem (SVP) is one of the mathematical foundations of lattice based cryptography. Lattice sieve algorithms are amongst the foremost methods of solving SVP. The asymptotically fastest known classical and quantum sieves solve SVP in a d-dimensional lattice in $2^{cd+o(d)}$ time steps with $2^{c'd+o(d)}$ memory for constants c, c'. In this work, we give various quantum sieving algorithms that trade computational steps for memory.

We first give a quantum analogue of the classical k-Sieve algorithm [Herold–Kirshanova–Laarhoven, PKC'18] in the Quantum Random Access Memory (QRAM) model, achieving an algorithm that heuristically solves SVP in $2^{0.2989d+o(d)}$ time steps using $2^{0.1395d+o(d)}$ memory. This should be compared to the state-of-the-art algorithm [Laarhoven, Ph.D Thesis, 2015] which, in the same model, solves SVP in $2^{0.2653d+o(d)}$ time steps and memory. In the QRAM model these algorithms can be implemented using $poly(d)$ width quantum circuits.

Secondly, we frame the k-Sieve as the problem of k-clique listing in a graph and apply quantum k-clique finding techniques to the k-Sieve.

Finally, we explore the large quantum memory regime by adapting parallel quantum search [Beals et al., Proc. Roy. Soc. A'13] to the 2-Sieve and giving an analysis in the quantum circuit model. We show how to heuristically solve SVP in 20.1037d+o(d) time steps using 20.2075d+o(d) quantum memory.

Category / Keywords: foundations / approximate k-list problem, cryptanalysis, distributed computation, grover's algorithm, lattice sieving, nearest neighbour algorithms, quantum cryptography, shortest vector problem, SVP

3.14 Quantum circuits for the CSIDH: optimizing quantum evaluation of isogenies

Tanja Lange (TU Eindhoven, NL)

URL https://quantum.isogeny.org/

Choosing safe post-quantum parameters for the new CSIDH isogeny-based key-exchange system requires concrete analysis of the cost of quantum attacks. The two main contributions to attack cost are the number of queries in hidden-shift algorithms and the cost of each query. This paper analyzes algorithms for each query, introducing several new speedups while showing that some previous claims were too optimistic for the attacker. This paper includes a full computer-verified simulation of its main algorithm down to the bit-operation level.

3.15 Quantum Period Finding with a Single Output Qubit-Factoring n-bit RSA with $n/2$ Qubits?

Alexander May (Ruhr-Universität Bochum, DE)

Joint work of Alexander May, Lars Schlieper
Main reference Alexander May, Lars Schlieper: "Quantum Period Finding with a Single Output Qubit – Factoring n-bit RSA with n/2 Qubits", CoRR, Vol. abs/1905.10074, 2019.
URL https://arxiv.org/abs/1905.10074

We study quantum period finding algorithms such as Simon and Shor (and its variants Ekerå-Håstad and Mosca-Ekert). For a periodic function f these algorithms produce – via some quantum embedding of f – a quantum superposition $\sum_x |x\rangle|f(x)\rangle$, which requires a certain amount of output qubits that represent $|f(x)\rangle$. We show that one can lower this amount to a single output qubit by hashing f down to a single bit in an oracle setting. Namely, we replace the embedding of f in quantum period finding circuits by oracle access to several embeddings of hashed versions of f. We show that on expectation this modification only doubles the required amount of quantum measurements, while significantly reducing the total number of qubits. For example, for Simon's period finding algorithm in some n-bit function $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$ our hashing technique reduces the required output qubits from n down to 1, and therefore the total amount of qubits from $2n$ to $n + 1$. We also show that Simon's algorithm admits real world applications with only $n + 1$ qubits by giving a concrete realization of a hashed version of the cryptographic Even-Mansour construction. Our oracle-based hashed version of the Ekerå-Håstad algorithm for factoring n-bit RSA reduces the required qubits from $(32 + o(1))n$ down to $(12 + o(1))n$. In principle our hashing approach also works for the Mosca-Ekert algorithm, but requires strong properties of the hash function family. A hashed version of Mosca-Ekert with as few as $O(\log n)$ qubits would imply classical polynomial time factoring. Therefore, the search for suitable hash functions might open a new path to factoring in $\mathcal{P}$.

3.16 Faster provable sieving algorits for SVP and CVP in ℓ_p norm

Priyanka Mukhopadhyay (University of Waterloo, CA)

We give an overview some old and new provable sieving algorithms for SVP and CVP in l_p norm.

3.17 An attack on LEDAcrypt

Ray Perlner (NIST – Gaithersburg, US) and Daniel C. Apon (NIST – Gaithersburg, US)

We present an attack on LEDAcrypt that modifies standard information set decoding to take advantage of the product structure of the private key. By carefully choosing the information set decoding we can explore a much larger fraction of the key space with each iteration. As a result, for all parameters sets there are large classes of weak keys that can be broken faster than expected based on previous analysis and claimed security (e.g. for category 5, $n_0 = 2$, where the attack is most powerful, 1 in 2 to the 45 keys can be broken by an attack costing the equivalent of 2 to the 52 AES operations, and for category 1, $n_0 = 4$, where the attack is least powerful, 1 in about 2 to the 40 keys can be broken by an attack costing the equivalent of 2 to the 50 AES operations.) For some parameter sets (category 5, $n_0 = 2$) the attack most likely also reduces security for the entire key space, although more rigorous analysis is needed to quantify this effect.

3.18 On the condition number of Macaulay matrices

Rachel Player (Royal Holloway University of London, GB),

Joint work of Rachel Player, Jean-Charles Faugère, and Ludovic Perret
Main reference J. C. Faugère, L. Perret, R. Player. On the condition number of Macaulay matrices. In preparation, 2019.

We present a work-in-progress and preliminary results with a view to promote discussion and push forward the work. This work is motivated by the potential impact for cryptanalysis of Chen and Gao's recent quantum algorithm [CG18] for solving Boolean systems of multivariate equations. The complexity of this algorithm depends on the condition number of a certain Macaulay matrix arising from the input Boolean system. The goal of this work is to provide experimental data to determine the size of the condition number in situations of cryptanalytic interest. We also provide a theoretical upper bound on the condition number of a Macaulay matrix that applies in certain cases.

3.19 Quantum speedups for lattice sieves are tenuous at best

John M. Schanck (University of Waterloo, CA)

Joint work of Martin R. Albrecht , Vlad Gheorghiu , Eamonn W. Postlethwaite , John M. Schanck
URL https://materials.dagstuhl.de/files/19/19421/19421.JohnM.Schanck.Preprint.pdf

Quantum variants of lattice sieve algorithms are often used to assess the security of lattice based cryptographic constructions. In this work we provide a heuristic, non-asymptotic, analysis of the cost of several algorithms for near neighbour search on high dimensional spheres. These algorithms are used in lattice sieves. We design quantum circuits for near neighbour algorithms and provide software that numerically optimises algorithm parameters according to various cost metrics. Using this software we estimate the cost of classical and quantum near neighbour search on spheres. We find that quantum search may provide a small speedup in dimensions of cryptanalytic interest, but only under exceedingly optimistic physical and algorithmic assumptions.

3.20 Quantum Merging Algorithms

André Schrottenloher (INRIA – Paris, FR)

Joint work of Maria Naya-Plasencia, Andre Schrottenloher
Main reference María Naya-Plasencia, André Schrottenloher: "Optimal Merging in Quantum k-xor and k-sum Algorithms", IACR Cryptology ePrint Archive, Vol. 2019, p. 501, 2019.
URL https://eprint.iacr.org/2019/501

The k-xor or Generalized Birthday Problem aims at finding, given k lists of bit-strings, a k-tuple among them XORing to 0. If the lists are unbounded, the best classical (exponential) time complexity has withstood since Wagner's CRYPTO 2002 paper. If the lists are bounded (of the same size) and such that there is a single solution, the dissection algorithms of Dinur et al. (CRYPTO 2012) improve the memory usage over a simple meet-in-the-middle. In this paper, we study quantum algorithms for the k-xor problem. With unbounded lists and quantum access, we improve previous work by Grassi et al. (ASIACRYPT 2018) for almost all k. Next, we extend our study to lists of any size and with classical access only. We define a set of "merging trees" which represent the best known strategies for quantum and classical merging in k-xor algorithms, and prove that our method is optimal among these. Our complexities are confirmed by a Mixed Integer Linear Program that computes the best strategy for a given k-xor problem. All our algorithms apply also when considering modular additions instead of bitwise xors. This framework enables us to give new improved quantum k-xor algorithms for all k and list sizes. Applications include the subset-sum problem, LPN with limited memory and the multiple-encryption problem.

3.21 Implementing Grover oracles for quantum key search on AES and LowMC

Fernando Virdia (Royal Holloway University of London, GB)

Joint work of Samuel Jaques, Michael Naehrig, Martin Roetteler, Fernando Virdia
Main reference Samuel Jaques, Michael Naehrig, Martin Roetteler, Fernando Virdia: "Implementing Grover oracles for quantum key search on AES and LowMC", IACR Cryptology ePrint Archive, Vol. 2019, p. 1146, 2019.
URL https://eprint.iacr.org/2019/1146

Grover's search algorithm gives a quantum attack against block ciphers by searching for a key that matches a small number of plaintext-ciphertext pairs. This attack uses O(sqrt(N)) calls to the cipher to search a key space of size N. Previous work in the specific case of AES derived the full gate cost by analyzing quantum circuits for the cipher, but focused on minimizing the number of qubits.

In contrast, we study the cost of quantum key search attacks under a depth restriction and introduce techniques that reduce the oracle depth, even if it requires more qubits. As cases in point, we design quantum circuits for the block ciphers AES and LowMC. Our circuits give a lower overall attack cost in both the gate count and depth-times-width cost models. In NIST's post-quantum cryptography standardization process, security categories are defined based on the concrete cost of quantum key search against AES. We present new, lower cost estimates for each category, so our work has immediate implications for the security assessment of post-quantum cryptography.

As part of this work, we release Q# implementations of the full Grover oracle for AES-128, -192, -256 and for the three LowMC instantiations used in Picnic, including unit tests and code to reproduce our quantum resource estimates. To the best of our knowledge, these are the first two such full implementations and automatic resource estimations.

Participants

- Gorjan Alagic
University of Maryland – College Park, US
- Daniel C. Apon
NIST – Gaithersburg, US
- Daniel J. Bernstein
University of Illinois – Chicago, US
- Jean-François Biasse
University of South Florida – Tampa, US
- Christian Bischof
TU Darmstadt, DE
- Xavier Bonnetain
INRIA – Paris, FR
- Harry Buhrman
CWI – Amsterdam, NL
- Jintai Ding
University of Cincinnati, US
- Martin Ekera
KTH Royal Institute of Technology – Stockholm, SE
- Philippe Gaborit
University of Limoges, FR
- András Gilyén
Caltech – Pasadena, US
- Maria Isabel González Vasco
King Juan Carlos University – Madrid, ES
- Sean Hallgren
Pennsylvania State University – University Park, US
- Akinori Hosoyamada
NTT – Tokyo, JP
- David Jao
University of Waterloo, CA
- Samuel E. Jaques
University of Oxford, GB
- Stacey Jeffery
CWI – Amsterdam, NL
- Antoine Joux
Sorbonne University – Paris, FR
- Elena Kirshanova
Immanuel Kant Baltic Federal University – Kaliningrad, RU
- Thijs Laarhoven
TU Eindhoven, NL
- Bradley Lackey
Microsoft Corporation – Redmond, US
- Tanja Lange
TU Eindhoven, NL
- Alexander May
Ruhr-Universität Bochum, DE
- Shaun Miller
Florida Atlantic University – Boca Raton, US
- Dustin Moody
NIST – Gaithersburg, US
- Michele Mosca
University of Waterloo, CA
- Priyanka Mukhopadhyay
University of Waterloo, CA
- Maria Naya-Plasencia
INRIA – Paris, FR
- Phong Q. Nguyen
ENS – Paris, FR
- Ray Perlner
NIST – Gaithersburg, US
- Edoardo Persichetti
Florida Atlantic University – Boca Raton, US
- Rachel Player
Royal Holloway University of London, GB
- Thomas Pöppelmann
Infineon Technologies AG – Neubiberg, DE
- Yu Sasaki
NTT – Tokyo, JP
- John M. Schanck
University of Waterloo, CA
- André Schrottenloher
INRIA – Paris, FR
- Nicolas Sendrier
INRIA – Paris, FR
- Yixin Shen
Paris Diderot University, FR
- Daniel C. Smith-Tone
NIST – Gaithersburg, US
- Rainer Steinwandt
Florida Atlantic University – Boca Raton, US
- Adriana Suárez Corona
University of León, ES
- Jean-Pierre Tillich
INRIA – Paris, FR
- Iggy van Hoof
TU Eindhoven, NL
- Fernando Virdia
Royal Holloway University of London, GB
- Thomas Wunderer
BSI – Bonn, DE
- Bo-Yin Yang
Academia Sinica – Taipei, TW

Report from Dagstuhl Seminar 19431

Theory of Randomized Optimization Heuristics

Edited by

Carola Doerr[1], Carlos M. Fonseca[2], Tobias Friedrich[3], and Xin Yao[4]

1 **Sorbonne University – Paris, FR,** carola.doerr@mpi-inf.mpg.de
2 **University of Coimbra, PT,** cmfonsec@dei.uc.pt
3 **Hasso Plattner Institute – Potsdam, DE,** friedrich@hpi.de
4 **Southern Univ. of Science and Technology – Shenzen, CN,** xiny@sustech.edu.cn

Abstract

This report documents the activities of Dagstuhl Seminar 19431 on "Theory of Randomized Optimization Heuristics". 46 researchers from Europe, Australia, Asia, and North America have come together to discuss ongoing research. This tenth edition of the seminar series had three focus topics: (1) relation between optimal control and heuristic optimization, (2) benchmarking optimization heuristics, and (3) the interfaces between continuous and discrete optimization. Several breakout sessions have provided ample opportunity to brainstorm on recent developments in the research landscape, to discuss and solve open problems, and to kick-start new research initiatives.

Seminar October 20–25, 2019 – http://www.dagstuhl.de/19431

2012 ACM Subject Classification Theory of computation → Bio-inspired optimization, Theory of computation → Evolutionary algorithms, Theory of computation → Optimization with randomized search heuristics

Keywords and phrases algorithms and complexity, evolutionary algorithms, machine learning, optimization, soft computing

Digital Object Identifier 10.4230/DagRep.9.10.61

Edited in cooperation with Patrick Spettel

1 Executive Summary

Carola Doerr
Carlos M. Fonseca
Tobias Friedrich
Xin Yao

Efficient optimization techniques affect our personal, industrial, and academic environments through the supply of well-designed processes that enable a best-possible use of our limited resources. Despite significant research efforts, most real-world problems remain too complex to admit exact analytical or computational solutions. Therefore, heuristic approaches that trade the accuracy of a solution for a simple algorithmic structure, fast running times, or an otherwise efficient use of computational resources are required. Randomized optimization heuristics form a highly successful and thus frequently applied class of such problem solvers. Among the best-known representatives of this class are stochastic local search methods, Monte Carlo techniques, genetic and evolutionary algorithms, and swarm intelligence techniques.

The theory of randomized optimization heuristics strives to set heuristic approaches on firm ground by providing a sound mathematical foundation for this important class of algorithms. Key challenges in this research area comprise optimization under uncertainty, parameter selection (most randomized optimization heuristics are parametrized), the role and usefulness of so-called *crossover* operations (i.e., the idea of creating high-quality solution candidates by recombining previously evaluated ones) and, more generally, performance guarantees for advanced heuristics such as population-based techniques, estimation-of-distribution algorithms, differential evolution, and others.

Dagstuhl Seminar 19431 on "Theory of Randomized Optimization Heuristics" was a continuation of the seminar series originally on "Theory of Evolutionary Algorithms". Today the field extends far beyond evolutionary algorithms – a development that previous Dagstuhl seminars have significantly influenced.

While the previous seminar 17191 had a very strong focus on methodological questions and techniques needed to analyze stochastic optimization heuristics, the present seminar had among its three main focus topics chosen to foster interaction with two strongly linked research communities that were not previously represented in the seminar series: stochastic control theory and empirical benchmarking of randomized optimization heuristics.

Recent work has shown that there is a very close link between the theory of randomized optimization heuristics and stochastic control theory, both regarding the nature of the "systems" of interest and the analytical techniques that have been developed in the two communities. At the seminar, we have explored these affinities through the two invited presentations of Luc Pronzato and Vivek Borkar, through contributed talks highlighting different aspects studied in both communities (e.g., the presentation on one-shot optimization by Olivier Teytaud), and through focussed breakout sessions, in particular the one fully dedicated to *Connection between the analysis of evolution strategies and estimation of distribution algorithms and the analysis of stochastic approximation and ordinary differential equations*, in which interesting similarities and differences between the two fields were identified.

The second focus topic of Dagstuhl Seminar 19431 was benchmarking of optimization heuristics. Benchmarking plays a central role in empirical performance assessment. However, it can also be an essential tool for theoreticians to develop their mathematically-derived ideas into practical algorithms, thereby encouraging a principled discussion between empirically-driven and theoretically-driven researchers. Benchmarking has been a central topic in several breakout sessions, for example those on *Competitions and Benchmarking*, *Algorithm Selection and Configuration*, but also the breakout session on *Multi-Objective Optimization.* A survey of best practices in empirical benchmarking has been kick-started in the breakout session on *Benchmarking: Best Practices and Open Issues.*

Discussing the mathematical challenges arising in the performance analysis of randomized heuristics has always been a central topic in this Dagstuhl seminar series. Among other achievements, important connections between continuous and discrete optimization have been established, most notably in the form of drift theorems, which are typically applicable regardless of the nature of the search space. Apart from such methodological advances, we have also observed two other trends bridging discrete and continuous optimization: (i) an increased interest in analyzing parameter-dependent performance guarantees, and (ii) the recent advances in the study of estimation of distribution algorithms, which borrow techniques from both discrete and continuous optimization theory. These topics have been discussed in the invited talk of Youhei Akimoto, in several contributed presentations, and in the breakout sessions on *Measuring Optimization Progress in an Invariant Way for Comparison-Based Algorithms* and on *Mixed-Integer Optimization.*

Apart from these focus topics, we have discussed a large number of different aspects related to the theoretical analysis of optimization heuristics, including brainstorming sessions on doing "good" research, organizing a repository to share lecture materials, and discussing the role of uncertainty in heuristic optimization, the connections between experimental design and one-shot optimization, the importance of neutral representations, and differences between stochastic gradient descent methods and evolution strategies, to give but a few examples.

Organization

The seminar hosted the following type of events:

- Five invited talks of 30 minutes each:
  - Youhei Akimoto on *Expected Runtime Bound for the (1+1)-Evolution Strategy*
  - Vivek Borkar on *Overview of Stochastic Approximation and Related Schemes*
  - Pietro S. Oliveto on *What is Hot in Evolutionary Computation (Part 2)*
  - Luc Pronzato on *Dynamical Search*
  - Carsten Witt on *What is Hot in Evolutionary Computation (Part 1)*
- 20 contributed talks of around 15-20 minutes
- Four "flash talks" of about 10 minutes
- Eleven parallel breakout sessions in various different formats, ranging from brainstorming on the purpose and future of theory research through actual problem solving on one-shot optimization to kick-starting a survey on best practices on benchmarking optimization heuristics.

All presentations were plenary, i.e., in a single session, while the breakouts were organized in parallel working groups, to allow for focused and specialized discussions. As in previous years, the breakout sessions were very well perceived, and can be considered a well-established format of this seminar series. As a result of these discussions, we are planning a workshop and a survey on benchmarking best practices. Several open problems have been proposed and discussed at the seminar, and we are confident that the seminar has helped to establish new collaborations.

Our traditional hike on Wednesday was a good opportunity to discuss in a less formal setting and to get to know each other. On Thursday evening, we had the special opportunity to hear Jonathan Rowe present activities of the Alain Turing Institute https://www.turing.ac.uk/, where he serves as Programme Director for Data Science for Science. Last, but not least, the wine-and-cheese party complemented the scientific activities with a relaxed social event.

We would like to thank the Dagstuhl team and all participants for making seminar 19431 a great success and a great pleasure to organize.

Carola Doerr (Sorbonne University – Paris, FR)
Carlos M. Fonseca (University of Coimbra, PT)
Tobias Friedrich (Hasso Plattner Institute – Potsdam, DE)
Xin Yao (Southern University of Science and Technology – Shenzen, CN)

2 Table of Contents

Executive Summary
Carola Doerr, Carlos M. Fonseca, Tobias Friedrich, and Xin Yao 61

Overview of Talks

Expected Runtime Bounds for $(1+1)$-ES
Youhei Akimoto 67

Precise Analysis for Plateaus
Denis Antipov and Benjamin Doerr 67

A Unified Invariance Formalism for Discrete and Continuous Optimization
Anne Auger 67

Evolution Strategies are NOT Gradient Followers
Hans-Georg Beyer 68

Stochastic Approximation: an Overview
Vivek Shripad Borkar 68

Variations on the Theme of the $(1+(\lambda,\lambda))$ GA
Maxim Buzdalov 69

Challenges of Mutation Rate Control in $(1+\lambda)$ Evolutionary Algorithm
Arina Buzdalova 69

Dynastic Potential Crossover Operator
Francisco Chicano 70

Adaptation of a Sampling Distribution for Metropolis-Hastings
Alexandre Chotard 70

Genetic Drift in EDAs
Benjamin Doerr 70

On Potential for Transfer of Results from Theory of Evolutionary Algorithms to Biology
Anton V. Eremeev 71

A (General) Definition of Invariance
Nikolaus Hansen 71

Gradient Descent and Evolution Strategies are Almost the Same (but don't behave almost the same)
Nikolaus Hansen 72

The UMDA on LeadingOnes Revisited
Martin S. Krejca 72

Runtime Analysis of Self-adaptive EAs
Per Kristian Lehre 72

Dynamic Linear Functions
Johannes Lengler 72

Automated Algorithm Configuration and Selection for Theoreticians
Manuel López-Ibánez 73

What's hot in EA theory II
Pietro S. Oliveto . . . 73

Dynamical Search: a Short Introduction
Luc Pronzato . . . 73

Open problems relating to Noisy OneMax
Jonathan E. Rowe . . . 74

Work at the Alan Turing Institute on "The Data Science Revolution in Scientific Research"
Jonathan E. Rowe . . . 74

Runtime in Integer Space Under Multiple Objectives
Günter Rudolph . . . 74

Analysis of Evolution Strategies Applied to a More General Conically Constrained Problem
Patrick Spettel . . . 74

Runtime Analysis of Probabilistic Crowding – Results beyond OneMax
Dirk Sudholt . . . 75

On the Linkage Equilibria of Weakly-Selective Steady-State GAs
Andrew M. Sutton . . . 75

Single-Iteration Evolutionary Computation (aka Fully Parallel Derivative-Free Optimization)
Olivier Teytaud . . . 75

Optimal Mixing Evolutionary Algorithms
Dirk Thierens . . . 76

Analysis of Artificial Genetic Representations with Neutrality
Vida Vukašinović . . . 76

What's Hot In EA Theory I
Carsten Witt . . . 76

Stochastic global optimization (SGO)
Anatoly Zhigljavsky . . . 77

Working groups

Breakout Session: Benchmarking – Best Practices and Open Issues
Thomas Bartz-Beielstein . . . 77

Breakout Session: Multiobjective Optimization
Dimo Brockhoff, Benjamin Doerr, Carola Doerr, Manuel López-Ibánez, Rolf H. Möhring, Günter Rudolph, Dirk Thierens, Markus Wagner, and Elizabeth Wanner 77

Breakout Session: Mixed-Integer-Nominal Optimization
Thomas Bäck . . . 79

Breakout Session: Open Problems
Benjamin Doerr and Frank Neumann . . . 79

Breakout Session: Analysis of Artificial Genetic Representations with Neutrality
Carlos M. Fonseca and Vida Vukašinović . . . 80

Breakout Session: Connection Between ES / EDA Analysis and Stochastic Approximation / ODE Theory
Tobias Glasmachers . . . 81

Breakout Session: Measuring Optimization Progress in an Invariant Way for Comparison-Based Algorithms
Tobias Glasmachers . . . 81

Breakout Session: Invariance
Nikolaus Hansen and Anne Auger . . . 81

Breakout Session: Drift Theory
Martin S. Krejca . . . 82

Breakout Session: Passing It On
Timo Kötzing . . . 82

Breakout Session: The Purpose of Theory Research
Timo Kötzing . . . 83

Breakout Session: Competitive Co-evolution
Per Kristian Lehre . . . 84

Breakout Session: Dynamic Linear Functions
Johannes Lengler . . . 86

Breakout Session: Algorithm Configuration and Selection
Pietro S. Oliveto . . . 87

Breakout Session: Competitions and Benchmarking
Olivier Teytaud and Carola Doerr . . . 87

Breakout Session: One-Shot Optimization
Olivier Teytaud and Carola Doerr . . . 88

Breakout Session: Permutation-based Problems
Christine Zarges . . . 88

Schedule . . . 89

Participants . . . 94

3 Overview of Talks

3.1 Expected Runtime Bounds for $(1+1)$-ES

Youhei Akimoto (University of Tsukuba, JP)

Main reference Daiki Morinaga, Youhei Akimoto: "Generalized drift analysis in continuous domain: linear convergence of (1 + 1)-ES on strongly convex functions with Lipschitz continuous gradients", in Proc. of the 15th ACM/SIGEVO Conference on Foundations of Genetic Algorithms, FOGA 2019, Potsdam, Germany, August 27-29, 2019, pp. 13–24, ACM, 2019.
URL https://doi.org/10.1145/3299904.3340303
Main reference Youhei Akimoto, Anne Auger, Tobias Glasmachers: "Drift theory in continuous search spaces: expected hitting time of the (1 + 1)-ES with 1/5 success rule", in Proc. of the Genetic and Evolutionary Computation Conference, GECCO 2018, Kyoto, Japan, July 15-19, 2018, pp. 801–808, 2018.
URL https://doi.org/10.1145/3205455.3205606

We presented recent results on the expected runtime bound for (1+1)-ES by drift analysis. We focused on what we want to show, what has been done, and what are still open. In the end, we had a discussion on what are the similarities and dissimilarities between drift analysis in continuous domain and discrete domain.

3.2 Precise Analysis for Plateaus

Denis Antipov (ITMO University – St. Petersburg, RU) and Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

Main reference Denis Antipov, Benjamin Doerr: "Precise Runtime Analysis for Plateaus", CoRR, Vol. abs/1806.01331, 2018.
URL https://arxiv.org/abs/1806.01331

Local optima and plateaus are the features of the fitness landscape which usually make a fitness function hard to be optimized by random search heuristics. While there are plenty of works considering the problem of escaping local optima, most of which are based on the jump functions, plateaus have not got this much attention in the community. In this talk we consider our results on analysis of the simple mutation-based EAs on a benchmark PLATEAU_k function, introduce the new methods of the analysis for the plateaus and discuss what are the obstacles for spreading these methods on the more complex algorithms.

3.3 A Unified Invariance Formalism for Discrete and Continuous Optimization

Anne Auger (INRIA Saclay – Palaiseau, FR)

Invariance is a general concept that is fundamental in many domains. In statistics, machine learning, decisions taken as a result of a procedure/algorithm based on data should not be affected by simple transformations on the input data like reordering or translation. Invariance is also essential in optimization where we do not want the performance of an algorithm to be greatly affected if e.g. the function optimized is translated or scaled by a positive factor. In this talk I will give a (unified) definition of invariance in the search space that holds in particular for discrete and continuous domains.

3.4 Evolution Strategies are NOT Gradient Followers

Hans-Georg Beyer (Fachhochschule Vorarlberg – Dornbirn, AT)

This talk addresses the question how Evolution Strategies (ES) explore high-dimensional $\mathbb{R}^N$ search spaces. A sometimes invoked picture tries to explain the working by some kind of gradient following strategy. On the other hand there are optimization algorithms that are labeled as ES, however, are actually gradient approximation strategies. It is shown that one of these novel "ESs" resembles the well-known Evolutionary Gradient Search strategy, published in the late 1990s by R. Salomon. Coming back to the question whether classical ESs are gradient approximating strategies, it is shown that this picture does not hold, neither when considering the search behavior of the population in the search space nor when investigating the mean value dynamics of the search process. It turns out that the ES devotes only small steps toward the optimizer in the search space while performing large step in the perpendicular $(N-1)$-dimensional subspace. The one-dimensional part, responsible for the fitness improvement, may be regarded as the exploitation part of the search process while the step in the perpendicular subspace may be regarded as exploration. The ratio of these two steps is roughly proportional to $1/\sqrt{N}$. This is different from vanilla gradient strategies and might explain in parts the success of ESs in highly multimodal optimization problems.

3.5 Stochastic Approximation: an Overview

Vivek Shripad Borkar (Indian Institute of Technology Bombay, IN)

The talk introduces the Robbins-Monro "stochastic approximation" algorithm and the "o.d.e." (for "ordinary differential equations") approach for its convergence analysis. Other theoretical issues such as avoidance of unstable equilibria, limit theorems, stability of iterates, etc. are briefly discussed. Variants such as constant stepsize, two time scale schemes, Markov noise, differential inclusion limits, and distributed asynchronous schemes are mentioned. As example, stochastic gradient and gradient-like schemes are presented. Finally, consensus algorithms are briefly discussed.

3.6 Variations on the Theme of the $(1 + (\lambda, \lambda))$ GA

Maxim Buzdalov (ITMO University – St. Petersburg, RU)

Joint work of Maxim Buzdalov, Anton Bassin

The $(1 + (\lambda, \lambda))$ genetic algorithm is an interesting theory-driven algorithm with many good properties, e.g. the $O(n)$ runtime on OneMax and the showcase of self-adjusting parameter tuning being a crucial part. However, it is still quite slow to conquer other territories, e.g. it is not a very brilliant player even for linear functions. This talk presents our preliminary work on changing this situation. In particular, we introduce a rather successful extension of the $(1 + (\lambda, \lambda))$ GA on problems defined on permutations, and show a few interesting consequences of that regarding how to understand the driving forces behind this algorithm. Another orthogonal idea is that the selection in this algorithm may be rethought based on statistical ideas. This instantly leads to the $O(n)$ runtime on the BinVal function and might pave the road towards wider applicability of this wonderful algorithm.

3.7 Challenges of Mutation Rate Control in $(1 + \lambda)$ Evolutionary Algorithm

Arina Buzdalova (ITMO University – St. Petersburg, RU)

Joint work of Arina Buzdalova, Kirill Antonov, Maxim Buzdalov, Carola Doerr, Irina Petrova, Anna Rodionova
Main reference Anna Rodionova, Kirill Antonov, Arina Buzdalova, Carola Doerr: "Offspring population size matters when comparing evolutionary algorithms with self-adjusting mutation rates", in Proc. of the Genetic and Evolutionary Computation Conference, GECCO 2019, Prague, Czech Republic, July 13-17, 2019, pp. 855–863, 2019.
URL https://doi.org/10.1145/3321707.3321827

It was empirically observed that efficiency of mutation rate control in $(1 + \lambda)$ EA depends on the specified lower bound. Particularly, with the growth of population size λ a higher mutation rate bound of $1/n$ is more efficient than $1/n^2$. However, it seems sensible that a successful adjustment mechanism should not be harmed by a more generous lower bound. We propose a simple modification that makes the 2-rate $(1 + \lambda)$ EA [1] much less sensitive to the lower bound. The open question is how to capture this improvement by theoretical analysis.

References

1 B. Doerr, C. Gießen, C. Witt, and J. Yang, "The $(1 + \lambda)$ Evolutionary Algorithm with Self-Adjusting Mutation Rate", Algorithmica 81, 2 (2019), 593–631.

3.8 Dynastic Potential Crossover Operator

Francisco Chicano (University of Málaga, ES)

An optimal recombination operator provides an optimal solution fulfilling the gene transmission property: the value of any variable in the offspring must be inherited from one of the parents. In the case of binary variables, the offspring of an optimal recombination operator is optimal in the smallest hyperplane containing the two parent solutions. In general, exploring this hyperplane is computationally costly, but if the objective function has a low number of nonlinear interactions among the variables, the exploration can be done in $O(4^\beta(n+m)+n^2)$ time, for problems with n decision variables, m subfunctions composing the objective function and where β is a constant. In this talk, we present a quasi-optimal recombination operator, called Dynastic Potential Crossover (DPX), that runs in $O(4^\beta(n+m)+n^2)$ time in any case and is able to act as an optimal recombination operator for low-epistasis combinatorial problems. We show some experimental results where the operator is integrated in DRILS (an ILS with recombination) and standard EA solving NKQ Landscapes and MAX-SAT.

3.9 Adaptation of a Sampling Distribution for Metropolis-Hastings

Alexandre Chotard (Calais University, FR)

A Metropolis-Hastings algorithm aims to emulate sampling from a target probability distribution π by using a proposal distribution q. As in optimization, one may adapt the proposal distribution using the points sampled so far, but one also has to care not to bias the resulting stationary distribution.

3.10 Genetic Drift in EDAs

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

Joint work of Benjamin Doerr, Weijie Zheng
Main reference Benjamin Doerr, Weijie Zheng: "Sharp Bounds for Genetic Drift in EDAs", CoRR, Vol. abs/1910.14389, 2019.
URL https://arxiv.org/abs/1910.14389

It has been observed in various mathematical runtime analyses of estimation-of-distribution algorithms that also in the complete absence of a fitness signal, the sampling distributions of the solution values develop a strong preferences for a single value. In this work, we quantify precisely this so-called genetic drift for the univariate EDAs cGA and PBIL (which includes the UMDA and the MMAS ant colony optimizer). Our results suggest how to choose the parameters of these algorithms such as to avoid genetic drift, which is useful both in applications and in research.

3.11 On Potential for Transfer of Results from Theory of Evolutionary Algorithms to Biology

Anton V. Eremeev (Institute of Scientific Information for Social Sciences RAS – Moscow, RU)

The well-known biotechnological procedure SELEX (Systematic Evolution of Ligands by EXponential enrichment) is considered as an experimental implementation of an evolutionary algorithm (EA). As a proof of concept, theoretical bounds on the expected EA runtime and on fraction of sufficiently fit individuals in population are applied in order to forecast the efficiency of SELEX in searching for a promoter sequence, including an enhancer. A comparison of theoretical bounds to the results of computational simulation indicates some cases where the theoretical runtime bounds and bounds on the frequency of highly fit individuals give favorable prediction, while simulation requires prohibitive computational resource. It is shown that further research is needed to extend applicability of the theoretical bounds for the expected runtime and to improve their tightness.

3.12 A (General) Definition of Invariance

Nikolaus Hansen (INRIA Saclay – Palaiseau, FR)

Invariance is arguably one of the single most important conceptions in science. Here, we attempt to give a concise definition of invariance in the context of randomized search algorithms.

▶ **Definition 1** (General Invariance). *Let $\mathcal{F}$ be the set of all functions on a given search space and $\mathcal{H}$ a mapping of $\mathcal{F}$ into its power set,*

$$\mathcal{H} : f \mapsto \mathcal{H}(f) \subset \mathcal{F} \text{ and } f \in \mathcal{H}(f) \text{ w.l.o.g. } .$$

We say that a search algorithm is invariant under $\mathcal{H}$ if for every pair of functions $f, h \in \mathcal{H}(f)$ there exists

- *a bijective search space transformation $\varphi_{f\to h}$ and*
- *for all (initial) algorithm states θ on f*
  - *a reachable (initial) state θ' on h and*
  - *a coupling for the random input*

such that for all time steps t the evaluated solutions on h and f, that is the search traces, are equivalent in that

$$x_t^{\theta',h} = \varphi_{f\to h}\left(x_t^{\theta,f}\right) \;.$$

3.13 Gradient Descent and Evolution Strategies are Almost the Same (but don't behave almost the same)

Nikolaus Hansen (INRIA Saclay – Palaiseau, FR)

Modifying the update equations of the iterate in gradient descent to become the update equations of the mean in evolution strategies requires only three surprisingly small changes when given a suitable representation. This insight allows to concisely scrutinize the possible reasons why (and when) gradient descent and evolution strategies behave very differently.

3.14 The UMDA on LeadingOnes Revisited

Martin S. Krejca (Hasso Plattner Institute – Potsdam, DE)

This talk will showcase some of the joint and ongoing work with Benjamin Doerr. When considering the univariate marginal distribution algorithm (UMDA) in a parameter regime with low genetic drift, it can be easily analyzed on the LeadingOnes function. This simplified analysis also improves the currently best known run time bound of the UMDA for that parameter regime.

3.15 Runtime Analysis of Self-adaptive EAs

Per Kristian Lehre (University of Birmingham, GB)

We will present ongoing work on runtime analysis of self-adaptive evolutionary algorithms.

3.16 Dynamic Linear Functions

Johannes Lengler (ETH Zürich, CH)

I believe that this type of test functions are interesting, fruitful, and tractable. I will explain what they are, why I believe that it's worth studying them, and what the background is. I would like to invite others to study them, either as collaboration or independently. After the talk, there will also be a breakout session on the topic.

Related papers are:
- https://doi.org/10.1109/SSCI.2018.8628785
- https://ieeexplore.ieee.org/abstract/document/8715464
- https://doi.org/10.1145/3299904.3340309

3.17 Automated Algorithm Configuration and Selection for Theoreticians

Manuel López-Ibánez (University of Manchester, GB)

High-performing optimizers have many parameters that need to be configured. There are many benefits of automatically configuring these parameters. The problem of automatic algorithm configuration (AC) is described in a formal mathematical manner, together with a brief description of irace, which is one method for tackling it. In addition, the problem of automatic algorithm selection (AS) is described and connected to automatic algorithm configuration. Finally, topics within AC/AS of potential interest for theoreticians are highlighted with links to recent works in this direction.

3.18 What's hot in EA theory II

Pietro S. Oliveto (University of Sheffield, GB)

In this introductory talk we provided an overview of recent trends, techniques and challenges in the theoretical runtime analysis of bio-inspired optimisation heuristics. We covered the latest results and open problems concerning generational and steady-state genetic algorithms and artificial immune systems.

3.19 Dynamical Search: a Short Introduction

Luc Pronzato (Laboratoire I3S – Sophia Antipolis, FR)

Joint work of Luc Pronzato, Henry P. Wynn, Anatoly A. Zhigljavsky
Main reference Luc Pronzato, Henry P. Wynn, Anatoly A. Zhigljavsky: "Dynamical Search – Applications of Dynamical Systems in Search and Optimization: Interdisciplinary Statistics", CRC Press, 1999.

Many algorithms that aim to determine the location of a target in $\mathbb{R}^d$ (typically, the minimizer of a given function) construct a sequence of regions, of decreasing sizes, that converge towards the (fixed) target. By renormalizing the region obtained at each iteration into a fixed base region, we obtain a new representation with a fixed region and a moving target inside. It is the evolution of this moving target over iterations that defines the dynamical system, whose behavior informs us about the performance of the algorithm. All the machinery of dynamical systems can be used, including ergodic theory, and Lyapunov exponents and entropies generated (Shannon and Rényi) can be associated with measures of performance. It may happen that worst-cases have zero ergodic measure, which opens the way to an acceleration of algorithms considered as worst-case optimal. The talk is based on the book [Pronzato, L., Wynn, H., Zhigljavsky, A., 2000. Dynamical Search. Chapman & Hall/CRC, Boca Raton] and mainly focuses on line-search algorithms (like the Golden-Section method); the more difficult cases cases of the ellipsoid and steepest-descent algorithms are briefly considered.

3.20 Open problems relating to Noisy OneMax

Jonathan E. Rowe (University of Birmingham, GB)

Noisy OneMax is an interesting test problem. We know some results on the runtime for various algorithms. There are still lots of interesting open problems, and the most efficient algorithm is still unknown.

3.21 Work at the Alan Turing Institute on "The Data Science Revolution in Scientific Research"

Jonathan E. Rowe (University of Birmingham, GB)

The use of big data methods in science has curious roots, from bioinformatics and paranormal psychology, to particle physics and social economics. These methods took a strange detour via advertising, social media and playing Go, but are now finding applications in research across the breadth of science and the humanities. We will look at a range of projects where AI is transforming research practice, and the role the Alan Turing Institute is playing in this revolution. We will then consider a number of challenges this approach presents, in which some traditional philosophical questions gain unexpected practical applications.

3.22 Runtime in Integer Space Under Multiple Objectives

Günter Rudolph (TU Dortmund, DE)

The talk described a research idea.

3.23 Analysis of Evolution Strategies Applied to a More General Conically Constrained Problem

Patrick Spettel (Fachhochschule Vorarlberg – Dornbirn, AT)

Joint work of Patrick Spettel, Hans-Georg Beyer

Theoretical predictions for the behavior of evolution strategies applied to a linear objective function with a specific conical constraint have recently been derived (work presented at GECCO 2019 and FOGA 2019, among others). The specialty of that problem is that the objective function gradient's direction coincides with the cone axis. Ongoing work tries to predict the behavior of a more general conically constrained problem, in which the objective function gradient does not coincide with the cone axis. The talk presents first steps and initial results.

3.24 Runtime Analysis of Probabilistic Crowding – Results beyond OneMax

Dirk Sudholt (University of Sheffield, GB)

Joint work of Edgar Covantes Osuna, Dirk Sudholt
Main reference Edgar C. Osuna, Dirk Sudholt: "Runtime Analysis of Crowding Mechanisms for Multimodal Optimisation", IEEE Transactions on Evolutionary Computation, pp. 1–1, 2019.
URL https://doi.org/10.1109/TEVC.2019.2914606

Premature convergence is a major challenge in evolutionary computation and many diversity-preserving mechanisms have been proposed to address this. In Probabilistic Crowding, an offspring competes against its parent in a fitness-proportional selection. I showed that a (μ+1) EA with probabilistic crowding does not perform much better than random search on OneMax. We then extended our results to a much more general problem class by introducing a notation of (α, β)-bounded gradients: the gradient towards the optimum is bounded by α in a Hamming ball of radius β around global optima. The improved results show that the algorithm is unable to evolve solutions close to global optima for all functions with bounded gradients and up to exponentially many optima.

3.25 On the Linkage Equilibria of Weakly-Selective Steady-State GAs

Andrew M. Sutton (University of Minnesota – Duluth, US)

I will present some recent work (with Carsten Witt) on the time it takes a steady-state genetic algorithm using uniform crossover and weak probabilistic tournament selection to approach linkage equilibrium, i.e., a state in which sampling a string from the population is very close to the factor distribution over allele frequencies. In this state, sampling from the population is similar to drawing a sample from an EDA, which is attractive from an analysis point of view. This comes at a cost, however, as the selection is likely unreasonably weak from an optimization perspective.

3.26 Single-Iteration Evolutionary Computation (aka Fully Parallel Derivative-Free Optimization)

Olivier Teytaud (Facebook – Paris, FR)

We have a few preliminary results, and we failed to derive a good parametrization for the best performing method.

3.27 Optimal Mixing Evolutionary Algorithms

Dirk Thierens (Utrecht University, NL)

This talk discussed the GOMEA algorithm, specifically the linkage tree model.

3.28 Analysis of Artificial Genetic Representations with Neutrality

Vida Vukašinović (Jozef Stefan Institut – Ljubljana, SI)

Joint work of Vida Vukašinović, Carlos M. Fonseca, Nino Bašić

Kimura's theory of evolution suggests the possibility of occurrence of so called neutral networks. The potential of neutral networks to establish alternative paths for the evolution of a population, and to lead to improved search quality, is the main motivation for the use of redundant representations in evolutionary computation, although not all redundant representations exhibit neutrality. We prepared a solid mathematical formalization of the binary representations developed by Fonseca and Correia (2005) and their equivalence classes. Those representations can exhibit various degrees of neutrality, connectivity, locality, and synonymity, all of which are properties known (or believed) to influence the performance of evolutionary algorithms. Based on this, we developed an efficient algorithm allowing for the exhaustive enumeration of a family of 15-bit representations involving 4 bits of redundancy. The questions of how to identify good, or at least promising, representations in such a large database, and how to automate their (theoretical and practical) evaluation, remain open.

3.29 What's Hot In EA Theory I

Carsten Witt (Technical University of Denmark – Lyngby, DK)

The purpose of this introductory talk was to give an overview of recent trends, techniques and challenges in the theoretical runtime analysis of evolutionary algorithms (EAs). We covered the exact analysis of EAs via drift theory, estimation-of-distribution algorithms and self-adjusting EAs. Topics for future research included monotone functions, multivariate estimation-of-distribution algorithms and a more comprehensive theory of self-adjusting EAs.

3.30 Stochastic global optimization (SGO)

Anatoly Zhigljavsky (Cardiff University, GB)

Main reference Anatoly A. Zhigljavsky: "Theory of Global Random Search". Kluwer Academic Press, Dordrecht, 1991, pp xviii+342
URL https://doi.org/10.1007/978-94-011-3436-1
Main reference Anatoly A. Zhigljavsky, Antanas Zilinskas: "Stochastic Global Optimization", Springer, 2008.

The talk was devoted to some open issues in the theory of global random search, in particular, to the rate of convergence of global random search algorithms in large dimensions and to the theory of evolutionary global random search algorithms. In particular, it was shown that for a large class of stationary evolutionary algorithms the asymptotic distribution of points approaches a stationary limiting distribution, which generalizes the celebrated Gibbs distribution.

4 Working groups

4.1 Breakout Session: Benchmarking – Best Practices and Open Issues

Thomas Bartz-Beielstein (TH Köln, DE)

The goal of this breakout session was to coordinate the working group for writing a commented survey article on "Benchmarking – Best Practices and Open Issues". The aim is to have a survey that is broadly accepted in the community. The outcome of the breakout session was a plan for collecting information and writing the article with the aim of publishing it in 2020.

4.2 Breakout Session: Multiobjective Optimization

Dimo Brockhoff (INRIA Saclay – Palaiseau, FR), Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR), Carola Doerr (Sorbonne University – Paris, FR), Manuel López-Ibánez (University of Manchester, GB), Rolf H. Möhring (TU Berlin, DE), Günter Rudolph (TU Dortmund, DE), Dirk Thierens (Utrecht University, NL), Markus Wagner (University of Adelaide, AU), and Elizabeth Wanner (Aston University – Birmingham, GB)

Time and date: 22.10.2019, 14:30 – 15:30

Participants: Dimo Brockhoff, Benjamin Doerr, Carola Doerr, Manuel López-Ibánez, Rolf H. Möhring, Günter Rudolph, Dirk Thierens, Markus Wagner, Elizabeth Wanner

We started with a short round, in which every participant briefly stated their involvement with multiobjective optimization and theoretical analyses in this context in particular. It turned out that the nine participants have quite heterogeneous backgrounds: about 2/3 claimed a theoretical background (about 1/3 did not do any theory before) and about 1/3 claimed experiences with non-CI methods.

The discussed topics in this working group can be categorized into previously researched (theoretical multiobjective optimization) topics and potential topics for future research. We will detail them in the following subsections but can conclude already here that, before starting to analyze any algorithm (runtime), we have to understand the underlying fundamental problems first.

4.2.1 Previous Research Topics

Compared to single-objective optimization, the theory of (population-based) multiobjective optimization is still in its infancy. Within the short time of the breakout session, we identified only the following, non-exhaustive list of topics that have been touched by previous research:

- fundamental aspects (not related to an algorithm)
  - approximation guarantees
  - optimal p-distributions
  - subset selection
  - properties of quality indicators
  - In discrete problems, the approximation of the Pareto front is polynomially equivalent to the approximation of the ideal point. This has been shown in [1] and means that – at least in theory – algorithms might just concentrate on computing or approximating the ideal point and then use the polynomial transformations of this paper to obtain an approximation to the whole Pareto front.
- computational geometry related problems such as hypervolume computations, see https://simco.gforge.inria.fr/doku.php?id=openproblems for a detailed list
- first runtime analyses

4.2.2 Topics for Future Research

In the remaining time of the breakout session, we collected potential topics for future research such as

- some topics where we don't know the complexity
  - computation of the hypervolume indicator in high dimension
  - algorithms for (bounded) archiving
- similarities between single- and multiobjective optimization, for example, what can be learned/transferred between the two scenarios? What are the differences?
- tradeoffs between different algorithm types (when is which better, e.g. correlation between the objectives); example: (Pareto) local search vs. scalarization, which one is better (and when)?
- Pareto-compliant indicators: how much can they disagree?
- already understanding properties of objective functions is hard in the multiobjective case:
  - The multiobjective quadratic assignment problem has very different instances but the instances of the (random) multiobjective knapsack problem are much less different. The question is why?
  - How do landscapes look like for certain quality indicators and/or different operators? In this context, Manuel brought up the study on NK landscapes in which it was shown empirically that more local optima exist with respect to dominance (between sets) compared to the number of local optima if the hypervolume indicator is the set quality criterion. Also the epsilon indicator shows more local optima compared to the hypervolume indicator. See [2] for details.

References

1 C. Büsing, K.-S. Goetzmann, J. Matuschke, and S. Stiller. Reference points and approximation algorithms in multicriteria discrete optimization. *European Journal of Operational Research*, 260(3):829–840, 2017.

2 A. Liefooghe, M. López-Ibáñez, L. Paquete, and S. Verel. Dominance, epsilon, and hypervolume local optimal sets in multi-objective optimization, and how to tell the difference. In H. E. Aguirre and K. Takadama, editors, *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2018*, pages 324–331. ACM Press, New York, NY, 2018. 10.1145/3205455.3205572.

4.3 Breakout Session: Mixed-Integer-Nominal Optimization

Thomas Bäck (Leiden University, NL)

This breakout session discussed and collected ideas for *mixed-integer-nominal optimization*. Different aspects have been discussed. As a starting point, it is interesting to look into the state of the art of mixed-integer nonlinear programming (MINLP). An idea for handling such problems in evolution strategies is to have a joint covariance matrix between integer and continuous variables. A further interesting question is what the typical problems in this area are. Different such problems were collected: There is work concerning landscape features [1]. Further problems were mentioned including optical multilayer systems (thickness and materials), car body safety optimization (thickness and materials), test case generation in software engineering (arguments for functions, Evosuite). For the theory community, a "standard" problem in this domain could be interesting.

References

1 Carola Doerr, Johann Dréo, and Pascal Kerschke. Making a case for (hyper-)parameter tuning as benchmark problems. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO'19)*, pages 1755–1764. ACM, 2019. 10.1145/3319619.3326857. URL https://doi.org/10.1145/3319619.3326857.

4.4 Breakout Session: Open Problems

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR) and Frank Neumann (University of Adelaide, AU)

The open problem session attracted a good 20 participants and nine open problems from all subdisciplines of the theory of randomized search heuristics. Each problem was presented in at most five minutes and then discussed for as long as the participants had to say something. A number of contradicting conjectures were made, which promises that we will soon see some interesting progress in one direction or the other. The problems can be found on the seminar page.

4.5 Breakout Session: Analysis of Artificial Genetic Representations with Neutrality

Carlos M. Fonseca (University of Coimbra, PT) and Vida Vukašinović (Jozef Stefan Institut – Ljubljana, SI)

The potential of neutral networks to establish alternative paths for the evolution of a population, and to lead to improved search quality, is the main motivation for the use of redundant representations in evolutionary computation, although not all redundant representations exhibit neutrality. We prepared a solid mathematical formalization of the binary representations developed by Fonseca and Correia (2005) and their equivalence classes. Those representations can exhibit various degrees of neutrality, connectivity, locality, and synonymity, all of which are properties known (or believed) to influence the performance of evolutionary algorithms. Based on this, we developed an efficient algorithm allowing for the exhaustive enumeration of a family of 15-bit representations involving 4 bits of redundancy. The representation database obtained in this manner contains over 4.58×10^{10} canonical representatives, each of which representing up to 20160 different representations. The questions of how to identify good, or at least promising, representations in such a large database, and how to automate their (theoretical and practical) evaluation, remain open. The aim of the proposed breakout session was to:

- Discuss current hypotheses about the role of neutrality in evolutionary search and how they may be investigated using this data
- Explore collaborations related to the runtime analysis of evolutionary algorithms based on such representations on simple problems
- Discuss other research opportunities offered by the availability of a database of this kind.

In the beginning, we explained the proposed representations into adequate details. Carlos showed the database as well he presented current feasibilities and obstacles in the database manipulation. During discussion on runtime analysis a justification of runtime analysis of evolutionary algorithms on simple problems based on the proposed representations was put under the question. Main arguments were that by using such representations for simple problems we could not expect better algorithm performance and proving that (1+1)-EA needs $\Omega(n \log n)$ function evaluations is not an impressive result. Nevertheless, we agreed that for deeper understanding what is the influence of proposed representations on the algorithm performance such study is essential. We were able to identify some concrete representations for which runtime analysis of (1+1)-EA on OneMax seems doable. By this an opportunity for further collaboration related to the runtime analysis with the researchers present at the breakout session is opened.

4.6 Breakout Session: Connection Between ES / EDA Analysis and Stochastic Approximation / ODE Theory

Tobias Glasmachers (Ruhr-Universität Bochum, DE)

Stochastic approximation and ODE methods are powerful tools to analyze stochastic algorithms that are formalized as a stochastic approximation of the solution of an underlying Ordinary Differential Equation. In this session we want to discuss how algorithms like Evolution Strategies (ES) (and at least some EDAs) can be casted in the framework of stochastic approximation methods and whether standard ODE methods apply or what is missing in current ODE method to be able to apply it to analyze ES and EDAs.

4.7 Breakout Session: Measuring Optimization Progress in an Invariant Way for Comparison-Based Algorithms

Tobias Glasmachers (Ruhr-Universität Bochum, DE)

Comparison-based or value-free algorithms ignore actual objective function values and instead only use pairwise "better or worse" comparisons. This property renders them invariant to strictly monotonically increasing transformations of objective values. Therefore, measuring quality and optimization progress in terms of function values is inappropriate since it does not exhibit the same invariance properties – possibly, unless there is a clear meaning attributed to these values in an application. An alternative approach is to consider the distance to the optimum instead. That choice is equally problematic unless the (then trivial) objective function is itself a function of that distance. The currently most promising way out of this dilemma is to consider the size (continuous case: Lebesgue measure) of sub-level sets.

4.8 Breakout Session: Invariance

Nikolaus Hansen (INRIA Saclay – Palaiseau, FR) and Anne Auger (INRIA Saclay – Palaiseau, FR)

The invariance breakout gathered participants from the different domains of research present at the seminar. The role of invariance and its importance was acknowledged for algorithm design. In the discrete search domain relatively mild invariance assumptions have lead to a proof of lower runtime bounds. The question was raised whether more results of that type should be expected or attempted. Invariance has also been instrumental for convergence proofs on continuous search spaces via Markov chains. The specific questions on how to model randomness was raised, where methods from stochastic approximation may prove to be useful in convergence proofs for evolutionary algorithms. We also scrutinized specific formulations of invariance, whether to consider the algorithm state in the formalization, and whether the (weak) notion of asymptotic invariance could be helpful.

4.9 Breakout Session: Drift Theory

Martin S. Krejca (Hasso Plattner Institute – Potsdam, DE)

We started with brainstorming some settings in which new drift theorems would be helpful but do not exist yet. The first proposal was a fixed-budget scenario, where one is interested in bounding the expected progress a process has made after a certain (known) time. The discussion suggested that tail bounds on the probability of the process not having reached a target state yet play an important role.

The next setting was drift in multiple dimensions, which is interesting when considering, for example, multiple (independent) processes that should all hit a target value. It was mentioned that the idea of super- and submartingales also generalizes to vectors of random variables. This might yield a useful approach to tackle this problem.

The discussion then moved to a scenario that frequently occurs in multiplicative drift: while the process is far away from the target, its drift is the dominating factor for the expected run time. However, when getting close to the target, the variance of the process is more crucial. It was discussed how the analyses of these two regimes could be combined. The conclusion was to consider the expected return time of the process in the regime of dominating variance, which basically amounts to a restart argument.

Afterward, a drift-like setting for stochastic domination was discussed. The idea was to consider a process where not the expected difference within a single time step is bounded (like in classical drift) but instead a stochastic domination is observed. It was decided that the problem needs to be formalized more rigorously in order to make precise statements.

In the end, the initial idea of drift in the fixed-budget setting was considered in a bit more detail. However, no final conclusion was reached until the session ended.

4.10 Breakout Session: Passing It On

Timo Kötzing (Hasso Plattner Institute – Potsdam, DE)

URL https://github.com/TeachingMetaheuristics

From time to time we all get PhD students or Master Students interested in joining our kind of research. Sometimes we want to teach our subject as a lecture. And also, sometimes researchers from other areas would like to understand better what we do. There are several books available to help such projects along, as well as plenty of other material: talks, scripts, collections of exercises and so on.

In this breakout we discussed in what fashion we could make all these resources widely available for anyone to use. We decided on the following:

- Open a dedicated GitHub repository.
- Let anyone submit more material to this repository.
- The material should be tagged and/or uploaded with a certain structure for easy browsing.
- Anybody who wants can write a "guide" which leads through a subset of the material, aiming at providing a certain expertise. For example
  - "Guide to doing run time analysis of search heuristics" could point to material covering arithmetic and stochastic inequalities, followed by pointing to drift theorems and some easy sample applications.

- Naturally, different material will have a different angle.

What we imagine this resource could be:

- An introduction to the field if read start to finish.
- Reading assignments for classes on the topic.
- A collection of useful assignments.
- A resource for looking up central results.
- Highly modular.
- Multi-authored, yet curated.

We agreed on the following.

- Timo and Thomas W. set up a GitHub.
- Timo sends an Email to all Dagstuhl seminar participants, inviting them to add their materials.
- Once some material is there, Timo sends an email inviting guides.

The github repository can be found at https://github.com/TeachingMetaheuristics.

4.10.1 Addendum by T. Weise

Maybe interesting in this context might be an automated tool chain for writing electronic books. If you host the book's sources in GitHub, the book can get automatically compiled and published to pdf, html, azw3, and epub and uploaded upon each commit: slides, early stage/incomplete example book project.

4.10.2 Addendum by M. Buzdalov

I have been teaching evolutionary computation for maybe three years already, and my other courses are about algorithms and data structures. This inevitably produced a crossover between them, in particular, in a form of *automatically tested programming assignments* for various aspects of evolutionary computation. These should be designed in such a way that solving them using non-evolutionary methods shall be inefficient or impossible. Surprisingly, *there are some*, especially with gray-box techniques. I think that:

- having more such problems is beneficial both for teaching the subject and for ourselves being confident that what we do is interesting for the general public;
- this would be a good practical driver to develop better algorithms and teach the subject properly.

4.11 Breakout Session: The Purpose of Theory Research

Timo Kötzing (Hasso Plattner Institute – Potsdam, DE)

We all do our research, and we all have a good gut feel for what constitutes good research. In this breakout, we brought this gut feel a bit more into the conscious realm and discuss what makes a result "good" and what would be considered less interesting. For example, knowing the lead constant of the expected optimization time of the 1+1 EA on OneMax is (to me?!) not so much of interest in itself; rather, (i) we gain understanding of the inner working principles of the 1+1 EA which (ii) allows us to get a *feel* for many other problems as well, the result generalizes, (iii) let's us dig deeper in related areas after having understood this part. It also (iv) lead us to develop tools (such as drift theory) which are applicable in other contexts as well.

In this breakout we determined the following ingredients for a paper to be worthwhile research.

- **Proper Execution**: The paper is well-written, experiments are clear, ideas are given, proofs are rigorous.
- **Connection to Scientific Community**: Works on topics also others are interested in, discusses own research in context of others', stimulates further research.
- **Academic Honesty**: Explains limits of applicability, does not oversell, does not cheat.
- **Novelty**: The work contains a new idea or a new angle, possibly disproofs commonly held beliefs.
- **Motivation**: Two core sources of a good motivation can be **applicability** (either internally for further research or externally providing valuable input to other scientific communities) and **explanation** of phenomena (also and importantly from other sciences); further worthy kinds of motivation include giving the broader picture, unifying results, settling important questions.

4.12 Breakout Session: Competitive Co-evolution

Per Kristian Lehre (University of Birmingham, GB)

Since long, the field of evolutionary computation has demonstrated empirically that competitive co-evolutionary algorithms can provide state-of-the art solution to certain types of optimisation problems, such as design of sorting networks. However, co-evolutionary algorithms often show pathological behaviour, such as disengagement, loss of gradient, cycling, and overspecialisation. There is currently no theory able to predict and explain this behaviour.

This breakout session explored the potential for runtime analysis of competitive co-evolutionary algorithms. To clarify what runtime means in this context, it is necessary to specify a solution concept, a class of co-evolutionary algorithms, and a class of games.

One well-known solution concept is **Nash equilibrium**. We also discussed **maximin-optimisation**, where we consider interactions between a set of candidate solutions $\mathcal{X} = \{0,1\}^n$ and a set of tests $\mathcal{Y} = \{0,1\}^n$, defined by an interaction function

$$g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}.$$

Here, $g(x, y)$ gives the *performance* of solution $x \in \mathcal{X}$ on test $y \in \mathcal{Y}$. Our goal is to find a solution $x \in \mathcal{X}$ which maximises the function

$$h(x) := \min_{y \in \mathcal{Y}} g(x, y), \tag{1}$$

i.e., to find the solution x which performs best when evaluated with respect to its worst-case test y.

Jon Rowe suggested a co-evolutionary setting with an underlying pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$ and two competing (1+1) EAs. Algorithm A attempts to optimise the function f by finding a good search point x, while Algorithm B attempts to fool Algorithm A by choosing a "wildcard" y. The algorithms interact via a function g defined as below, which Algorithm A attempts to maximise, and Algorithm B attempts to minimise:

$$g(x, y) = f\left(\max(x_1, y_1), \ldots, \max(x_n, y_n)\right). \quad (2)$$

Algorithm 1 Co-evolutionary (1+1) EA

Require: Fitness function $f : \{0,1\}^n \to \mathbb{R}$
Require: Interaction function $g : \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$ as in Eq. (2)

Sample initial search points $x, y \sim Unif(\{0,1\}^n)$
while stopping condition not met **do**
  Obtain x' by flipping each bit in x with prob $1/n$.
  Obtain y' by flipping each bit in y with prob $1/n$.
  if $g(x', y') \geq g(x, y)$ **then**
    $x \leftarrow x'$
  end if
  if $g(x', y') \leq g(x, y)$ **then**
    $y \leftarrow y'$
  end if
end while

It is an open problem to determine for what functions f Algorithm 1 finds an optimal solution x in expected polynomial time. It might be necessary to choose strict inequalities when updating the current search points x and y.

Per Kristian Lehre suggested a framework for population-based co-evolutionary algorithms (Algorithm 2), where a specific algorithm is obtained by choosing a specific operator $\mathcal{D}$. Implicitly, the dynamics is governed by a two-player normal form game with payoff matrices G and H respectively. We make the following assumptions:

1. The strategy spaces $\mathcal{X}$ and $\mathcal{Y}$ are finite and exponentially large in some parameter n, e.g., $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$. Thus, G and H are non-differentiable.
2. The functions G and H can be non-convex.
3. The algorithm is limited to "black box access" to G and H.
4. We have a "solution concept" given by a subset $\mathcal{S} \subset \mathcal{X} \times \mathcal{Y}$
5. $\mathcal{D}$ is non-deterministic, i.e., we need to take into account stochastic effects.

Algorithm 2 Co-evolutionary algorithm

Require: Payoff matrices $G, H : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$
Require: Population size $\lambda \in \mathbb{N}$

for i in $1, \ldots, \lambda$ **do**
  $P_0(i) := (x, y, G(x, y), H(x, y))$ where $(x, y) \sim Unif(\mathcal{X} \times \mathcal{Y})$.
end for
for t in $0, \ldots, \lambda$ **do**
  for i in $1, \ldots, \lambda$ **do**
    $P_{t+1}(i) := (x, y, G(x, y), H(x, y))$ where $(x, y) \sim \mathcal{D}(P_t)$
  end for
end for

First, in steps 1–3, the algorithm samples λ initial pairs of learners and teachers (x, y) uniformly at random, and evaluates the payoff $G(x, y)$ of the learner x, and the payoff of the teacher $H(x, y)$. In each generation t, in steps 5–7, the algorithm samples and evaluates λ new pairs of learners and teachers from a probability distribution $\mathcal{D}(P_t)$ which depends on the current interaction outcomes $P_t \in (\mathcal{X} \times \mathcal{Y} \times \mathbb{R} \times \mathbb{R})^\lambda$.

A (pure) solution concept corresponds to a subset $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Y}$ of the strategy space. The objective of the algorithm is to discover a pair of individuals (x, y) in this set.

▶ **Definition 1** (Runtime).

$$T_{A,\mathcal{S}} := \min\{t \in \mathbb{N} \mid \exists j \in [\lambda] \text{ such that } P_t(j) \in \mathcal{S}\}.$$

Per Kristian also suggested the following maximin-optimisation benchmark problem. The utility function for the prey is $u_1(x, y) = d(x, y)$, while the utility function for the predator is $u_2(x, y) = -d(x, y)$, where for any parameter $\varepsilon \geq 0$,

$$d(x, y) := (|y|_1 - \varepsilon|x|_1)^2 ,$$

and $|z|_1 := \sum_{i=1}^n z_i$ for all bitstrings $z \in \{0, 1\}^n$.

There is a large literature on related concepts in biology, economics, and theoretical computer science. Vivek S. Borkar discussed results from evolutionary game theory.

4.13 Breakout Session: Dynamic Linear Functions

Johannes Lengler (ETH Zürich, CH)

Dynamic Linear Functions are a benchmark where we use a linear function with positive weights, but every few rounds (or every round), the weights are redrawn and everything in the population is re-evaluated. This sounds rather trivial, but it is not. I believe that this is an extremely rich and rewarding topic, and in the talk preceding the breakout session, I will explain where this belief comes from. In a nutshell, they are the easier (less technical) siblings of the monotone functions that have surprised us so many times in the least years.

Goal There are three goals that I hope to achieve:
- Discuss which settings of these functions are most interesting. (E.g., how often should the weights change.)
- Discuss which research questions would be most interesting.
- Motivate other people to work on the topic, either in a collaboration with me, or independently.

Length 30-60 minutes.

Method Group discussion.

Outcome We discussed first how we could categorize the algorithms that fail on dynamic linear functions, or monotone functions. We also discussed the black-box complexity of dynamic linear functions, which is in $\Omega(n/\log n) \cap O(n)$.
We then moved on to discuss algorithms that would be interesting to study on dynamic linear functions. A recurring theme were EDAs like the compact Genetic Algorithm cGA. There were split opinions on how efficient the cGA would be on these benchmarks. Of

particular interest might be whether dynamic linear functions could actually be easier than the static instance of BinVal, i.e., whether noise might actually make optimization easier in this case.
John Rowe pointed out that dynamic linear function have a very peculiar aspect in that the noise level increases as the algorithm moves closer to the optimum, since one-bits contribute noise, whereas zero-bits don't. Finally, we discussed possible extensions, in particular a model in which not all weights are redrawn every round, but rather only a subset of weights is redrawn.

4.14 Breakout Session: Algorithm Configuration and Selection

Pietro S. Oliveto (University of Sheffield, GB)

Main reference George T. Hall, Pietro Simone Oliveto, Dirk Sudholt: "On the impact of the cutoff time on the performance of algorithm configurators", in Proc. of the Genetic and Evolutionary Computation Conference, GECCO 2019, Prague, Czech Republic, July 13-17, 2019, pp. 907–915, 2019.
URL http://dx.doi.org/10.1145/3321707.3321879

In this brief breakout session we discussed the state-of-the-art in the time complexity analysis of algorithm configurators. Most of the discussion concerned which performance measures should be preferred to compare the effectiveness of different parameter settings. Time to optimality, Best identified fitness, and time to fixed targets were considered.

4.15 Breakout Session: Competitions and Benchmarking

Olivier Teytaud (Facebook – Paris, FR) and Carola Doerr (Sorbonne University – Paris, FR)

GECCO and other conferences are hosting several workshops on benchmarking evolutionary algorithms. In addition, a number of competitions are proposed. At the moment, there is little coordination between the workshops and competitions, and we have discussed if it makes sense to coordinate efforts and/or to share best practices and pitfalls.

In the first session the focus of the discussion has centered around the question whether competitions are useful for our understanding of algorithmic behavior, or whether they encourage too much overfitting. In the discussion, most/all participants agreed that contributions to the benchmark environment (e.g., suggestion of new benchmark problems, additional features for a software-based analysis, a critical discussion of different statistics, etc.) are at least as important as the development of high-performing algorithms. Participants agree that such ideas should be "rewarded" as well. A comparison has been made to the Pytorch, which is used in Machine Learning, and which benefits from a user-friendly platform. Another examples that has been mentioned in this context is OpenML, which has similar goals than what we consider a widely accepted benchmarking environment.

At the end of the session we have discussed the idea to have an award committee, which is different from and independent of the organizing committee of the competition, and which judges the contributions made to EC-centered benchmarking environments.

4.16 Breakout Session: One-Shot Optimization

Olivier Teytaud (Facebook – Paris, FR) and Carola Doerr (Sorbonne University – Paris, FR)

In one-shot optimization, aka *single-iteration evolution* or *fully parallel optimization*, the user selects a population, evaluates it, and has to base all future decisions only on the quality of these points. In recent work, O. Teytaud and co-authors have analyzed the setting in which an optimal solution is chosen at random from a Gaussian distribution. They could prove that, unlike one might have guessed, it is better to sample only one (namely, the center of the distribution) rather than sampling n times from the same Gaussian distribution [1]. In the breakout session we have proven that sampling the middle point is not optimal. We have started to compute the optimal distribution, but will need to resume this discussion offline.

Participants: Thomas Bartz-Beielstein, Alexandre Chotard, Carola Doerr, and Olivier Teytaud.

References

1 Marie-Liesse Cauwet, Camille Couprie, Julien Dehos, Pauline Luc, Jérémy Rapin, Morgane Rivière, Fabien Teytaud, and Olivier Teytaud. Fully parallel hyperparameter search: Reshaped space-filling. *CoRR*, abs/1910.08406, 2019. URL http://arxiv.org/abs/1910.08406.

4.17 Breakout Session: Permutation-based Problems

Christine Zarges (Aberystwyth University, GB)

The aim of this breakout session was to discuss current work and potential research directions for permutation-based problems.

Starting with a brief recap of the GECCO 2019 workshop on "Evolutionary Computation for Permutation Problems", the discussion first evolved around different types of permutation-based problems (e.g., total ordering, partial ordering, adjacency) and example problems (e.g., travelling salesperson, linear ordering, linear and quadratic assignment, OneMax variants as introduced in the talk by Maxim Buzdalov, Lyndon factorisation).

Afterwards the group focussed its discussion on metrics and permutation spaces based on the following two publications:

- Ekhine Irurozki: Sampling and learning distance-based probability models for permutation spaces. PhD Thesis, University of the Basque Country, 2014.
- Tommaso Schiavinotto and Thomas Stützle: A review of metrics on permutations for search landscape analysis. Computers & OR 34(10): 3143-3153 (2007)

Carlos Fonseca also pointed out that the API developed by working group 4 of COST Action CA15140 contains some common neighbourhood definitions that could be useful for future work.

Participants: Francisco Chicano, Anton V. Eremeev, Carlos Fonseca, Andrei Lissovoi, Dirk Thierens, Christine Zarges

5 Schedule

Monday

– 09:00	Breakfast
09:00 – 09:15	Welcome and seminar opening
09:15 – 10:00	Participant introduction I
10:00 – 10:30	Coffee break
10:30 – 11:00	**Carsten Witt** on *What's Hot in EA Theory I*
11:00 – 11:20	**Benjamin Doerr** on *Genetic Drift in EDAs*
11:20 – 12:00	Participant introduction II
12:15 – 13:30	Lunch
13:30 – 14:30	Time for individual discussions
14:30 – 15:00	**Luc Pronzato** on *Dynamical Search*
15:00 – 15:20	Participant introduction III
15:20 – 15:30	Short announcement concerning the group work
15:30 – 16:00	Cake
16:00 – 16:15	Organization of group work
16:15 – 17:15	**Breakout Sessions:**
	Passing It On (Timo Kötzing)
	Measuring Optimization Progress in an Invariant Way for Comparison-Based Algorithms (Tobias Glasmachers)
17:15 – 17:45	Participant introduction IV
17:45 – 18:00	Debrief from the breakout sessions
18:00 – 19:00	Dinner
19:30 –	Opening of the art exhibit "Lost Places" by the German artist Winfried Groke

Tuesday

– 09:00	Breakfast
09:00 – 09:20	**Johannes Lengler** on *Dynamic linear functions*
09:20 – 09:40	**Vida Vukašinović** on *Analysis of artificial genetic representations with neutrality*
09:40 – 10:00	**Olivier Teytaud** on *Single-iteration evolutionary computation, also known as fully parallel derivative-free optimization*
10:00 – 10:30	Coffee break
10:30 – 11:00	**Pietro Oliveto** on *What's hot in EA theory II*
11:00 – 11:30	**Vivek Borkar** on *Overview of stochastic approximation and related schemes*
11:30 – 12:00	**Youhei Akimoto** on *Expected runtime bounds for* $(1+1)$*-ES*
12:15 – 13:30	Lunch
13:30 – 14:00	Time for individual discussions
14:00 – 14:30	Organization of group work
14:30 – 15:30	**Breakout Sessions:**
	Neutral Representation (Carlos Fonseca, Vida Vukašinović)
	IGO/Stochastic Optimization (Anne Auger, Tobias Glasmachers)
	Multi-Objective Optimization (Dimo Brockhoff)
	Dynamic Linear Functions (Johannes Lengler)
15:30 – 16:00	Cake
16:00 – 16:10	Flash talk: **Günter Rudolph** on *Runtime in integer space under multiple objectives*
16:10 – 17:30	**Breakout Sessions:**
	Drift Analysis (Martin Krejca)
	Benchmarking and Competition (Olivier Teytaud, Carola Doerr)
17:30 – 18:00	Debrief and announcements
18:00 – 19:00	Dinner
19:00 –	Individual discussions

Wednesday

– 09:00	Breakfast
09:00 – 09:20	**Patrick Spettel** on *Analysis of evolution strategies applied to a more general conically constrained problem*
09:20 – 09:40	**Anne Auger** on *A unified invariance formalism for discrete and continuous optimization*
09:40 – 10:00	**Niko Hansen** on *A (general) definition of invariance*
10:00 – 10:30	Coffee break
10:30 – 10:50	**Anatoly Zhigljavsky** on *Stochastic global optimization (SGO)*
10:50 – 11:10	**Martin Krejca** on *The UMDA on LeadingOnes revisited*
11:10 – 11:30	**Dirk Sudholt** on *Runtime analysis of diversity mechanisms – recent results*
11:30 – 12:00	Organization of group work
12:15 – 13:30	Lunch
13:30 – 15:30	Hike
15:30 – 16:00	Cake
16:00 – 18:00	**Breakout Sessions:**
	Algorithm Configuration and Selection (Pietro Oliveto)
18:00 – 19:00	Dinner
19:00 –	Individual discussions

Thursday

– 09:00	Breakfast
09:00 – 09:20	**Francisco Chicano** on *Dynastic Potential Crossover*
09:20 – 09:40	**Denis Antipov** on *Precise Analysis for Plateaus*
09:40 – 10:00	**Hans-Georg Beyer** on *Evolution Strategies are NOT Gradient Followers*
10:00 – 10:10	Group Picture
10:10 – 10:45	Coffee break
10:45 – 11:05	**Maxim Buzdalov** on *Variations on the Theme of the* $(1 + (\lambda, \lambda))$ *GA*
11:05 – 11:25	**Per Kristian Lehre** on *Runtime Analysis of Self-adaptive EAs*
11:25 – 11:45	**Anton Eremeev** on *On potential for transfer of results from theory of evolutionary algorithms to biology*
11:45 – 11:55	**Jon Rowe** on *Open Questions Relating to Noisy OneMax*
12:15 – 13:30	Lunch
13:30 – 14:30	Time for individual discussions
14:30 – 15:30	**Breakout Sessions:**
	Permutation-based problems (Christine Zarges)
	The Purpose of Theory Research (Timo Kötzing)
	Benchmarking Survey (Thomas Bartz-Beielstein)
	Invariance (Niko Hansen, Anne Auger)
15:30 – 16:00	Cake
16:00 – 17:30	**Breakout Sessions:**
	Mixed-Integer-Nominal Optimization (Thomas Bäck)
	Open Problems (Benjamin Doerr and Frank Neumann)
	One-shot Optimization (Olivier Teytaud)
	Competitive Co-evolution (Per Kristian Lehre)
17:30 – 18:00	Debrief from breakout sessions
18:00 – 19:00	Dinner
19:30 – 20:00	**Jon Rowe** on *Work at the Alan Turing Institute on "The Data Science Revolution in Scientific Research*
20:00 – 20:30	Individual discussions
20:30 –	Wine & cheese party

Friday

– 09:00	Breakfast
09:20 – 09:40	**Manuel López-Ibáñez**
	Automated Algorithm Configuration and Selection for Theoreticians
09:40 – 10:00	**Andrew Sutton** on *On the Linkage Equilibria of Weakly-Selective Steady-State GAs*
10:00 – 10:10	**Dirk Thierens** on *Optimal Mixing Evolutionary Algorithms*
10:10 – 10:30	Coffee break
10:30 – 10:40	**Arina Buzdalova** on *Challenges of mutation rate control in* $(1 + \lambda)$ *EA*
10:40 – 11:00	**Niko Hansen** on *Gradient Descent and Evolution Strategies are Almost the Same*
11:00 – 11:20	**Alexandre Chotard** on *Adaptation of a Sampling Distribution for Metropolis-Hastings*
11:20 – 12:00	Closing session, feedback, and goodbye
12:15 – 13:30	Lunch
13:30 –	Individual departures

Participants

- Youhei Akimoto
University of Tsukuba, JP
- Denis Antipov
ITMO University –
St. Petersburg, RU
- Anne Auger
INRIA Saclay – Palaiseau, FR
- Thomas Bäck
Leiden University, NL
- Thomas Bartz-Beielstein
TH Köln, DE
- Hans-Georg Beyer
Fachhochschule Vorarlberg –
Dornbirn, AT
- Vivek Shripad Borkar
Indian Institute of Technology
Bombay, IN
- Dimo Brockhoff
INRIA Saclay – Palaiseau, FR
- Maxim Buzdalov
ITMO University –
St. Petersburg, RU
- Arina Buzdalova
ITMO University –
St. Petersburg, RU
- Francisco Chicano
University of Málaga, ES
- Alexandre Chotard
Calais University, FR
- Benjamin Doerr
Ecole Polytechnique –
Palaiseau, FR
- Carola Doerr
Sorbonne University – Paris, FR
- Anton V. Eremeev
Institute of Scientific Information
for Social Sciences RAS –
Moscow, RU
- Carlos M. Fonseca
University of Coimbra, PT
- Tobias Friedrich
Hasso-Plattner-Institut –
Potsdam, DE
- Tobias Glasmachers
Ruhr-Universität Bochum, DE
- Nikolaus Hansen
INRIA Saclay – Palaiseau, FR
- Thomas Jansen
Aberystwyth University, GB
- Timo Kötzing
Hasso-Plattner-Institut –
Potsdam, DE
- Martin S. Krejca
Hasso-Plattner-Institut –
Potsdam, DE
- Per Kristian Lehre
University of Birmingham, GB
- Johannes Lengler
ETH Zürich, CH
- Andrei Lissovoi
University of Sheffield, GB
- Manuel López-Ibánez
University of Manchester, GB
- Rolf H. Möhring
TU Berlin, DE
- Frank Neumann
University of Adelaide, AU
- Pietro S. Oliveto
University of Sheffield, GB
- Luc Pronzato
Laboratoire I3S –
Sophia Antipolis, FR
- Jonathan E. Rowe
University of Birmingham, GB
- Günter Rudolph
TU Dortmund, DE
- Ofer M. Shir
Tel-Hai College –
Upper Galilee, IL
- Patrick Spettel
Fachhochschule Vorarlberg –
Dornbirn, AT
- Dirk Sudholt
University of Sheffield, GB
- Andrew M. Sutton
University of Minnesota –
Duluth, US
- Olivier Teytaud
Facebook – Paris, FR
- Dirk Thierens
Utrecht University, NL
- Vida Vukašinovic
Jozef Stefan Institut –
Ljubljana, SI
- Markus Wagner
University of Adelaide, AU
- Elizabeth Wanner
Aston University –
Birmingham, GB
- Thomas Weise
Hefei University, CN
- Carsten Witt
Technical University of Denmark
– Lyngby, DK
- Xin Yao
Southern Univ. of Science and
Technology – Shenzen, CN
- Christine Zarges
Aberystwyth University, GB
- Anatoly Zhigljavsky
Cardiff University, GB

Report from Dagstuhl Seminar 19432

Analysis of Autonomous Mobile Collectives in Complex Physical Environments

Edited by
Mario Gleirscher[1], Anne E. Haxthausen[2], Martin Leucker[3], and Sven Linker[4]

1 **University of York, GB,** mario.gleirscher@york.ac.uk
2 **Technical University of Denmark – Lyngby, DK,** aeha@dtu.dk
3 **Universität Lübeck, DE,** leucker@isp.uni-luebeck.de
4 **University of Liverpool, GB,** s.linker@liverpool.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19432 "Analysis of Autonomous Mobile Collectives in Complex Physical Environments". Our working hypothesis for this seminar was that for systems of such complexity and criticality, the trustworthy certification and the successful operation in society will strongly benefit from the coordinated application of several rigorous engineering methods and formal analysis techniques. In this context, we discussed the state-of-the-art based on the working example of a Smart Farm. Our aim was to understand the practical challenges and the capabilities and limitations of recent formal modelling and analysis techniques when tackling these challenges, and to initiate a special research community on the verification of autonomous collectives.

Seminar October 20–23, 2019 – http://www.dagstuhl.de/19432
2012 ACM Subject Classification Computer systems organization → Robotic autonomy, Computing methodologies → Model development and analysis, Theory of computation → Logic and verification, Theory of computation → Timed and hybrid models
Keywords and phrases autonomous collectives, control engineering, formal verification, hybrid systems, uncertainty and risk
Digital Object Identifier 10.4230/DagRep.9.10.95
Edited in cooperation with Frederik Forchhammer Foldager

1 Executive Summary

Mario Gleirscher (University of York, GB)
Anne E. Haxthausen (Technical University of Denmark – Lyngby, DK)
Martin Leucker (Universität Lübeck, DE)
Sven Linker (University of Liverpool, GB)

Motivation

Autonomous vehicles (AVs) are facing strong proof obligations. Individual AVs can be part of a *collective* (e.g. a platoon of utility vehicles on a farm field, a truck convoy on a highway, a convoy of passenger vehicles on urban road, an in-door aerial platoon, a railway convoy) and act within a heterogeneous environment of other collectives, for example, pedestrians, bicyclists, and motorcyclists. Multiple AVs might have to correctly and reliably negotiate their order of passing a crossing or reliably and robustly arrange in a certain work layout on agricultural land. Individuals and collectives in such environments, whether controlled in a centralised or distributed way, are subjected to change, uncertainty, and defects. Moreover, *complex environments* typically deny a comprehensive segregation of physical space and, hence, involve interactions with entities out of control (e.g. human-controlled machines, pedestrians, animals) and mostly also out of sight of an individual machine's (short-range) sensors.

Objective

This seminar was centred around an application challenge, the **Smart Farm**. Participants were encouraged to discuss how their research addresses typical **engineering tasks** (ETs; upper layer in Fig. 1) to be accomplished for the given challenge or for similar challenges. These tasks include

1. the identification, modelling, and analysis of operational situations in complex environments
2. real-time coordination, composition, and reconfiguration of machine collectives with a focus on (i) interaction with human-operated systems, humans, animals, infrastructure and (ii) situation-specific centralised or distributed control regimes
3. the determination of strongest safety and performance guarantees with a focus on (i) the estimation of upper resilience bounds of machine collectives and lower reliability bounds of individual machines and (ii) the determination of strongest guarantees under partial state knowledge, with minimal infrastructural support, and under reduced controllability.

In the discussions of how the ETs can be accomplished best, we also aimed at investigating **abstractions of defects and uncertainties**, for example:

- controller, communication, and infrastructure failures (e.g. erroneous vehicle-to-X connection and communication, deficient road infrastructure),
- undesired interference or disturbance of autonomous operation (e.g. malicious and unintended misuse; controller, communication, and infrastructure attacks),
- practical sensor uncertainties, actuator perturbations, and partial state knowledge.

Defects and uncertainties are crucial for constructing *realistic models* of the behavioural spectrum of mobile collectives and yet abstract enough to perform *practical reasoning*. Likewise, such models allow the necessary freedom to express ideal and actual behaviour, independent of whether such behaviour is desirable. This freedom can involve the use of non-deterministic models. In any case, a *(property) specification* would label some of the observable behaviours as *desirable*, some as *undesirable*, others somewhere in between (cf. quantitative verification). The more complete and precise such a specification, the better the distinction between correct, undesirable, and other classes of behaviours of a collective.

Our **overall objective** with this seminar was to gain a common understanding of acceptable safety and performance of autonomous mobile collectives in presence of defects and other uncertainties typically occurring in complex open environments. The **overarching approach** of all seminar contributions was the **formal analysis and verification of behavioural correctness** under these assumptions (lower layer in Fig. 1) by using techniques such as, e.g. theorem proving, model checking, run-time verification, and model-based testing.

Our **central assumption for this seminar** was that the given application challenge or any similar challenges render individual methods for the analysis and verification of such systems insufficient. For example, in control-theoretic models such collectives are modelled by differential equations. Interaction within and among collectives and with their

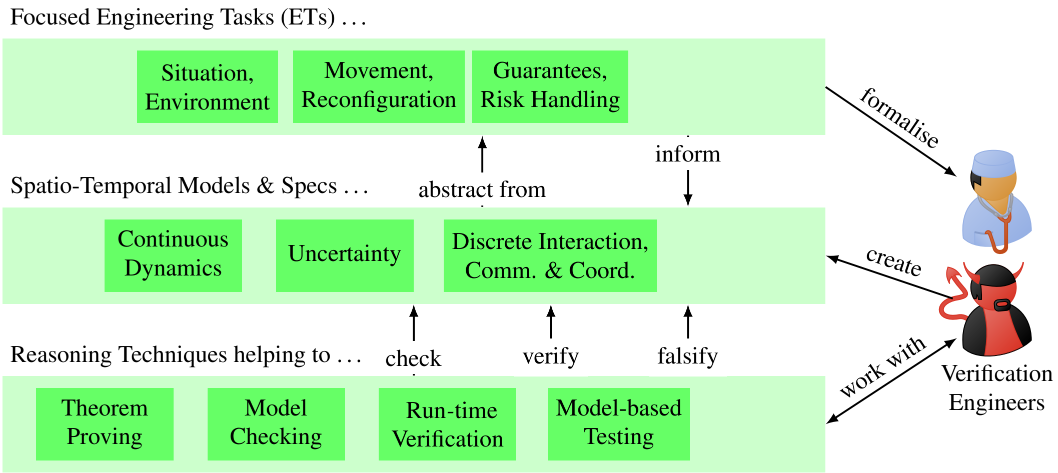

Figure 1 Topic structure of the seminar.

environment, governing these equations, cannot be easily encoded. Approaches that express such interactions well, however, typically struggle with the detailed description of the physical laws the AVs need to adhere to. Hence, for ensuring correct behaviour in such a setting, layered abstractions, corresponding models, and **specialised reasoning techniques have to be combined**.

Organisation

Before the seminar, we provided each participant with material about the *application challenge* (see Section 4.1) together with list of *engineering tasks* and *research questions*. We encouraged the participants to apply their approach, if available, to at least one of the ETs of the application challenge and to answer at least one of the research questions. Alternatively, participants were invited to present any research and practical experiences related to the seminar topic and the challenge. Everyone was given the opportunity to give a full-length talk. Table 1 shows the seminar structure, the talks, and further sessions. After the welcome session, participants introduced themselves to the group. The rest of the seminar was organised into *talk sessions* and *break-out sessions*.

Talks

In the talk sessions, we investigated several **research questions** from different angles. We had talks about (1) industry challenges, (2) the analysis and verification of properties of individual autonomous vehicles (two sessions), (3) the analysis and verification of properties of autonomous collectives, and (4) the modelling of uncertainty for the (quantitative) property verification of critical autonomous systems. Nine talks dealt with an **introduction of a specific verification approach** suitable for tackling an aspect of the application challenge, including a summary of the state-of-the-art of this approach. Four talks were about **industrial examples** of a nature similar to the Smart Farm, highlighting technical challenges, encountered issues, and perceived practical obstacles. Five talks focused on the **application of a particular approach to a particular aspect of the Smart Farm**, addressing some of the research questions.

In the following, we list the main questions and the participants whose talks highlighted a particular aspect of the corresponding question. For more details, see the list of talk abstracts below.

Table 1 Seminar schedule.

	Monday	Tuesday	Wednesday
9:00 9:30	Introductions	***Industry Challenges*** J. Brauer: *Verification of Autonomous Transport Systems - Some Industrial Prospects* S. Fröschle: *Trustworthy identity and key management for mobile systems in transportation*	*Break-out session*
10:00–10:30	break	break	break
10:30 10:50 11:10 11:30	***Individual Properties*** P.G. Larsen/F. Foldager: *A Journey Towards a Fleet of Autonomous Robots for Agricultural Field Operations* J.B. Jeannin: *Collision avoidance and path replanning of individual farm robots* A. Fantechi: *Safety aspects of autonomous systems* P.C. Ölveczky: *Formal modeling and analysis of real-time systems using Real-Time Maude*	***Uncertainty Modelling*** K.G. Larsen: *Synthesis of Safe, Optimal and Small Strategies for Advanced Driver Assistance using UPPAAL Stratego* D. Parker: *Probabilistic model checking for safety and performance guarantees* R. Calinescu: *Stochastic modelling underpinning the engineering of trustworthy autonomous systems* M. Gleirscher: *Risk Structures*	*Break-out and discussion* *Closing discussion*
12:15–13:30	lunch	lunch	lunch
13:30 13:50 14:10 14:30	***Collective Properties*** M. Waga: *Optimization of the watering schedule by run-time and design-time analysis* É. André: *White-box and black-box quantitative verification of timing properties* P. Ribeiro: *Modelling and Verification using RoboChart* (spare)	***Individual Properties*** C. Heinzemann: *Context Analysis and Requirements Derivation with SCODE* S. Bogomolov: *Trusted Autonomous Systems: Verification Meets Falsification* S. Mitsch: *Modular Verification of Cyber-Physical Systems in KeYmaeraX* (spare)	
15:00–15:30	break	break	
15:30 16:00 16:30 17:00	*Break-out session* *Discussion of results*	*Break-out session* *Discussion of results*	
18:00	dinner	dinner	

1. How can each ET be solved? How can we achieve safety in presence of distribution, mobility, and uncertainty? Which mechanisms fit best to ensure safety in the *application challenge*?
   Frederik Foldager and Peter Gorm Larsen
2. How do we model the systems and verify *safety and progress* properties? Can we always find acceptable Pareto optima over safety and performance, at traffic level, at the level of a collective, and for individual machines?
   Étienne André, Sergiy Bogomolov, Kim Larsen, David Parker
3. How can we *exploit the structure* of practical AVs and collectives to craft specific verification techniques (e.g. prevent state space explosion, identify fundamental theorems)?
   Stefan Mitsch, Pedro Ribeiro
4. Which benefits do we gain from *integrating* design-time verification, model-based testing, and run-time verification?
   Mario Gleirscher, Masaki Waga
5. How can verification techniques be incorporated into the *development process* of AVs?
   Jörg Brauer, Radu Calinescu, Alessandro Fantechi, Peter Csaba Ölvecky

6. Which *complications* arise from the verification of AVs and how can we mitigate the impact of these complications, particularly, during practical verification?
   Sibylle Fröschle, Christian Heinzemann

Break-Out Sessions

To stimulate interaction, we created *break-out groups* on each seminar day and on the following topics: challenges of verifying autonomous collectives, the challenge of uncertainty (using, e.g. quantitative verification, parametric model checking), abstractions of space & uncertainty, the impact of IT security issues on AV safety, and safe platooning. Additionally, several smaller groups (sometimes consisting of only two participants) met to discuss combinations and extensions of the topics they presented in their respective talks.

One break-out group focused on creating a big picture of the **challenges of verifying autonomous mobile collectives** in the Smart Farm. The identified problems include

- estimation of behavioural properties (e.g. exact arrival times of agents, dead-lock freedom of the plan), real-time interleaving of sensing and control, and finding the "sweet spot" between precision and performance when used at run-time,
- model checking at scale, when to use online or offline analysis for verification and synthesis (e.g. synthesis of distributed safety controllers for automatic repair/fallback),
- useful architectural abstractions, compositionality, and refinement (e.g. how to safely partition the tasks of a mission between system components or whole robots?),
- security of communication and robustness of control to communication glitches (e.g. how to integrate a jamming model into overall system verification?),
- languages/models for dealing with system failures (e.g. how to cope with failures of individual autonomous vehicles in the context of a collective?) and component failures (e.g. how to safely integrate machine learning into autonomous systems?), and
- safety in the presence of uncertainty (e.g. how to quantify uncertainty?, how to deal with uncertainty in parameters and in the structure of the system and the environment?).

Another group investigated **the challenge of uncertainty in modelling**, discussing how uncertainty (e.g. due to partial observability) can be dealt with in automated verification and how techniques such as quantitative verification can be used to solve verification problems with uncertainties in the considered parameters. Depending on the Smart Farm aspect to be tackled, state-of-the-art approaches include the use of interval abstractions for parameters, the calculation of confidence intervals for verification results, and the use of counterexample-guided abstraction refinement.

The break-out session on **space and uncertainty** stretched over all three days, and was concerned with the possible ways to specify spatial aspects, as well as how to incorporate uncertainty into such specifications. Our discussion proceeded on different topics. We discussed, which types of sensors allow robotic systems to gain spatial knowledge, and what levels of uncertainty can be expected. Based on this, we examined whether several layers of space are necessary and beneficial to specify both the systems and their desired properties (e.g., a discrete layer for planning high-level actions and a continuous layer, on which more local properties are ensured by controllers, as for example obstacle avoidance). Furthermore, we compared the different types of uncertainty, the level of spatial layers they occur on, and their impact on systems in the Smart Farm. This included a discussion of how much knowledge needs to be globally available, and what can be kept locally at the level of each individual entity. We realised that while the modelling scenario allowed for different levels of space and uncertainty, it was not easy and straightforward to identify necessary and interesting spatial properties to analyse. Hence, we agreed that the case study needs to allow for more degrees of freedom (e.g., different routes to reach physical targets, to permit several alternative plans).

The session about **IT security of farm collectives** focused on the aspect of communication security. First, the group identified the typical communication requirements between the actors of a smart farm such as: between a robot and a supervisory control (perhaps including a drone), between two robots that carry out a task on the same field (e.g. to carry out the task cooperatively or for collision avoidance), between a sensor and a control centre (e.g. for watering). Altogether, it became clear that the operation of a smart farm critically depends on the secure and timely communication between the various actors. It is also clear that in the setting of the smart farm the actors must communicate over wireless channels. Hence, the usual threats against communication over an open medium apply, e.g. message spoofing and manipulation, eavesdropping and jamming. On the one hand, this requires us to employ appropriate security protocols and key management, which can guarantee origin and message authenticity as well as confidentiality. On the other hand, this requires further measures against availability attacks such as jamming. The group focused on the threat of jamming. While jamming cannot be prevented in an open system the general idea was to take a 'detect and mitigate' approach. For example, jamming can be detected by the absence of regular 'heartbeat' signals and by combination with visual channels. Mitigation strategies involve raising an alarm and removing the jamming device in a timely fashion while ensuring the system is not overly susceptible to false positives and denial-of-service attacks. Neither detection nor mitigation seemed trivial when discussed in detail. On the positive side, the verification methods and tools presented at the seminar could be used to evaluate possible strategies, and perhaps, even to synthesise them. Later on the group joined the break-out group on platooning, where communication is particularly critical.

In the break-out session on **safe platooning on the farm**, we discussed

1. the handling of *planned events* being part of the normal operation of a platoon (e.g. several farm vehicles, lorries and harvesters, form a platoon including leader election; a lorry wants to join or leave a harvesting platoon; a platoon with two consecutive lorries needs to be rearranged; a lorry decides to leave the platoon) and
2. the detection of *critical (not necessarily undesired) events* to be dealt with or to recover from during normal operation (e.g. a foreign vehicle, a farmer's car, enters the platoon area; communication error because of a jamming attack or a hardware failure disturbs the platoon controller; the current leader looses trustworthiness, e.g. because of being hacked, by deviating from the common goal of the platoon; farm workers enter the working area of the platoon).

Our discussions lead to a deeper understanding of the intricacies, both from the perspectives of different verification approaches and from the viewpoint of certification obligations. The results of our discussion are suitable for the identification of *formal properties* to be used as proof obligations in certification activities as well as the modelling of so-called protocol automata *describing the inter- and intra-modal behaviour* required to handle some of the mentioned events. Such models can then serve as a basis for hazard and risk assessment activities as well as for safety verification.

Outcomes and Conclusions

Our expectations for this first seminar were modest. We wanted to learn from each others' perspectives, to discuss available approaches, and to identify the hardest and most relevant **open challenges**.

Our discussions opened **paths to an integration and application of the presented theories and models** (middle layer in Fig. 1), particularly, continuous models (e.g. timed and hybrid automata), uncertainty models (e.g. Markov chains, probabilistic automata), communication and coordination models (e.g. timed process algebra). We investigated the use of such models in the context of various reasoning techniques (e.g.theorem proving, model checking, run-time verification, model-based testing). These discussions lay a basis for the *derivation of guidelines* on how the approaches, when applied to systems such as the Smart Farm, can be combined and/or enhanced to tackle the identified problems *in practical contexts subject to certification efforts*.

The attendees were from various fields such as formal verification, testing, certification, mechanical and control engineering, and embedded IT security, working at universities, in industry-oriented research institutes, or directly in industry. In this setting, we were able to **share experiences and insights from various application domains** (e.g. smart farming, smart energy systems, train/railway systems, automotive and transportation), to discuss issues of the Smart Farm scenario, and to examine potential research directions. Particularly, we observe that commonalities among the used approaches give rise to an *integrated and more versatile approach*. Our participants from industry receive the opportunity to convert any of these insights into lasting process improvements in their safety-critical domains. We expect our findings to be *relevant to regulatory authorities* in these domains.

In overall, we believe this seminar was an important step to **foster collaboration** of researchers and practitioners experienced with the different models and reasoning techniques, and to **initiate a research community** focusing on autonomous collectives of similar or even higher complexity than the Smart Farm. To that end, we are planning **further meetings** of the seminar's participants in the near future, to allow for further refinement of the models, and combinations of the methods presented. Additionally, we will further improve and **extend the modelling scenario**, so that a particular combination of specification and verification approaches can be explored in more detail. Eventually, we intend to collect our findings possibly in a special issue of a suitable journal.

Funding and Acknowledgements. Sven Linker was supported by the Engineering and Physical Sciences Research Council programme grant EP/N007565/1 (S4: Science of Sensor Systems Software). Mario Gleirscher was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Grant no. 381212925. We are grateful to Sibylle Fröschle for summarising the results of the IT security session. Further thanks go to Frederik Foldager for collecting and compiling the abstracts. We would like to spend sincere gratitude to all participants for their contributions and for their support and active engagement in making this seminar an insightful experience.

2 Table of Contents

Executive Summary
Mario Gleirscher, Anne E. Haxthausen, Martin Leucker, and Sven Linker 95

Overview of Talks

White-box and black-box quantitative verification of timing properties
Étienne André 103

Trusted Autonomous Systems: Verification Meets Falsification
Sergiy Bogomolov 104

Verification of Autonomous Transport Systems – Some Industrial Prospects
Jörg Brauer 104

Stochastic modelling underpinning the engineering of trustworthy autonomous systems
Radu Calinescu 104

Safety aspects of autonomous systems
Alessandro Fantechi 105

A Journey Towards a Fleet of Autonomous Robots for Agricultural Field Operations
Frederik Foldager and Peter Gorm Larsen 105

Trustworthy Identity and Key Management for Mobile Systems in Transportation
Sibylle Fröschle 106

Risk Structures: Specification Templates for Controller Synthesis
Mario Gleirscher 106

Context Analysis and Requirements Derivation with SCODE
Christian Heinzemann 106

Synthesis of Safe, Optimal and Compact Strategies using UPPAAL Stratego
Kim Guldstrand Larsen 107

Modular Verification of Cyber-Physical Systems in KeYmaera X
Stefan Mitsch 107

Probabilistic Model Checking for Safety and Performance Guarantees
David Parker 108

Modelling and Verification using RoboChart
Pedro Ribeiro 109

Optimization of the watering schedule by run-time and design-time analysis
Masaki Waga 109

Formal modeling and analysis of real-time systems using Real-Time Maude
Peter Csaba Ölveczky 110

Open problems

Specification of the Application Challenge
Mario Gleirscher, Anne E. Haxthausen, Martin Leucker, and Sven Linker 110

Participants 116

3 Overview of Talks

3.1 White-box and black-box quantitative verification of timing properties

Étienne André (University of Paris North, FR)

Joint work of Masaki Waga, Étienne André, Ichiro Hasuo
Main reference Masaki Waga, Étienne André, Ichiro Hasuo: "Symbolic Monitoring Against Specifications Parametric in Time and Data", in Proc. of the Computer Aided Verification – 31st International Conference, CAV 2019, New York City, NY, USA, July 15-18, 2019, Proceedings, Part I, Lecture Notes in Computer Science, Vol. 11561, pp. 520–539, Springer, 2019.
URL https://doi.org/10.1007/978-3-030-25540-4_30

In this talk, I will envision two parts: on a white box model, i.e., on a formal model of (part of) the system, I will propose to use parametric timed model checking techniques to formally evaluate the correctness of (some of) the timing aspects, but also to evaluate their robustness, i.e., the effect of infinitesimal variations on the system correctness. That is, how critical can be some timing parameters, such as *del_t* or *gps_t*, to the system correctness? The formalism used will be parametric timed automata [1].

Then, on a black box model (obtained by either concrete execution or, more likely, on simulation using tools such as Simulink), I will propose efficient run-time verification techniques to *monitor* the system behavior, again taking into consideration the timing aspects and their robustness. On the one hand, on a "shorter-time scale", the absence of collisions, but also the *robust* absence of collisions (i.e., situations of "near collisions") should be monitored. On the other hand, on a "longer-time scale", the absence of rotten ripes, and their robust counterpart ("near-rotten" situations) should be monitored. The ultimate goal is to not only perform a Boolean monitoring, but to detect problematic timeframes, and to provide them with a quantitative measure of the property. This implies to be able to write specifications in some quantitative formalism sufficiently expressive to allow to detect such failure, together with some robustness values. The formalism used could be (variants of) *parametric timed data automata*, a formalism recently proposed with Ichiro Hasuo and Masaki Waga [2].

References

1 Rajeev Alur, Thomas A. Henzinger, and Moshe Y. Vardi. Parametric real-time reasoning. In Rao Kosaraju, David S. Johnson, and Alok Aggarwal (eds.), STOC'93, ACM, pages 592–601, 1993. DOI: 10.1145/167088.167242
2 Masaki Waga, Étienne André and Ichiro Hasuo. Symbolic monitoring against specifications parametric in time and data. In Işil Dillig and Serdar Tasiran (eds.), CAV'19, Springer LNCS 11561, pages 520-539, July 2019. DOI: 10.1007/978-3-030-25540-4_30

3.2 Trusted Autonomous Systems: Verification Meets Falsification

Sergiy Bogomolov (Newcastle University, GB)

Joint work of Sergiy Bogomolov, Goran Frehse, Amit Gurung, Dongxu Li, Georg Martius, Rajarshi Ray
Main reference Sergiy Bogomolov, Goran Frehse, Amit Gurung, Dongxu Li, Georg Martius, Rajarshi Ray: "Falsification of hybrid systems using symbolic reachability and trajectory splicing", in Proc. of the 22nd ACM International Conference on Hybrid Systems: Computation and Control, HSCC 2019, Montreal, QC, Canada, April 16-18, 2019, pp. 1–10, ACM, 2019.
URL https://doi.org/10.1145/3302504.3311813

Falsification algorithms for hybrid systems aim at finding trajectories that violate a given safety property. This is a challenging problem, and the practical applicability of current falsification algorithms still suffers from their high time complexity. In contrast to falsification, verification algorithms aim at providing guarantees that no such trajectories exist. Recent symbolic reachability techniques are capable of efficiently computing linear constraints that enclose all trajectories of the system with reasonable precision. In this talk, we present an approach which leverages the power of symbolic reachability algorithms to improve the scalability of falsification techniques. Recent approaches to falsification reduce the problem to a nonlinear optimization problem. We propose to reduce the search space of the optimization problem by adding linear state constraints computed by a reachability algorithm. We showcase the efficiency of our approach on a number of standard hybrid systems benchmarks demonstrating the performance increase in speed and the number of falsifiable instances.

3.3 Verification of Autonomous Transport Systems – Some Industrial Prospects

Jörg Brauer (Verified Systems International GmbH – Bremen, DE)

Coming from industry, most of our projects are to some extent based on development standards such as the RTCA DO-178 for avionics systems, which have not really been set up with adaptive or autonomous systems in mind. In this talk, we focus on some aspects of how safety certification and autonomy do not really match up, and what we can do about it.

3.4 Stochastic modelling underpinning the engineering of trustworthy autonomous systems

Radu Calinescu (University of York, GB)

Stochastic modelling is a powerful tool for establishing performance, dependability and other key properties of systems and processes during design, verification and at run-time. However, the usefulness of this tool depends on the accuracy of the models being analysed, on the efficiency of the analysis, and on the ability to find models corresponding to effective system and process architectures and configurations. This talk will describe how recent approaches to stochastic model learning, analysis and synthesis address major challenges posed by these prerequisites, extending the applicability of stochastic modelling to autonomous systems.

3.5 Safety aspects of autonomous systems

Alessandro Fantechi (University of Florence, IT)

The talk will review the currently considered/implemented techniques and policies for safety enforcement of autonomous railway vehicles, with the aim to derive a more general conceptual model encompassing the principles upon which safety of autonomous vehicles is assessed

Notions of uncertainty over positioning and speed metering of autonomous vehicles are also inherited from what is currently investigated in the railway domain, and generalised to the three-dimensional case.

The sketched concepts are then instantiated on the provided benchmark, as a contribution to develop an analytic safety assessment process.

3.6 A Journey Towards a Fleet of Autonomous Robots for Agricultural Field Operations

Frederik Foldager (Aarhus University, DK) and Peter Gorm Larsen (Aarhus University, DK)

In this presentation, we provide an overview of the collaboration between a proactive Danish SME called Agrointelli and Aarhus University to make the vision of a fleet of autonomous robots for arable farming a reality. The work surrounds a full-scale robot called Robotti which is now sold commercially. The journey includes both a series of different joint research projects involving many other institutions as well as considerations of commercial and business development. We will give an introduction to how we have modelled the soil-machine interaction using the Discrete Element Method on a component level, as well as explaining the models that have been made both of the dynamics of the robot, its complex physical environment, in particular in relation to different soil-types and the model of the different levels of the discrete event controllers on a systems level. Many of these have been combined using a technology called co-simulation which also includes capabilities for exploring alternative designs in a virtual setting as well as connecting it to 3D visualization engines. Some of these models are naturally commercially sensitive but we are also able to share a purely public version of these multi-models. Our current research involves supporting this with a digital twin capability in a real-time fashion and scaling up to a fleet of robots operating in collaboration with humans. We expect to close the presentation with some research challenges that we currently see as the most prominent ones.

3.7 Trustworthy Identity and Key Management for Mobile Systems in Transportation

Sibylle Fröschle (OFFIS – Oldenburg, DE)

In this talk I will talk about the importance and challenges of trustworthy identity and key management for mobile autonomous systems, and illustrate this by examples from the automotive, aerospace, and maritime domain. I will then present current research on how to answer these challenges, including how to obtain verifiable security and resilience guarantees on the system-of-systems layer. Finally, I will report on practical experiences within the working group "Identity management and security" of the Maritime Connectivity Platform (MCP).

3.8 Risk Structures: Specification Templates for Controller Synthesis

Mario Gleirscher (University of York, GB)

Main reference Mario Gleirscher: "Run-Time Risk Mitigation in Automated Vehicles: A Model for Studying Preparatory Steps", in Proc. of the Proceedings First Workshop on Formal Verification of Autonomous Vehicles, FVAV@iFM 2017, Turin, Italy, 19th September 2017, EPTCS, Vol. 257, pp. 75–90, 2017.
URL http://dx.doi.org/10.4204/EPTCS.257.8

To achieve desirable safety, autonomous systems will have to detect, predict, and reduce risk by incorporating risk models and risk handling mechanisms that enhance their mission controllers. Complex environments and the missing fallback to human operators pose tough challenges to the engineering of risk handlers, particularly, to the hazard analysis and risk modelling leading to such handlers. This talk will discuss research on an algebraic framework for risk modelling and analysis. It will also be highlighted how one can use a specific risk model to derive proof obligations for mission controllers with safety mechanisms.

3.9 Context Analysis and Requirements Derivation with SCODE

Christian Heinzemann (Robert Bosch GmbH – Stuttgart, DE)

Autonomous systems, particular autonomous driving systems, need to cope with complex environments and are subject to a multitude of influences that have an impact on the necessary behavior of a system. To this end, key questions are what constitutes a correct behavior in a given situation and how to derive a complete-as-possible set of requirements for an autonomous system in a given environment (or context)? In my talk, I will outline an approach based on essential analysis (also known as morphological analysis) for capturing influence factors from a system's context and for deriving a set of top-level requirements (or modes of operation) that denote an expected system reaction to a specific combination of external influences. The approach guarantees that the derived top-level requirements (or modes of operation) are consistent and complete with respect to the known and specified influences.

3.10 Synthesis of Safe, Optimal and Compact Strategies using UPPAAL Stratego

Kim Guldstrand Larsen (Aalborg University, DK)

In this talk I gave an overview of the UPPAAL tool suite with outset in the Smart Farming Benchmark of the seminar. The classical version of UPPAAL allows for a Timed Automata model of the timed behaviour of robots capturing their movement on the road as well as entering and leaving collection point and field. In particular, timing properties may be verified here given best and worst case timing information.

A refinement of the timed automata model interpret delays stochastically giving rise to Stochastic Timed Automata. Here expected and probabilistic threshold properties may be settled using the statical model checking engine of UPPAAL SMC.

In the setting of two robots, we model their joint behaviour as a (product) Timed Game. This allows for synthesis of most permissive safety controllers, where crashes between robots is guaranteed to be avoided.

Finally, we add stochastic components for weather prediction and hybrid components in terms of differential equations describing the growth of crops in the field. Given this overall model – a stochastic hybrid game – we use the reinforcement learning method of UPPAAL Stratego to obtain a near optimal sub-strategy of the no-crash safety strategy.

3.11 Modular Verification of Cyber-Physical Systems in KeYmaera X

Stefan Mitsch (Carnegie Mellon University – Pittsburgh, US)

Joint work of Stefan Mitsch, Andre Platzer, Brandon Bohrer, Yong Kiam Tan, Nathan Fulton, Andreas Müller, Wieland Schwinger, Werner Retschitzegger, Jan-David Quesel, Marcus Völp, Magnus O. Myreen

Cyber-physical systems (CPS) combine cyber aspects such as communication and computer control with physical aspects such as motion in space; they have many important applications, e.g., in robotics, aerospace, and automotive domains, but require careful designs to meet stringent safety demands. Formal verification techniques justify such safety properties but need to handle mathematical models of CPSs called hybrid systems, i.e., those that combine the discrete dynamics of stepwise controller computations with the continuous dynamics of their differential equations. Modularity principles for the design and formal verification of cyber-physical systems are especially beneficial when a system consists of many cooperating entities that together must satisfy some safety criteria. This talk discusses how differential dynamic logic (dL) for hybrid systems can be used to model and verify CPS in a modular fashion. Its theorem prover KeYmaera X provides compositional verification techniques for hybrid systems, which not only handle nonlinear systems but also use invariants to reduce the verification of larger systems to subsystems. For very large models, component-based modeling can be used to split large models into multiple component models with local responsibilities to further reduce modeling complexity.

References

1 Brandon Bohrer, Yong Kiam Tan, Stefan Mitsch, Magnus O. Myreen, and André Platzer. VeriPhy: Verified controller executables from verified cyber-physical system models. In Dan Grossman, editor, *PLDI*, pages 617–630. ACM, 2018.
2 Nathan Fulton, Stefan Mitsch, Jan-David Quesel, Marcus Völp, and André Platzer. KeYmaera X: An axiomatic tactical theorem prover for hybrid systems. In Amy Felty and Aart Middeldorp, editors, *CADE*, volume 9195 of *LNCS*, pages 527–538, Berlin, 2015. Springer.
3 Nathan Fulton and André Platzer. Safe reinforcement learning via formal methods: Toward safe control through proof and learning. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *AAAI*. AAAI Press, 2018.
4 Andreas Müller, Stefan Mitsch, Werner Retschitzegger, Wieland Schwinger, and André Platzer. Tactical contract composition for hybrid system component verification. *STTT*, 20(6):615–643, 2018. Special issue for selected papers from FASE'17.
5 Andreas Müller, Stefan Mitsch, Wieland Schwinger, and André Platzer. A component-based hybrid systems verification and implementation tool in keymaera X (tool demonstration). In Roger D. Chamberlain, Walid Taha, and Martin Törngren, editors, *Cyber Physical Systems. Model-Based Design – 8th International Workshop, CyPhy 2018, and 14th International Workshop, WESE 2018, Turin, Italy, October 4-5, 2018, Revised Selected Papers*, volume 11615 of *LNCS*, pages 91–110. Springer, 2018.
6 Stefan Mitsch and André Platzer. ModelPlex: Verified runtime validation of verified cyber-physical system models. *Form. Methods Syst. Des.*, 49(1-2):33–74, 2016. Special issue of selected papers from RV'14.
7 André Platzer. A complete uniform substitution calculus for differential dynamic logic. *J. Autom. Reas.*, 59(2):219–265, 2017.

3.12 Probabilistic Model Checking for Safety and Performance Guarantees

David Parker (University of Birmingham, GB)

This talk gives an overview of the state of the art in probabilistic model checking, with a particular focus on the theme of the seminar: formally analysing collections of autonomous robots. I will describe some recent related applications of these techniques, including synthesising autonomous mobile robot plans with probabilistic guarantees and verifying adaptive mission plans for unmanned underwater vehicles. Motivated by the application challenge for the seminar, I will also summarise some recent directions on verification for partially observable models, stochastic games and multi-robot systems.

3.13 Modelling and Verification using RoboChart

Pedro Ribeiro (University of York, GB)

Joint work of Pesro Ribeiro, James Baxter, Ana Cavalcanti, Madiel Conserva, Simon Foster, Wei Li, Alvaro Miyazawa, Pedro Ribeiro, Augusto Sampaio, Jon Timmis, Jim Woodcock
Main reference Alvaro Miyazawa, Pedro Ribeiro, Wei Li, Ana Cavalcanti, Jon Timmis, Jim Woodcock: "RoboChart: modelling and verification of the functional behaviour of robotic applications", Software and Systems Modeling, Vol. 18(5), pp. 3097–3149, 2019.
URL http://dx.doi.org/10.1007/s10270-018-00710-z

Designing robotic systems can be very challenging, yet controllers are often specified using informal notations with development driven primarily by simulations and physical experiments, without clear relation to abstract models of requirements. Our goal is to support roboticists in writing models and applying modern verification techniques using a language familiar to them. To that end, we consider RoboChart, a domain-specific modelling language based on UML, but with a restricted set of constructs to enable a simplified formal semantics and automated reasoning. It supports the specification of reactive, timed and probabilistic behaviours. We illustrate how RoboChart can be used to specify the behaviour of individual robots in the context of the smart farm. We pursue an analysis of the collective using a discrete model of the environment and the model-checker FDR.

References

1 A. Miyazawa, P. Ribeiro, W. Li, A. Cavalcanti, J. Timmis, and J. C. P. Woodcock. Robochart: modelling and verification of the functional behaviour of robotic applications. *Software & Systems Modeling*, 18(5):3097–3149, Oct 2019.
2 A. Miyazawa, P. Ribeiro, W. Li, A. Cavalcanti, and J. Timmis. Automatic property checking of robotic applications. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3869–3876, Sep. 2017.
3 A. Cavalcanti, A. Sampaio, A. Miyazawa, P. Ribeiro, M. Conserva Filho, A. Didier, W. Li, and J. Timmis. Verified simulation for robotics. *Science of Computer Programming*, 174:1 – 37, 2019.

3.14 Optimization of the watering schedule by run-time and design-time analysis

Masaki Waga (National Institute of Informatics – Tokyo, JP)

By design-time verification of a real-time model (e.g., timed automata or time Petri Nets), we can verify if there are any potential deadline misses. However, to confirm the verified deadline is reasonable, we have to model the environment, or we have to exploit some empirical knowledge (e.g., previous environmental data). In this talk, I will talk about a data-driven approach to confirm the deadline in the watering by robots. Typically, we modeled the watering by the robots and the change of the water level of the fields, and show how to obtain the safe set of the watering intervals by symbolic monitoring, which is one of the run-time verification methods. As an example of the watering strategy, we also show that a simple round-robin strategy can be modeled by a parametric timed automaton and the worst-case watering interval can be obtained e.g., by IMITATOR.

3.15 Formal modeling and analysis of real-time systems using Real-Time Maude

Peter Csaba Ölveczky (University of Oslo, NO)

Real-Time Maude is a tool that extends the rewriting-logic-based Maude system to support the executable formal modeling and analysis of real-time systems. Real-Time Maude is characterized by its general and expressive, yet intuitive, specification formalism, and offers a spectrum of formal analysis methods, including: rewriting for simulation purposes, search for reachability analysis, and both untimed and metric temporal logic model checking. Real-Time Maude is particularly suitable for specifying real-time systems in an object-oriented style, and its flexible formalism makes it easy to model different forms of communication.

This modeling flexibility, and the usefulness of Real-Time Maude for both simulation and model checking, has been demonstrated on many advanced state-of-the-art applications, including both distributed protocols of different kinds and industrial embedded systems. Furthermore, Real-Time Maude's expressiveness has also been exploited for defining the formal semantics of MDE languages for real-time/embedded systems, including Ptolemy discrete-event models, a subset of the avionics modeling standard AADL, and a timed extension of the MOMENT2 model transformation framework. Real-Time Maude thereby provides formal model checking capabilities for these languages for free, and such analysis has been integrated into the tool environment of a number of modeling languages.

This talk gives a high-level overview of Real-Time Maude and some of its applications. The talk also briefly discusses what features of Real-Time Maude and associated Maude-based tools are suitable for certain aspects of the smart farm case study (e.g., object orientation to model robots, the ability to define complex data types and functions to model, e.g., areas and collision courses, and so on) and for which aspects of the case study the tool environment seems less suitable (e.g., complex continuous behaviors).

4 Open problems

4.1 Specification of the Application Challenge

Mario Gleirscher (University of York, GB), Anne E. Haxthausen (Technical University of Denmark – Lyngby, DK), Martin Leucker (Universität Lübeck, DE), and Sven Linker (University of Liverpool, GB)

The following material was provided to and used by the seminar participants to present their approach in the context of common application domain.

4.1.1 Purpose of this Specification

In the following, we describe a scenario, where several autonomous robots solve a common task. The intention behind this description is to provide a framework for the discussion within the seminar. To that end, we invite you to model parts of the scenario with formalisms of your choice. However, we **do not expect** that you model the whole scenario, but encourage you to pick the parts that you are interested in. Furthermore, even if your formalisation for certain aspects is not complete, we appreciate comments on whether this is due to your choice of formalism, or for other reasons.

The *main goal of this exercise* is to identify common ground between different formalisations and approaches, and how they could be used in combination to enhance modelling and analysis of such scenarios. In other words,

1. when similar aspects of this challenge have been modelled by different seminar participants, we expect to discuss the *differences as well as advantages and disadvantages* of each approach, and
2. when complementary aspects have been modelled, we expect a discussion of *how these models are related* and together contribute to the assurance of the overall plant.

4.1.2 The Challenge

The scenario we consider is an instance of *smart farming*. A local farm consists of several fields and green houses, where fruit and vegetables are grown. The farm and the fields are connected by public roads, which may (and will) be used by the general public, as well as the agricultural machines.

Each field is covered by sensors detecting the moisture levels of the ground. The farm employs several different autonomous robots: On the one hand, we have *worker robots*, which are used both for maintenance, that is to repair other robots, as well as for plant care, that is to cut, water, and fertilise the plants. On the other hand, we have *transportation robots*, which harvest, collect the harvested plants and transport them to delivery stations. Robots of each category can be used for all of the tasks within the category. For example, any worker robot can water plants, or be used to cut the plants. For worker robots, the farm uses both flying robots, as well as robots driving on the ground, while all transportation robots are ground-based.

We assume that there is *no central controlling element*, and that the robots do not have the full knowledge about everything in the environment.

However, the farm still employs humans who maintain the machines, and who may take over some of the responsibilities (e.g., harvesting fields or cutting plants). Hence, the robots need to take the *presence of humans* into account, and need to adapt their behaviour accordingly. In particular, this means it is always possible that *manually operated machines* (for maintenance, plant care, harvesting or transportation, or simply other traffic) may be present in the farm and/or on the roads, as well as humans outside of any vehicles.

Goals

- Ensure safety of all entities involved, in particular working personnel and general public using connecting streets
  - Low-level safety: obstacle avoidance, collision avoidance
  - High-level safety: exclusive access to working areas
  - Avoidance of other hazards
- Optimise yield of farm and reduce potential losses during fertilisation, watering harvest and transportation

Table 2 Information on the actors in the smart farm.

Entity	Purpose	Number	Information Type
Field	Grow vegetables (salad, potatoes, turnips) or grains (wheat, rye)	4	Global
Green House	Grow vegetables or fruit (bell peppers, tomatoes, cucumbers, peaches)	2	Global
Worker Robot (Flying/Ground)	Plant crops, water and fertilise fields, repair other robots	3/2	Local
Harvester/ Transporter	Harvest plants and transport goods between farms/greenhouses and delivery station	3	Local

4.1.3 A cutout of typical activities in the smart farm

Example use case

1. Field X is *empty*
2. Robot A drives to X and plants potatoes
3. Robot B waters X
4. Robot B applies fertiliser to X
5. Field X is now in state *growing*, while steps 3 and 4 may be repeated
6. When field X (or rather the sensors on field X) sends message that plants are ripe (state *harvest*): Robot C comes to harvest potatoes
7. Robot C delivers the potatoes to the farm collection point

Example of an emergency scenario

1. Robot D detects utility vehicle on its path
2. Robot D avoids crash by replanning path

Further example of an emergency scenario

1. Robot E crashes into road/field-side ditch and gets immobile or collides with an object and gets damaged
2. Maintenance service is notified
3. Unoccupied worker takes care or issue will be delegated to supervisory control

4.1.4 Actors in the smart farm

Table 2 contains all different actor types of the smart farm. The first and second column contains the name and purpose of each category of actors, while the third column contains the number of single entities in each category. The final column denotes, whether the information about entities in this category is available to all other entities (global information), or only within each single entity (local information).

4.1.5 Layout of the smart farm

The fields and green houses are all connected to the collecting point on public streets. However, while the green houses can be approached separately and independent from each other, the fields share a common road for the approach. That is, the layout can be imagined as in Fig. 2: C denotes the collecting point, the rectangles marked by G are green houses and the rectangles marked with F are the fields. The lines in between indicate the road structure.

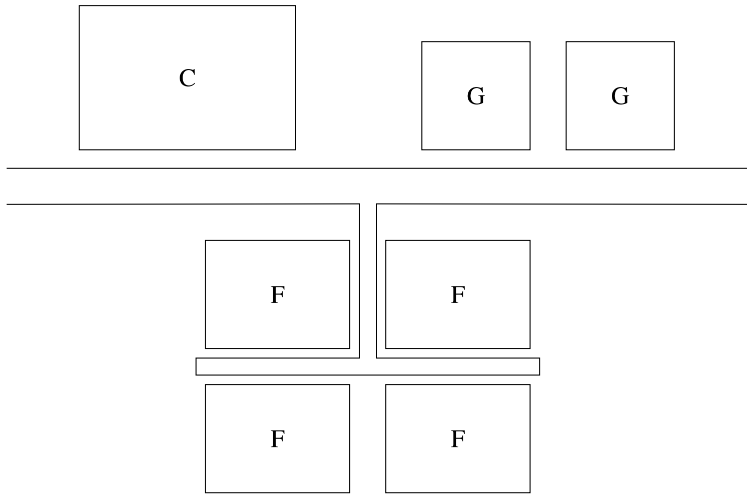

Figure 2 Layout of the smart farm.

4.1.6 Modelling Parameters according to Abstraction Level

For a more structured discussion, we distinguish several levels of details for this system, the environment and the hazards, which refer to the level of detail for the physicality of the system. The different levels are

1. Discrete
2. Real-Time
3. Physical

The first level contains the purely discrete aspects of the system components. That is, communication channels, structure and data, as well as possible (discrete) states of each autonomous entity. The second level incorporates real-time aspects of the behaviour, for example durations and time bounds. The third and final level includes more physical laws, for example in the form of differential equations. All of these models may include probabilistic aspects, or, in the case of real-time and physical models, limits on how exact durations and time bounds can be satisfied.

Generally, we assume that suitable sensors provide information about the different entities, and that this information may be shared via suitable channels (message passing, ...) For simplicity, we assume that this information is always correct, if not stated otherwise.

In order to focus and integrate the modelling approaches during the seminar, we strongly encourage you to use the following modelling parameters that are supposed to represent the variables of the Smart Farm state space. However, if you need to change these parameters, please be transparent about this in your model and its presentation.

Parameters and Parameter Types for Discrete Modelling:

- Map (areas/road segments):
  - state: *occupied, empty*

  You can assume that there is an attributed map available (to all vehicles) with geometry data (precision .5 meters). Depending on the activity and on a per-vehicle basis, SLAM[1] might be used to update volatile attributes of the area in the mapping information (local

[1] Simultaneous localisation and mapping

to a vehicle). Markers with high precision (+/-10cm) at convenient but practical places of the map can also be used for mapping and positioning.

- Resource (Field/Greenhouse):
  - contents: *peppers, salad, turnips, potatoes, wheat, rye, empty*
  - state: *harvest, growing, empty*
  - water level: *low, good*
  - fertiliser level: *low, good*
  - Invariants: $state\,empty \implies (contents\,empty \wedge water\_level\,good \wedge fertiliser\_level\,good)$
- Worker Robots:
  - cargo_type: *water, fertiliser*
  - movement: *ground, flying*
  - cargo: *full, empty*
  - or alternatively cargo: *(finite) set of values in [0,1], where 0 means empty, 1 means full*
- Harvester/Transporter:
  - state: *harvesting, transport to drop off*
  - cargo: *full, empty*
  - alternatively cargo: *(finite) set of values in [0,1], where 0 means empty, 1 means full*

Parameters and Parameter Types for (Distributed) Real-Time Modelling:

- Resource
  - The only real-time aspects for the resources would be the duration plants need to grow. However, since the time-scale of these durations is very different from communication and other aspects, we refrain from any further specification of this aspect.
- Communication
  - message delay: *del_t* seconds from sending to reception
  - localisation messages may have different delays:
    - global positioning (GPS, precision $\pm 2m$): *gps_t*
    - local positioning (with respect to finite set of fixed markers, precision $\pm .01m$): *loc_t*
- Worker Robots
  - filling up the cargo bay from empty to full: *care_fill_t*
- Harvester/Transporter
  - filling up the cargo bay from empty to full: *trans_fill_t*
- Relations between parameters: $del\_t < care\_fill\_t < trans\_fill\_t$

Parameters and Parameter Types for Continuous/Physical Modelling:

- Vehicles
  - speed
  - position
  - maxaccel
  - maxdecel
  - maxspeed: 30 kph
  - $1m \leq length \leq 5m$
  - $300kg \leq weight \leq 5000kg$
- Human traffic on public streets (bicycles and cars)
  - $15kph < speed < 60kph$
- Human traffic on farm streets (bicycles and pedestrians)
  - $3kph < speed < 20kph$

Failure probabilities

- Resources
  - Rotting goods: *.02/h*
- Ground Based Vehicle
  - Failure rate: .05/h
- Flying Vehicle
  - Failure rate: .1/h
- Message loss
  - *p_m_loss*
- Probabilities of humans (on bicycles or in cars) on public streets
  - *p_h_public*
- Probabilities of humans (pedestrians, or on bicycles) on farm streets:
  - *p_h_farm*

Feel free to refine these uncertainties (e.g. probability of vehicles on roads) by introducing further parameters.

4.1.7 Properties

The following properties of the Smart Farm control scheme refine the goal of the seminar.

Safety Constraints (depending on environment: Public street, street between fields, ...)

- Public road
  - avoids vehicles from general public (cars driving through, bicyclists, pedestrians)
  - avoids colliding with other utility vehicles
- Rural road
  - avoids working personnel (trained, but may still make errors)
  - avoids colliding with other utility vehicles
- Fields
  - avoids colliding with other utility vehicles

These constraints should depict the variety of collision situations to be encountered in the Smart Farm. It is of course possible to cover these constraints with a generalised constraint of the form: "Avoid collision with any moving vehicle or person or any static object in the Smart Farm."

Productivity Requirements (Liveness, Progress)

- Resources
  - Harvest ripe goods timely (alternatively: plants shall not rot on the fields/in the green houses)

Participants

- Étienne André
University of Paris North, FR
- Andreas Bauer
KUKA Systems GmbH – Augsburg, DE
- Sergiy Bogomolov
Newcastle University, GB
- Jörg Brauer
Verified Systems International GmbH – Bremen, DE
- Radu Calinescu
University of York, GB
- Alessandro Fantechi
University of Florence, IT
- Frederik Foldager
Aarhus University, DK
- Sibylle Fröschle
OFFIS – Oldenburg, DE
- Mario Gleirscher
University of York, GB
- Anne E. Haxthausen
Technical University of Denmark – Lyngby, DK
- Christian Heinzemann
Robert Bosch GmbH – Stuttgart, DE
- Jean-Baptiste Jeannin
University of Michigan – Ann Arbor, US
- Kim Guldstrand Larsen
Aalborg University, DK
- Peter Gorm Larsen
Aarhus University, DK
- Martin Leucker
Universität Lübeck, DE
- Sven Linker
University of Liverpool, GB
- Stefan Mitsch
Carnegie Mellon University – Pittsburgh, US
- Laura Nenzi
University of Trieste, IT
- Peter Csaba ölveczky
University of Oslo, NO
- David Parker
University of Birmingham, GB
- Pedro Ribeiro
University of York, GB
- Masaki Waga
National Institute of Informatics – Tokyo, JP

Report from Dagstuhl Seminar 19442

Programming Languages for Distributed Systems and Distributed Data Management

Edited by

Carla Ferreira[1], Philipp Haller[2], and Guido Salvaneschi[3]

1 **Universidade NOVA de Lisboa, PT,** carla.ferreira@fct.unl.pt
2 **KTH Royal Institute of Technology – Stockholm, SE,** phaller@kth.se
3 **TU Darmstadt, DE,** salvaneschi@st.informatik.tu-darmstadt.de

Abstract

Programming language advances have played an important role in various areas of distributed systems research, including consistency, communication, and fault tolerance, enabling automated reasoning and performance optimization. However, over the last few years, researchers focusing on this area have been scattered across different communities such as language design and implementation, (distributed) databases, Big Data processing and IoT/edge computing – resulting in limited interaction. The goal of this seminar is to build a community of researchers interested in programming language techniques for distributed systems and distributed data management, share current research results and set up a common research agenda. This report documents the program and the outcomes of Dagstuhl Seminar 19442 "Programming Languages for Distributed Systems and Distributed Data Management."

Seminar October 27–31, 2019 – http://www.dagstuhl.de/19442

2012 ACM Subject Classification Software and its engineering → General programming languages, Computer systems organization → Distributed architectures, Information systems → Data management systems, Computer systems organization → Embedded and cyber-physical systems

Keywords and phrases Programming Languages, Distributed Systems, Data Management

Digital Object Identifier 10.4230/DagRep.9.10.117

1 Executive Summary

Carla Ferreira
Philipp Haller
Guido Salvaneschi

Developing distributed systems is a well-known, decades-old problem in computer science. Despite significant research effort dedicated to this area, programming distributed systems remains challenging. The issues of consistency, concurrency, fault tolerance, as well as (asynchronous) remote communication among heterogeneous platforms naturally show up in this class of systems, creating a demand for proper language abstractions that enable developers to tackle such challenges.

Over the last years, language abstractions have been a key for achieving the properties above in many industrially successful distributed systems. For example, MapReduce takes advantage of purity to parallelize task processing; complex event processing adopts declarative programming to express sophisticated event correlations; and Spark leverages functional programming for efficient fault recovery via lineage. In parallel, there have been notable advances in research on programming languages for distributed systems, such as conflict-free replicated data types, distributed information-flow security, language support for safe distribution of computations, as well as programming frameworks for mixed IoT/cloud development.

However, the researchers that have been carrying out these efforts are scattered across different communities which include programming language design, type systems and theory, database systems and database theory, distributed systems, systems programming, data-centric programming, and web application development. This Dagstuhl Seminar brought together researchers from these different communities.

The seminar focused on answering the following major questions:

- Which abstractions are required in emergent fields of distributed systems, such as mixed cloud/edge computing and IoT?
- How can language abstractions be designed in a way that they provide a high-level interface to programmers and still allow fine-grained tuning of low-level properties when needed, possibly in a gradual way?
- Which compilation pipeline (e.g., which intermediate representation) is needed to address the (e.g., optimization) issues of distributed systems?
- Which research issues must be solved to provide tools (e.g., debuggers, profilers) that are needed to support languages that target distributed systems?
- Which security and privacy issues come up in the context of programming languages for distributed systems and how can they be addressed?
- What benchmarks can be defined to compare language implementations for distributed systems?

The seminar accomplished the goal of bringing together the research communities of databases, distributed systems, and programming languages. The list of participants includes 24 academic and industrial researchers from Austria, Belgium, France, Germany, Portugal, Sweden, Switzerland, UK, and USA, with complementary expertise and research interests. The group had a balanced number of senior researchers and junior researchers, as well as a strong industrial representation.

The scientific program comprised 28 sessions. The sessions devoted to individual presentations included 16 short talks with a maximum duration of 15 minutes and 6 long contributed talks with a maximum duration of 35 minutes. In addition, the seminar included 2 plenary sessions and 4 group sessions. The first two days of the seminar were dedicated to research talks, but it was ensured that each talk had allocated time for discussions and exchange of ideas. In the two following mornings there were 3 plenary sessions and 2 parallel group sessions. The topics for these sessions were proposed and selected after a lively discussion between participants, where the most popular sessions were promoted to plenary and the remaining occurred in two parallel sessions. The scientific sessions discussed and collected open questions on the topics of: programming models and abstractions; security and privacy; static guarantees, type systems, verification; distributed computing for the edge; time, synchrony, and consistency; and persistency and serialization. There was also a social topic discussing further actions to bring the three communities together. Even though there are overlapping research interests, there is a difference of values between communities that needs to be acknowledged and tackled. Participants agreed on the goal of organizing follow-up events to further strengthen the connection among the database, the distributed systems and the programming languages communities. In particular, the importance of extending future events to Ph.D. students, for instance with an integrated Summer School, has been discussed.

2 Table of Contents

Executive Summary
Carla Ferreira, Philipp Haller, and Guido Salvaneschi, ... 117

Overview of Talks

Aggregation ≠ Replication
Carlos Baquero 121

Stateful serverless programming
Sebastian Burckhardt 122

Programming for autonomy
Amit Chopra 122

Access control for highly-available transactional data stores
Annette Bieniusa 123

Automating the deployment of complex distributed systems
Uwe Breitenbücher 123

Scaling distributed systems reliably
Natalia Chechina 124

Cloud + Big Data: Implications for structured data platforms
Surajit Chaudhuri 124

Programming elastic services with AEON and PLASMA
Patrick Eugster 125

Verification of message-passing programs
Damien Zufferey 125

Selected challenges in concurrent and distributed programming
Philipp Haller 126

HipHop.js
Manuel Serrano 126

Distributed systems – The next level
Schahram Dustdar 126

Actors revisited for predictable systems
Edward A. Lee 127

Toward high-level programming for distributed systems
Laurent Prosperi 127

Engineering distributed data-intensive applications
Guido Salvaneschi 128

Just-right consistency & The programming continuum
Marc Shapiro 128

Debugging of actor programs using Rebeca model checking tool
Marjan Sirjani 129

Designing distributed systems with piecewise relative observable purity
Peter Van Roy 129

How can concurrent data structures inspire distributed data structures and how to implement efficient language prototypes "for free"
Aleksandar Prokopec 130

Invariant-preserving applications for weakly consistent replicated databases
Carla Ferreira 130

The global object tracker (GoT)
Rohan Achar 131

Data programming for ML and Data Science – Challenges for data management, compilers, and distributed systems
Volker Markl 131

Participants 133

3 Overview of Talks

3.1 Aggregation $\neq$ Replication

Carlos Baquero (University of Minho – Braga, PT, cbm@di.uminho.pt)

Both distributed aggregation and replication for high availability are techniques that can help tackle geo-replication, offline operation and edge/fog computing. Distributed aggregation often shares many properties in common with CRDT style convergent replication, but they are not the same concept. The main difference is that in replication there is an abstraction of a single replicated state that can be updated in the multiple locations where a replica is present. This state is not owned by any given replica, but any replica can evolve it by applying operations that transform the shared state. This notion applies both in strong consistency and high availability settings. The difference being that in highly available replication the replicas are allowed to diverge and later reconcile. Another factor is that operations that lead to state changes are often the result of the activity an external user that interacts with the system, e.g. adjusting the target room temperature up by 2 degrees. As such, different users, can do conflicting actions, either concurrently or in sequence (most of us did in their childhood on/off light switching fights with other kids and adults).

Distributed data aggregation refers to several data aggregation techniques that are common in sensor network settings and datacenter infrastructure monitoring. In contrast to replication, each node/location has access to its own local data, e.g. CPU utilisation or a local measurement of humidity levels, and typically this data can evolve continuously. Also, the data to be aggregated is often not directly controlled by users, it usually results from an external physical process or the result of complex system evolutions. Thus, each sensing node usually has exclusive access to a local input value that evolves in time. The aggregation process is then tasked with collecting and transforming this information, e.g. calculating the average or the maximum value, and making it available at a specified location (sink) or disseminating it back to the nodes (by broadcasting the aggregate result). In aggregation the source of truth for each individual measurement is in the actual node that provided it.

Sometimes the two concepts have in common the notion of data merging. In state-based CRDTs operations are reflected in a semi-lattice state that can be combined with others with a join function. In data aggregation there is also often a notion of joining data together, but there is an additional aspect of data reduction and summarisation that is usually not present in CRDT designs. To add to the confusion, it's is possible to combine the two concepts in a single system, as we did in the design of Scalable Eventually Consistent Counters, that combines a hierarchical CRDT design with a global aggregation and reporting facet. However, ignoring corner cases, the difference can be quite clear and recognising it can help in selecting the right tools. A final take-away example is to consider the control of room temperature: The plus/minus control that sets the set point temperature can be captured by a CRDT; The combining of different temperature sensors across the room to obtain the average temperature is distributed aggregation.

3.2 Stateful serverless programming

Sebastian Burckhardt (Microsoft Research Lab – Redmond, US, sburckha@microsoft.com)

Serverless programming models, such as AWS Lambdas or Azure Functions, simplify the development of elastic cloud services by automating low-level aspects of deployment, VM management, and monitoring. However, building a stateful application from stateless functions still poses some challenges for developers, such as handling partial execution failures, or enforcing proper synchronization of conflicting operations. In Azure Durable Functions we offer several features to aid developers in that regard: orchestrations provide reliable workflows, entities provide reliable application objects, and critical sections provide reliable multi-object synchronization. The resulting programming model combines aspects of both the actor model and shared memory, but can execute reliably in a distributed serverless context, and is guaranteed to not deadlock.

3.3 Programming for autonomy

Amit Chopra (Lancaster University, UK, amit.chopra@lancaster.ac.uk)

How do we program systems that involve multiple autonomous principals?

To address the question, we must understand what autonomy means. Autonomy means decentralization: principals in a system exercise independent decision making and engage via arms-length communications. However, not all engagements can be correct (if they were, we would have no system). This motivates the notion of norms as the basis for determining the correctness of their engagements. Norms act as a counterbalance to autonomy: do what you please but not everything goes.

Norms must be operationalized in a decentralized setting via information protocols. An information protocol specifies the ordering and occurrence of events in a decentralized setting by specifying causality and integrity constraints. An information protocol can be correctly enacted by endpoints over an asynchronous, unordered communication infrastructure based only upon local knowledge. This is a significant departure from existing work in computing, which typically does not specify causality and instead relies on stronger infrastructure assumptions (e.g., pairwise FIFO or causal delivery).

In a nutshell, any specification of a system of autonomous principals must be based upon norms and information protocols. A rigorous study of these ideas and programming based upon them will enable exciting novel kinds of systems, e.g., based upon agreements and contracts – what most business on our planet is based upon.

3.4 Access control for highly-available transactional data stores

Annette Bieniusa (TU Kaiserslautern, DE, bieniusa@cs.uni-kl.de)

Access control systems for data stores regulate which users are allowed to read or update a specific item. For long term deployments, it is typically required that these policies can dynamically change as the system and its user base evolves. In this talk, we discuss the challenges these adaptable security policies raises in highly available data stores that allow for concurrent modifications and tolerate partial network partitions. By formally deriving the consistency guarantees for access control and data modifications, we formulate the requirements on the involved system components and their interplay. We further present ACGreGate, a Java framework for implementing correct access control layers for the transactional CRDT store AntidoteDB. This is joint work with Mathias Weber and Arnd Poetzsch-Heffter.

3.5 Automating the deployment of complex distributed systems

Uwe Breitenbücher (Universität Stuttgart, DE, uwe.breitenbuecher@iaas.uni-stuttgart.de)

The automation of application deployment is critical because deploying systems manually is too error-prone, time-consuming, and costly. Therefore, several deployment automation technologies have been developed in recent years. However, to deploy complex distributed systems, it is often necessary to combine several of these deployment technologies as their capabilities differ considerably. Unfortunately, such an integration is a complex technical issue as each technology has its own deployment metamodel and API. Our first step to tackle this issue was the introduction of the Essential Deployment Metamodel (EDMM), which is a normalized metamodel for deployment models that can be mapped to the 13 most important deployment technologies including, e. g., Terraform, CloudFormation, and TOSCA. However, the current EDMM Transformation Framework only supports transforming an EDMM model into one certain deployment technology, which restricts its applicability, as typically multiple deployment technologies need to be combined for deploying complex systems. Therefore, we are working on an extension that is capable of automatically splitting and transforming one EDMM model to several deployment models supported by different deployment technologies. Moreover, the extension also generates an imperative workflow model that invokes the different deployment technologies involved with the corresponding deployment models. This enables to fully automate the deployment of complex distributed systems.

3.6 Scaling distributed systems reliably

Natalia Chechina (University of Bournemouth – Poole, UK, nchechina@bournemouth.ac.uk)

Erlang is a well-known programming language in the areas of distributed databases and large-scale messaging applications, e.g., WhatsApp with 1.5Bn monthly users. However, when it comes to safety critical systems and robotics in particular, people who never worked with Erlang are sceptical regarding its usefulness and the applicability of its principles. In this talk I will share research and findings of applying Erlang's non-defensive programming approach and "let it crash" philosophy to enable fault tolerance and scalability of robots. I will also share findings regarding Communication Scaling Limit Volume (CSLV) which states that in a team of robots the volume of data remains constant and is in direct proportion with the number of nodes, size and number of messages.

3.7 Cloud + Big Data: Implications for structured data platforms

Surajit Chaudhuri (Microsoft Research Lab – Redmond, US, surajitc@microsoft.com)

The combination of the Cloud and Big Data has led to significant architectural rethinking in the database community because of the need to accommodate requirements of Compute Elasticity and the diversity of the Big Data Platforms that encompass SQL Data Warehousing, Spark, and other emerging platforms that support distributed ML. There is also increased urgency to support Approximate Data Analysis as data volumes continue to grow exponentially. Another long standing pain point further amplified by Big Data is data cleaning and data transformation, an essential pre-processing step for querying as well as advanced analytics to generate valuable insight. Despite much research activities, we don't yet have a DSL that has both broad applicability and helps lower the complexity of this important step for the programmers. In this talk, we will review these disruptions and challenges and sketch a few of the promising directions. Specifically, we will discuss the progress we have made in approximate query processing through injection of two new sampling operators in the query language and incorporating them in the Query Optimization step. However, leveraging such operators require sophistication and so we are still far from democratizing approximate query processing. In the area of data cleaning and transformation, we will examine the promise of Program Synthesis. For many of these problems, there is a unique opportunity for Programming Language researchers and database researchers to work together to address the open challenges.

3.8 Programming elastic services with AEON and PLASMA

Patrick Eugster (University of Lugano, CH, patrick.thomas.eugster@usi.ch)

Implementing distributed services that automatically scale in and out in response to workload changes in order to run efficiently in third-party cloud datacenters is a hard task for programmers. In this talk we present two contributions towards simplifying the development of such elastic services. The first contribution is a variant of the actor programming model which provides strong consistency (i.e., serializability) without hampering the actor model's strong potential for scalability – a prerequisite for elasticity. That is, programmers can perform calls across multiple actors in so-called "events" without interference with other events. Our model leverages a DAG-based arrangement of actors with a novel corresponding synchronization protocol in order to efficiently execute such events, showing substantial speedups over traditional 2-phase locking while similarly avoiding races and deadlocks. The second, independent, contribution consists in augmenting the actor programming model with a second "layer" of programming to support fine-grained elasticity. That is, this layer allows programmers to specify high-level program conditions hinting to scalability bottlenecks (e.g., CPU usage beyond a certain threshold, too high rate of messages between certain actors), and corresponding mitigation actions (e.g., migrate certain actors to hosts with available CPU cycles, co-locate actors with other actors they interact with). As we show, policies expressed in this way consisting in only a few lines allow applications to substantially reduce resource usage and/or improve performance by better distributing load.

3.9 Verification of message-passing programs

Damien Zufferey (MPI-SWS – Kaiserslautern, DE, zufferey@mpi-sws.org)

In this talk, I will show how we can harness the synergy between programming languages and verification methods to help programmers build reliable software. Often there is a mismatch between what the programming model allows and its applications. Better programming models can (1) remove unneeded expressive power and (2) make it easy, for a verifier, to decompose the program into smaller parts which can be verified separately. I will first look at fault-tolerant distributed algorithms where we integrate a scoping mechanism for communication as part of the program syntax. The key insight is the use of communication-closure (logical boundaries in a program that messages should not cross) to structure the code. This structure element greatly simplify the programming and verification of fault-tolerant distributed algorithms. Then I will explain how we can use session types to reason about cyber-physical systems in combination with assume-guarantee reasoning. Assume-guarantee reasoning is designed to compose the behaviors of multiple components (bottom-up composition). On the other hand, session types are carefully designed to make a global specification projectable on the individual components in the systems (top-down decomposition).

3.10 Selected challenges in concurrent and distributed programming

Philipp Haller (KTH Royal Institute of Technology – Stockholm, SE, phaller@kth.se)

We present three challenges in concurrent and distributed programming, as well as recent results addressing them. The first challenge consists of ensuring fault-tolerance properties in typed programming languages. The main question is how to enforce fault-tolerance properties for well-typed programs, as opposed to specific algorithms or systems. Towards addressing this question, we present the first correctness results for a typed calculus with first-class lineages. The second challenge consists of using data with different consistency properties safely within the same distributed application. To address this challenge, we propose a novel type system which provides a noninterference guarantee: mutations of potentially-inconsistent data cannot be observed via access to consistent data types. As a third challenge we propose the design of a concurrent domain-specific language for parallelizing static analysis problems.

3.11 HipHop.js

Manuel Serrano (INRIA Sophia Antipolis, FR, Manuel.Serrano@inria.fr)

HipHop is a synchronous reactive language for the web and IoT. It adds synchronous concurrency and preemption to Hop, which is itself an asynchronous multitier extension of JavaScript. Inspired from Esterel, HipHop simplifies the programming of non-trivial temporal behaviors as found in complex web interfaces or IoT controllers and the cooperation between synchronous and asynchronous activities. HipHop is compiled into plain sequential JavaScript and executes on unmodified runtime environments. In this presentation we show two examples to present and discuss HipHop: a simple web login form to introduce the language and show how it differs from JavaScript, and a real life example, an interactive music system that show why concurrency and preemption help programming such temporal applications. A live demo of the musical application will be given.

3.12 Distributed systems – The next level

Schahram Dustdar (TU Wien, AT, dustdar@dsg.tuwien.ac.at)

As humans, things, software and AI continue to become the entangled fabric of distributed systems, systems engineers and researchers are facing novel challenges. In this talk, we analyze the role of Edge, Cloud, and Human-based Computing as well as AI in the co-evolution of distributed systems for the new decade. We identify challenges and discuss a roadmap that these new distributed systems have to address. We take a closer look at how a cyber-physical fabric will be complemented by AI operationalization to enable seamless end-to-end distributed systems.

3.13 Actors revisited for predictable systems

Edward A. Lee (University of California – Berkeley, US, eal@berkeley.edu)

Concurrent and distributed software based on publish-and-subscribe and actors are sometimes used to realize distributed cyber-physical systems. Broadly, these mechanisms compose software components that have private state and communicate with each other via message passing. However, the underlying message-passing mechanisms are less deterministic than they could be. In this talk, I described some simple challenge problems that are common in distributed cyber-physical systems and extremely difficult to solve using either actors or publish-and-subscribe. I offered an alternative model of computation that we call "reactors" that solves these problems simply and elegantly and that is able to leverage decades of results from the real-time systems community. The reactors model is being implemented in a coordination language called Lingua Franca. A key feature is that extends messages with logical timestamps that provide a semantic ordering and a semantic notion of simultaneity. By leveraging synchronized clocks, an efficient distributed implementation guarantees determinacy when network latencies and clock synchronization error remain below assumed bounds.

3.14 Toward high-level programming for distributed systems

Laurent Prosperi (Panthéon-Sorbonne University – Paris, FR, laurent.prosperi@inria.fr)

Programming distributed systems is arduous because of failures, asynchrony and trade-offs (e.g. CAP). Moreover, requirements will depend on the audience, for instance ranging from productivity to control. Our work aims at mastering the complexity of building distributed systems while keeping fine-grain control and enhancing dependability. Distributed abstractions will help mastering the complexity, a distributed abstraction is composed of specifications and a set of implementations having their own patterns of distribution. Fine-grain control will be achieved by allowing the programmer to create new abstractions (or use a custom implementation of an existing one) and by exposing runtime behaviors (e.g. fault-tolerance) as a first class citizen, by expressing them as distributed abstractions. Dependability will be ensured at the level of distributed abstractions, by providing, and dynamic checking, formal specifications (internal behaviors, ports, concurrent interactions and consistency) of each of them and of their composition. We believe that the situation is ripe for a new programming environment composed of i) a specification language and ii) a related programming language with embedded dynamic checking of the specifications. Moreover, we will enable the reusability of existing code base (by implementing our approach as a DSL) and of existing systems by allowing assembling pre existing distributed components (e.g. a datastore) as an implementation of a distributed abstraction.

3.15 Engineering distributed data-intensive applications

Guido Salvaneschi (TU Darmstadt, DE, salvaneschi@st.informatik.tu-darmstadt.de)

Over the last few years, ubiquitous connectivity has led to data being constantly generated at an unprecedented rate. As a result, large amounts of data are constantly being processed in an heterogenous infrastructure which stems from the convergence of edge (IoT, mobile) and cloud computing. This poses fundamental engineering challenges on software design, especially with respect to fault tolerance, data consistency, and privacy.

In this presentation, we discuss recent research results we achieved in this context at various levels. We describe an innovative programming framework that improves and simplifies the design of data intensive applications. We also present the use of our programming framework on real-world case studies, emphasising how to achieve fault tolerance and data consistency. Finally, we propose how to account for privacy in the software engineering process for data intensive distributed applications.

3.16 Just-right consistency & The programming continuum

Marc Shapiro (Panthéon-Sorbonne University – Paris, FR, marc.shapiro@acm.org)

Just-right consistency. In a distributed data store, the CAP theorem forces a choice between strong consistency (CP) and availability and responsiveness (AP). To address this issue, we take an application-driven approach, Just-Right Consistency (JRC). JRC derives a consistency model that is sufficient to maintain the application invariants, otherwise remaining as available as possible. JRC leverages application invariant-maintaining patterns. Two, ordered updates and atomic grouping, are compatible with concurrent and asynchronous updates, orthogonally to CAP. In contrast, checking a data precondition on partitioned state is CAP-sensitive. However, if two updates do not negate each other's precondition, they may legally execute concurrently. Updates must synchronise only if one negates the precondition of the other. The JRC approach is supported by the CRDT data model that ensures that concurrent updates converge; by Antidote, a cloud-scale CRDT data store that guarantees transactional causal consistency; and by the CISE static analyser that verifies whether application invariants are guaranteed.

The programming continuum, from core to edge and back. Current cloud architectures, centralised in a few massive data centres, are increasingly moving towards support of edge resources, including localised data centres, points-of-presence, 5G tower micro-DCs, IoT gateways, and far-edge devices. Computing models offered across this spectrum differ vastly, from database-centric in the core, to stream- and notification based at the far edge. When a database system supports notifications, and vice-versa, these are tacked on as an afterthought and not well integrated. Data-sharing models themselves range from weakly to strongly consistent, with blockchains being a bit of both. Indeed, at this scale, CAP and the conflict between correctness and availability is inescapable. Security is often a second-class citizen in distributed system design, as is deployment, monitoring and run-time control. However, we argue that there is no good reason for this proliferation of incompatible models. Developers need access to the full power of distributed computing; they need a common programming model across the whole spectrum, forming a programming continuum. For instance, data access and notifications can be designed to be mutually consistent. Replication can be available-first (based on CRDTs) but designed to seamlessly support stronger synchronisation when required by application semantics. A large system being a composition of parts, composable verification techniques are a key to success. The designer should be able to create and reason about distributed abstractions. To enforce these abstractions, and for security reasons, requires arms-length isolation boundaries. This may use encryption and to branching/merging consistency models, inspired by distributed version control and blockchains. Deployment and monitoring can be programmed using the same abstractions as ordinary computations. The above can be implemented in many different (but mutually compatible) ways, for instance in the core vs. at the far edge.

3.17 Debugging of actor programs using Rebeca model checking tool

Marjan Sirjani (Mälardalen University – Västerås, SE, marjan.sirjani@mdh.se)

Rebeca is designed as an imperative actor-based language with the goal of providing an easy to use language for modeling concurrent and distributed systems, with formal verification support. Timed Rebeca is an extension of Rebeca in which network and computational delays, periodic events, and required deadlines can be expressed in the model. Model checking and simulation tools are built based on the formal semantics of the language. For deadlock-freedom and schedulability analysis special efficient techniques in state space exploration is proposed by exploiting the isolation of method execution in the model. I will briefly show how these models can be used in safety assurance and performance evaluation of different systems, like Network on Chip architectures, sensor network applications, and network protocols. Then I will show how Rebeca can be used for debugging and model-driven development of distributed event-based asynchronous systems.

3.18 Designing distributed systems with piecewise relative observable purity

Peter Van Roy (Catholic University of Louvain, BE, pvr@info.ucl.ac.be)

There exists a useful purely functional subset of distributed programming. Purely functional distributed computations do not interact with the real world (because all inputs must be known in advance), but they support message asynchrony and reordering, and can be used to build networks of communicating agents. General distributed programming consists of purely functional distributed programming plus interaction points for real-world interactions. We are working on a design language, called PROP (Piecewise Relative Observable Purity) to specify distributed systems explicitly as a purely functional core plus interaction points. We aim to turn this into a practical tool that can leverage the powerful techniques available to functional programming for distributed systems design.

3.19 How can concurrent data structures inspire distributed data structures and how to implement efficient language prototypes "for free"

Aleksandar Prokopec (Oracle Labs – Zurich, CH, aleksandar.prokopec@gmail.com)

Most balanced search trees use key comparisons to guide their operations, and achieve logarithmic running time. By relying on numerical properties of the keys, interpolation search achieves lower search complexity and better performance. Although interpolation-based data structures were investigated in the past, their non-blocking concurrent variants have received very little attention so far. In this talk, I describe the first non-blocking implementation of the classic interpolation search tree data structure. For arbitrary key distributions, the data structure ensures amortized O(log n) insertion and deletion. Furthermore, when input key distributions are smooth, lookups run in expected O(log log n) time, and insertion and deletion run in amortized O(log log n). I then hypothesize that the design of this data structure can influence the design of distributed search data structures, and achieve similar performance benefits.

In the second part of the talk, I describe how we implemented GraalWasm – an engine for the WebAssembly language by extending GraalVM. I start by describing the GraalVM stack, and the WebAssembly language, and I then describe the internals of GraalWasm. The talk is meant as an inspiration for people who want to use GraalVM for rapid prototyping of programming language implementations when high performance is required – my hope is that this talk should be of a particular interest for people working on query languages for databases or distributed systems.

3.20 Invariant-preserving applications for weakly consistent replicated databases

Carla Ferreira (Universidade Nova de Lisboa – Lisbon, PT, carla.ferreira@fct.unl.pt)

Building trustworthy cloud applications is inherently complex and error-prone, and requires developers with a high level of expertise. In this talk, I discuss sound analyses techniques that leverage recent theoretical advances to avoid altogether coordinating the execution of operations. The approach consists of modifying operations in a way that application invariants are ensured to be always maintained. When no conflicting updates occur, the modified operations present their original semantics. Otherwise, it uses sensible and deterministic conflict resolution policies that preserve the invariants of the application.

3.21 The global object tracker (GoT)

Rohan Achar (University of California – Irvine, US)

Object state synchronization between components of distributed applications containing several collaborating or competing components with highly mutable, long-lived, and replicated state is a challenging research area. As an organizing principle for such replicated objects, we propose the Global Object Tracker (GoT) model, an object-oriented programming model whose design and interfaces mirror those found in decentralized version control systems: a version graph, working data, diffs, commit, checkout, fetch, push, and merge. We have implemented GoT in a framework called Spacetime, written in Python. The advantages offered by GoT is the communication of expressive state updates that have low latency of propagation, and observability and thus, reasoning, over all the state changes that happen in the application.

3.22 Data programming for ML and Data Science – Challenges for data management, compilers, and distributed systems

Volker Markl (TU Berlin, DE)

Over the past decade, data management has steadily grown in complexity, with scientific institutions and enterprises building novel analytics that draw on theory and best practices in relational database management, graph analysis, machine learning, signal processing, statistical science, and mathematical programming. This heterogeneity of analytics problems has spurred the development of a diverse ecosystem of data analytics engines, each tailored to a specific paradigm and use case. Examples of such engines include relational database systems (e.g., Postgres, MySQL, MonetDB), tools for numerical analysis (e.g., Matlab, R, NumPy), emerging distributed data processing engines (e.g., Hadoop, Spark, Flink), distributed key-value stores (e.g., HBase, Cassandra), as well as specialized graph-processing systems (e.g., Neo4J, Giraph, GraphLab). Each of these engines has specific advantages and disadvantages; however, picking the right one – or the right combination – for a given problem can be a daunting task for a data analyst. In addition, we are observing an increase in the diversification of the hardware landscape, promising to improve data processing performance: (i) heterogeneous processors configurations that combine diverse architectures (e.g., CPUs, GPUs, vector processors, FPGAs), (ii) the availability of fast, high-capacity flash storage, (iii) the emergence of non-volatile memory technology disrupting the traditional memory hierarchy, and (iv) the continued evolution of network interconnects. Furthermore, the growing number of available hardware virtualization and "infrastructure-as-a-service" solutions implies that specially-tailored hardware configurations will now be readily available to basically anyone, at the click of a button. However, this increase in variety makes it far more difficult to identify the hardware configuration that exploits hardware properties optimally for a target problem. The growing heterogeneity at the model (e.g., matrices, tables, graphs) and language (Matlab, SQL, Java), the system, and the hardware level is making efficient data analysis increasingly formidable. Specifying and tuning data analysis programs (DAPs) requires analysts to manually find the optimal combination of programming models, runtime engines, and hardware configurations from a vast number of possible alternatives. Therefore, today's data analyst must be a "jack-of-all-trades," i.e., proficient in a multitude of different systems and languages, comprehend data and processing models, grasp the intricacies of tuning parameters and their corresponding performance impact, and able to map a given analysis task to the ideal combination of systems with the most effective hardware configuration. This rare combination of skills is one of the key reasons behind the severe shortage of capable data analysts. Reducing the complexity of the data analysis process, the entry barrier, and the cost of analyzing large amounts of data at scale is one of the most important goals in data management research today. The only reasonable way to reduce this complexity is to automate the manual design and implementation choices data scientists regularly face in this heterogeneous environment (e.g., identifying the algorithms, system components, and tuning parameters). Achieving this automation is the "holy grail" of data science, but it is possible, if we combine several data processing technologies established by the scientific/systems community, drawing from skills in programming languages, compiler technology, database systems and distributed systems. To this end, we envision to a principled algebraic model for scalable data science systems, akin to relational algebra in database systems, that will enable precisely define, study, and solve in a consistent way issues pertaining to the automatic optimization, distribution, parallelization, and hardware adaptation of entire data analysis pipelines. To do that, we need to carry the concepts of declarative languages, optimizing transformation rules and query optimizers and concepts over to the world of data science with DAPs beyond relational algebra.

Participants

- Rohan Achar
University of California – Irvine, US
- Carlos Baquero
University of Minho, PT
- Annette Bieniusa
TU Kaiserslautern, DE
- Uwe Breitenbücher
Universität Stuttgart, DE
- Sebastian Burckhardt
Microsoft Research – Redmond, US
- Surajit Chaudhuri
Microsoft Research – Redmond, US
- Natalia Chechina
University of Bournemouth-Poole, GB
- Amit K. Chopra
Lancaster University, GB
- Schahram Dustdar
Technische Universität Wien, AT
- Patrick Thomas Eugster
University of Lugano, CH
- Carla Ferreira
New University of Lisbon, PT
- Torsten Grust
Universität Tübingen, DE
- Philipp Haller
KTH Royal Institute of Technology – Stockholm, SE
- Edward A. Lee
University of California – Berkeley, US
- Heather Miller
Carnegie Mellon University – Pittsburgh, US
- Aleksandar Prokopec
Oracle Labs Switzerland – Zürich, CH
- Laurent Prosperi
Sorbonne University – Paris, FR
- Guido Salvaneschi
TU Darmstadt, DE
- Manuel Serrano
INRIA – Valbonne, FR
- Marc Shapiro
Sorbonne University – Paris, FR
- Marjan Sirjani
Mälardalen University – Västerås, SE
- Peter Van Roy
UC Louvain, BE
- Nobuko Yoshida
Imperial College London, GB
- Damien Zufferey
MPI-SWS – Kaiserslautern, DE

Report from Dagstuhl Seminar 19443

Algorithms and Complexity in Phylogenetics

Edited by

Magnus Bordewich[1], Britta Dorn[2], Simone Linz[3], and Rolf Niedermeier[4]

1 **Durham University, GB,** m.j.r.bordewich@durham.ac.uk
2 **Universität Tübingen, DE,** britta.dorn@uni-tuebingen.de
3 **University of Auckland, NZ,** s.linz@auckland.ac.nz
4 **TU Berlin, DE,** rolf.niedermeier@tu-berlin.de

Abstract

Phylogenetics is the study of ancestral relationships between species. Its central goal is the reconstruction and analysis of phylogenetic trees and networks. Even though research in phylogenetics is motivated by biological questions and applications, it heavily relies on mathematics and computer science. Dagstuhl Seminar 19443 on *Algorithms and Complexity in Phylogenetics* aimed at bringing together researchers from phylogenetics and theoretical computer science to enable an exchange of expertise, facilitate interactions across both research areas, and establish new collaborations. This report documents the program and outcomes of the seminar. It contains an executive summary, abstracts of talks, short summaries of working groups, and a list of open problems that were posed during the seminar.

Seminar October 27–31, 2019 – http://www.dagstuhl.de/19443

2012 ACM Subject Classification Theory of computation → Parameterized complexity and exact algorithms, Mathematics of computing → Graph algorithms, Applied computing → Molecular evolution

Keywords and phrases Approximation algorithms, Evolution, Parameterized algorithms, Phylogenetic trees and networks

Digital Object Identifier 10.4230/DagRep.9.10.134

Edited in cooperation with Kristina Wicke, Universität Greifswald, DE

1 Executive Summary

Magnus Bordewich (Durham University, GB)
Britta Dorn (Universität Tübingen, DE)
Simone Linz (University of Auckland, NZ)
Rolf Niedermeier (TU Berlin, DE)

Disentangling the evolutionary relationships between species dates back at least to Charles Darwin and his voyage on board the Beagle. Ever since, the research area of phylogenetics focusses on the reconstruction and analysis of rooted leaf-labeled trees, called phylogenetic (evolutionary) trees, to unravel ancestral relationships between entities like species, languages, and viruses. However, processes such as horizontal gene transfer and hybridization challenge the model of a phylogenetic tree since they result in mosaic patterns of relationships that cannot be represented by a single tree. Indeed, it is now widely acknowledged that rooted leaf-labeled digraphs with underlying cycles, called phylogenetic networks, are better suited to represent evolutionary histories.

Biological questions and applications motivate much of the research in phylogenetics. Nevertheless, most of the software that is routinely used by evolutionary biologists has its roots in theoretical research areas which include algorithms, computational complexity, graph theory, algebra, and probability theory. With a shift from phylogenetic trees towards more complex graphs, the development of new algorithms for phylogenetic networks is currently an active area of research that requires deep insight from computer science and mathematics.

The objective of the seminar was to facilitate interactions between the two research communities of (i) computational and mathematical phylogenetics and (ii) theoretical computer science with a focus on algorithms and complexity. Specifically, its goal was to advance the development of novel algorithms (with provable performance guarantee) to reconstruct and analyze phylogenetic networks that are grounded in techniques from theoretical computer science such as parameterized and approximation algorithms.

This four-day seminar brought together 27 researchers from ten countries, whose research spans theoretical computer science and algorithms, (discrete) mathematics, and computational and mathematical phylogenetics. The seminar program included six overview talks, nine research talks (one of which via Skype), a rump session for short five-minute contributions, and slots for discussions and group work on open problems. More specifically, the overview talks provided introductions to techniques and current trends in parameterized algorithms, combinatorial decompositions, and enumeration algorithms on one hand, and introductions to spaces of phylogenetic trees and networks, and the reconstruction of networks from smaller networks and trees on the other hand. Additionally, each overview talk included open questions and challenges that provided a foundation for discussions and group work throughout the week. The research talks, of which three were given by postgraduate students, covered topical streams of research, including phylogenetic split theory, the placement of phylogenetic problems in higher classes of the polynomial hierarchy, new insight into the popular so-called Tree Containment problem, and phylogenetic diversity and biodiversity indices. Moreover, five working groups were formed on the second day of the seminar. While the research projects that were initiated in these groups are ongoing, some groups obtained first results during the seminar that were presented on the last day.

By building on initially existing synergies between the two research communities, the seminar has taken a leap towards developing new and fostering existing collaborations between both communities. Collaborative work was encouraged and put into practice over formal and informal discussions as well as three group work sessions. Since a significant number of open problems in phylogenetics require the combined expertise of experts in phylogenetics and theoretical computer science, we expect the collaborations formed at Schloss Dagstuhl to make progress on problems across the traditional discipline boundaries and, ideally, lead to joint peer-reviewed journal or conference publications.

To conclude, this seminar has acknowledged that exchange and connection between the two research communities of theoretical computer science and phylogenetics is fruitful for both sides. Techniques and methods from algorithms and complexity as well as theoretical considerations in general enable, account for, and foster new insights in problems from phylogenetics. Conversely, the specific features and problem structures appearing in the context of phylogenetic trees and networks provide novel theoretical challenges and new directions for foundational research in algorithms and computational complexity.

We thank all participants for their contributions and for openly sharing their ideas and research questions that led to a positive working atmosphere and many discussions throughout the seminar. Furthermore, we sincerely thank the team of Schloss Dagstuhl for their excellent support and communication as well as for providing an enjoyable seminar environment.

2 Table of Contents

Executive Summary
Magnus Bordewich, Britta Dorn, Simone Linz, and Rolf Niedermeier 134

Overview of Talks

Reconstructing equidistant phylogenetic networks from distance matrices
Allan Bai 138

Deciding whether two phylogenetic networks embed the same trees is hard
Janosch Döcker 138

Orthology relations
Katharina T. Huber 139

New kernels for TBR distance (or, equivalently, the maximum agreement forest problem): theory and practice
Steven Kelk 139

An introduction to fixed-parameter algorithms: basic techniques and recent ideas
Christian Komusiewicz and André Nichterlein 140

Some open problems in phylogenetic split theory
Vincent Moulton 140

Continuous spaces of phylogenetic trees and networks
Megan Owen 141

Algorithmic tree and network problems in phylogenetics
Charles Semple 141

Searching for optimal phylogenetic histories
Katherine St. John 142

Applying SNAQ with 5-net
Nihan Tokaç 142

Combinatorial decompositions and enumeration algorithms
Alexandru Tomescu 143

Constructing phylogenetic networks from smaller networks
Leo van Iersel 143

Tree Containment with polytomies
Mathias Weller 144

Phylogenetic diversity and biodiversity indices on phylogenetic networks
Kristina Wicke 144

Global-Local Clustering
Norbert Zeh 145

Working Groups

Longest Common Subsequence for similar strings
Laurent Bulteau, Mark E. L. Jones, Rolf Niedermeier, and Till Tantau 146

Hybridization for many trees with non-identical leaf sets
Britta Dorn, Christian Komusiewicz, Catherine McCartin, André Nichterlein, Mathias Weller, and Norbert Zeh 147

Hybridization Number for multiple multifurcating trees
Mark E. L. Jones and Vincent Moulton 148

Maximum agreement subtrees
Katherine St. John, Magnus Bordewich, Simone Linz, Megan Owen, Charles Semple, and Kristina Wicke 149

Constructing phylogenetic networks from trinets
Leo van Iersel, Vincent Moulton, Leen Stougie, and Nihan Tokac 149

Open Problems

List of open problems
Leo van Iersel 150

Participants 151

3 Overview of Talks

3.1 Reconstructing equidistant phylogenetic networks from distance matrices

Allan Bai (University of Canterbury – Christchurch, NZ)

Main reference Magnus Bordewich, Katharina T. Huber, Vincent Moulton, Charles Semple: "Recovering normal networks from shortest inter-taxa distance information", Journal of Mathematical Biology, Vol. 77(3), pp. 571–594, 2018.
URL http://dx.doi.org/10.1007/s00285-018-1218-x

A phylogenetic network is a rooted acyclic directed graph, where the set of all vertices with out-degree zero are called leaves. A phylogenetic network is equidistant if all paths from the root to the leaves are equal. It has been shown in the past that certain classes of phylogenetic networks are determined by the distances between the leaves, up to a certain equivalence. In this talk, I will give an overview of the known algorithms for reconstructing normal and tree-child networks. I will also present my research on reconstructing shortcut free networks using minimum distance matrices.

3.2 Deciding whether two phylogenetic networks embed the same trees is hard

Janosch Döcker (Universität Tübingen, DE)

Joint work of Janosch Döcker, Simone Linz, Charles Semple
Main reference Janosch Döcker, Simone Linz, Charles Semple: "Displaying trees across two phylogenetic networks", Theor. Comput. Sci., Vol. 796, pp. 129–146, 2019.
URL https://doi.org/10.1016/j.tcs.2019.09.003

Phylogenetic networks are frequently used to represent the ancestral relationships between a collection of extant species. Noting that each phylogenetic network N embeds a collection of phylogenetic trees, we refer to this collection as the display set of N. A well-studied and biologically relevant problem asks, given a phylogenetic network N and a phylogenetic tree T, whether T is contained in the display set of N. We study the computational complexity of several questions related to the display sets of two given phylogenetic networks. In particular, we show that deciding whether two phylogenetic networks have the same display set is computationally hard and, more specifically, that it can be placed on the second level of the polynomial-time hierarchy.

3.3 Orthology relations

Katharina T. Huber (University of East Anglia – Norwich, GB)

Joint work of Katharina T. Huber, Guillaume Scholz
Main reference Katharina T. Huber, Guillaume E. Scholz: "Beyond Representing Orthology Relations by Trees", Algorithmica, Vol. 80(1), pp. 73–103, 2018.
URL https://doi.org/10.1007/s00453-016-0241-9

Reconstructing the evolutionary past of a family of genes is an important aspect of many genomic studies. To help with this, simple relations on a set of sequences called orthology relations may be employed [1]. In addition to being interesting from a practical point of view, they are also attractive from a theoretical perspective in that e.g. a characterization is known for when such a relation is representable by a certain type of phylogenetic tree [2]. Perhaps not surprisingly, real biological data however hardly ever satisfies that characterization. Starting with a brief introduction into the area, we review some recent results concerning such relations [3].

References

1 M. Hellmuth, N. Wieseke, M. Lechner, H-P. Lenhof, M. Middendorf, and P.F. Stadler. *Phylogenomics with paralogs* PNAS, 112:2058-2063, 2015.
2 M. Hellmuth, M. Hernandez-Rosales, K.T. Huber, V. Moulton, P.F. Stadler, and N. Wieseke. *Orthology relations, symbolic ultrametrics and cographs.* Journal of Mathematical Biology, 66:39-420, 2013.
3 K.T. Huber, and E.G. Scholz. *Beyond representing orthology relations by trees.* Algorithmica, 80:73-103, 2018.

3.4 New kernels for TBR distance (or, equivalently, the maximum agreement forest problem): theory and practice

Steven Kelk

Joint work of Steven Kelk, Simone Linz, Rim Van Wersch
Main reference Steven Kelk, Simone Linz: "A Tight Kernel for Computing the Tree Bisection and Reconnection Distance between Two Phylogenetic Trees", SIAM J. Discrete Math., Vol. 33(3), pp. 1556–1574, 2019.
URL http://dx.doi.org/10.1137/18M122724X

Given two phylogenetic (i.e. evolutionary) trees, the TBR-distance between them is the minimum number of "Tree Bisection and Reconnection" topological rearrangement moves required to transform one tree into another. This problem is NP-hard but was shown to be FPT in 2001 by Allen and Steel [1], who showed that polynomial-time reduction rules can be applied to reduce instances to size $28k$, where k is the TBR distance. In this talk we show that the kernelization strategy proposed by Allen and Steel [1] actually reduces the trees to size $15k - 9$, and that this is tight. The sharpened analysis is made possible by exploiting the equivalence of the TBR-distance problem to the problem of embedding the two trees parsimoniously into an undirected graph. Combining this equivalence with an older equivalence (that of "maximum agreement forests") then yields a whole suite of new polynomial-time reduction rules which further shrink the trees to size $11k - 9$. We have also implemented the new reduction rules and describe briefly the results of preliminary experiments indicating that the new rules do, in practice, lead to further reductions in kernel size.

References

1 B. Allen, and M. Steel. *Subtree transfer operations and their induced metrics on evolutionary trees.* Annals of Combinatorics, 5:1-15, 2001.

3.5 An introduction to fixed-parameter algorithms: basic techniques and recent ideas

Christian Komusiewicz (Universität Marburg, DE) and André Nichterlein (TU Berlin, DE)

Main reference Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, Saket Saurabh: "Parameterized Algorithms", Springer, 2015.
URL https://doi.org/10.1007/978-3-319-21275-3
Main reference Fedor V. Fomin, Daniel Lokshtanov, Saket Saurabh, Meirav Zehavi: "Kernelization: Theory of Parameterized Preprocessing", Cambridge University Press, 2019.
URL https://doi.org/10.1017/9781107415157

We first review two algorithmic techniques for developing fixed-parameter algorithms: search tree algorithms and kernelization. Then we describe methods for showing fixed-parameter intractability. Finally, we discuss three more recent issues in fixed-parameter algorithms:
1. identification of good parameters via parameter hierarchies,
2. FPT-approximation, and
3. parameterized local search.

3.6 Some open problems in phylogenetic split theory

Vincent Moulton (University of East Anglia – Norwich, GB)

Split networks are one of the most commonly used type of phylogenetic network [2, 3]. Underlying these networks are combinatorial structures called split systems [1], which have a rich associated mathematical and computational theory. In this talk, we will give a brief introduction to split systems and split networks, and present some open problems related to these structures. These problems include:
1. What is the complexity of deciding whether or not a split system is flat [4]?
2. Is there an efficient algorithm to compute phylogenetic diversity for weakly compatible split systems [7]?
3. Does the 1-skeleton of the tight-span of a totally-split decomposable metric contain an optimal realization for the metric [5]?
4. Is a maximum linearly independent split system orderly if and only if it is circular [6]?

References

1 H.-J. Bandelt, A. Dress. *A canonical decomposition theory for metrics on a finite set.* Advances in Mathematics, 92:47-105, 1992.

2 D. Bryant, V. Moulton. *Neighbor-net: an agglomerative method for the construction of phylogenetic networks.* Molecular Biology and Evolution, 21:255-65, 2004.

3 D. Huson, D. Bryant. *Application of phylogenetic networks in evolutionary studies.* Molecular Biology and Evolution, 23:254-67, 2005.

4 M. Balvocute, A. Spillner, V. Moulton. *FlatNJ: A novel network-based approach to visualize evolutionary and biogeographical relationships*. Systematic Biology, 63:383-96, 2014.
5 S. Herrmann, J. Koolen, A. Lesser, V. Moulton, T. Wu. *Optimal realizations of two-dimensional, totally-decomposable metrics*. Discrete Mathematics, 338:1289-99, 2015.
6 V. Moulton, A. Spillner. *Order distance and split systems*. https://arxiv.org/abs/1910.10119
7 A. Spillner, B. Nguyen, V. Moulton. *Computing phylogenetic diversity for split systems*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 5:235-44, 2008.

3.7 Continuous spaces of phylogenetic trees and networks

Megan Owen (Lehman College – New York, US)

Joint work of Daniel G. Brown, Gillian Grindstaff, Megan Owen
Main reference Daniel G Brown, Megan Owen: "Mean and Variance of Phylogenetic Trees", Systematic Biology, Vol. 69(1), pp. 139–154, 2019.
URL https://doi.org/10.1093/sysbio/syz041
Main reference Gillian Grindstaff, Megan Owen: "Geometric comparison of phylogenetic trees with different leaf sets", CoRR, Vol. abs/1807.04235, 2018.
URL https://arxiv.org/abs/1807.04235

A metric phylogenetic tree is a phylogenetic tree with lengths on its edges. Since evolutionary processes such as the coalescent depend on tree edge lengths, it is important to have a framework for analyzing both the tree topology and edge lengths together. One such framework is a continuous geometric space of phylogenetic trees, which has the metric trees as its points, and which accounts for the intrinsic properties of the trees through the geometry of the space. The most well-known such space is the Billera-Holmes-Vogtmann (BHV) treespace, and I will describe it, algorithms on it, and how it can be used to analyze tree data. I will also describe several other continuous treespaces, including an approach to analyzing trees with over-lapping leaf sets, and discuss network spaces.

3.8 Algorithmic tree and network problems in phylogenetics

Charles Semple (University of Canterbury – Christchurch, NZ)

Joint work of Mihaela Baroni, Magnus Bordewich, Janosch Döcker, Stefan Grünewald, Peter Humphries, Simone Linz, Vincent Moulton, Charles Semple
Main reference Janosch Döcker, Simone Linz, Charles Semple: "Displaying trees across two phylogenetic networks", Theor. Comput. Sci., Vol. 796, pp. 129–146, 2019.
URL http://dx.doi.org/10.1016/j.tcs.2019.09.003
Main reference Peter J. Humphries, Simone Linz, Charles Semple: "Cherry Picking: A Characterization of the Temporal Hybridization Number for a Set of Phylogenies", Bulletin of Mathematical Biology, Vol. 75(10), pp. 1879–1890, 2013.
URL http://dx.doi.org/10.1007/s11538-013-9874-x

Phylogenetic networks are a particular type of rooted, acyclic digraph and are used in computational biology to represent the non-treelike evolutionary history of extant species. Non-treelike (reticulate) processes in evolution include lateral gene transfer and hybridisation. Although evolution is not necessarily treelike at the species-level, at the level of genes, we typically assume treelike evolution. Thus phylogenetic networks are often viewed as an amalgamation of gene trees (phylogenetic trees representing the evolutionary history of single genes). From this viewpoint, two of the most well-studied computational problems concerning phylogenetic networks is that of (i) determining the minimum number of reticulations for a network to embed a given set of conflicting trees and (ii) deciding whether or not a given network embeds a given tree. In this talk, I will present an overview of these problems and variations of them.

3.9 Searching for optimal phylogenetic histories

Katherine St. John (CUNY Hunter College – New York, US)

Main reference Katherine St. John: "Review Paper: The Shape of Phylogenetic Treespace", Systematic Biology, Vol. 66(1), pp. e83–e94, 2016.
URL https://doi.org/10.1093/sysbio/syw025

Evolutionary histories, or phylogenies, form an integral part of much work in biology. In addition to the intrinsic interest in the interrelationships between species, phylogenies are used for drug design, multiple sequence alignment, and even as evidence in a recent criminal trial. A simple representation for a phylogeny is a rooted, binary tree, where the leaves represent the species, and internal nodes represent their hypothetical ancestors. In this talk, we outline the optimality criteria used for evaluating phylogenetic trees and organizing the search space, the space of n-leaf trees. We classify the most popular metrics and the resulting treespaces. We examine the choice of metrics on the success of the search on finding the optimal trees, as well as the complexity of the algorithms, with emphasis on those problems that yield tractable or fixed parameter tractable algorithms.

3.10 Applying SNAQ with 5-net

Nihan Tokaç (Antalya International University, TR)

Joint work of Céline Scornavacca, Nihan Tokaç

Phylogenetic networks are a necessary tool to represent the evolutionary history of species including horizontal gene transfers, hybridizations or gene flow. In [1], Solís-Lemus and Ané have inferred phylogenetic networks in a pseudolikelihood framework. In this work, the pseudolikelihood of a network is based on the likelihood formulas of its 5-taxon subnetworks instead of 4-taxon subnetworks

References

1 C. Solís-Lemus, C. Ané. *Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting*. PLoS Genetics, 12(3), e1005896, 2016.

3.11 Combinatorial decompositions and enumeration algorithms

Alexandru Tomescu (University of Helsinki, FI)

Main reference Alexandru I. Tomescu, Paul Medvedev: "Safe and Complete Contig Assembly Via Omnitigs", in Proc. of the Research in Computational Molecular Biology – 20th Annual Conference, RECOMB 2016, Santa Monica, CA, USA, April 17-21, 2016, Proceedings, Lecture Notes in Computer Science, Vol. 9649, pp. 152–163, Springer, 2016.
URL http://dx.doi.org/10.1007/978-3-319-31957-5_11

A combinatorial decomposition is a characterization of an object in terms of same objects, but of smaller size. We introduce the classical decomposition of labeled DAGs by sources and show how it can be used for counting, approximate counting and random generation. We then introduce some classical complexity measures of an enumeration algorithm, present best-k enumeration algorithms and introduce the more recent topic of safe and complete algorithms, which was introduced to formalize the genome assembly problem. A safe algorithm is one outputting only partial solutions that are common to all solutions to a problem. A particular variant of the notion of safety has been previously studied under the title of persistency.

3.12 Constructing phylogenetic networks from smaller networks

Leo van Iersel (TU Delft, NL)

Joint work of Leo van Iersel, Vincent Moulton, Katharina Huber, Taoyang Wu, Celine Scornavacca, James Oldman, Sjors Kole
Main reference Katharina T. Huber, Leo van Iersel, Vincent Moulton, Céline Scornavacca, Taoyang Wu: "Reconstructing Phylogenetic Level-1 Networks from Nondense Binet and Trinet Sets", Algorithmica, Vol. 77(1), pp. 173–200, 2017.
URL http://dx.doi.org/10.1007/s00453-015-0069-8

A common approach towards reconstructing large phylogenetic trees is to combine various phylogenetic trees on different, overlapping leaf label sets to a single phylogenetic tree on all leaf labels. A similar approach has been proposed for phylogenetic networks. In this case, the input consists of phylogenetic networks with different but overlapping leaf labels sets, and the goal is to construct a phylogenetic network that contains each of the input networks. Unfortunately, it has been shown that phylogenetic networks are in general not uniquely determined by their subnetworks. This contrasts the situation for phylogenetic trees, which are uniquely determined by their sets of contained 3-leaf trees, which are called triplets. Nevertheless, it has been shown that certain restricted classes of phylogenetic networks are uniquely determined by their sets of trinets, which are 3-leaf networks. For the severely restricted class of level-1 networks, it is even possible to reconstruct the network given all its trinets in polynomial time. However, for non-dense trinet sets, this problem is already NP-hard and the only existing algorithms are an efficient heuristic and an exponential-time exact algorithm. Open problems in this area include the questions whether there exist fixed-parameter algorithms and whether more general classes of networks are still uniquely determined by their trinets.

3.13 Tree Containment with polytomies

Mathias Weller University Paris-Est – Marne-la-Vallée, FR)

Joint work of Matthias Bentert, Mathias Weller
Main reference Mathias Weller: "Linear-Time Tree Containment in Phylogenetic Networks", in Proc. of the Comparative Genomics – 16th International Conference, RECOMB-CG 2018, Magog-Orford, QC, Canada, October 9-12, 2018, Proceedings, Lecture Notes in Computer Science, Vol. 11183, pp. 309–323, Springer, 2018.
URL http://dx.doi.org/10.1007/978-3-030-00834-5_18
Main reference Matthias Bentert, Josef Malík, Mathias Weller: "Tree Containment With Soft Polytomies", in Proc. of the 16th Scandinavian Symposium and Workshops on Algorithm Theory, SWAT 2018, June 18-20, 2018, Malmö, Sweden, LIPIcs, Vol. 101, pp. 9:1–9:14, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.
URL http://dx.doi.org/10.4230/LIPIcs.SWAT.2018.9

In this work, we consider the Tree Containment problem, asking whether a given rooted phylogentic tree is embeddable (leaf-label respecting topological minor) in a given phylogenetic network. We improve previously known results by presenting a linear-time algorithm for a broad class of networks, properly including the class of reticulation visible networks. We also show parameterized algorithms for a parameter that is stronger (that is, smaller in all instances) than the "level" of the network. All results work for so-called "hard polytomies" meaning high-degree nodes that represent large species fan outs. We further consider the more biologically relevant case of "soft polytomies", where high-degree nodes represent uncertainty in the tree (branches with low support are often contracted after construction of the phylogeny) and show algorithms and NP-hardness for some classes of networks.

3.14 Phylogenetic diversity and biodiversity indices on phylogenetic networks

Kristina Wicke (Universität Greifswald, DE)

Joint work of Mareike Fischer, Kristina Wicke
Main reference Kristina Wicke, Mareike Fischer: "Phylogenetic diversity and biodiversity indices on phylogenetic networks", Mathematical Biosciences, Vol. 298, pp. 80–90, 2018.
URL https://doi.org/10.1016/j.mbs.2018.02.005

Facing a major extinction crisis and the inevitable loss of biodiversity at the same time with limited financial means, it is often necessary to prioritize species for conservation. Existing approaches for prioritization, e.g. the Fair Proportion index, the Equal Splits index or the Shapley value, are based on phylogenetic trees and rank species according to their contribution to overall phylogenetic diversity (PD). However, in many cases evolution is not treelike and thus, phylogenetic networks have come to the fore as a generalization of phylogenetic trees, allowing for the representation of non-treelike evolutionary events, such as horizontal gene transfer or hybridization.

While phylogenetic diversity and PD indices have been studied in great detail for trees, research on how these concepts might be used in the context of phylogenetic networks is still in its infancy. In this talk, I will thus introduce phylogenetic diversity and PD indices for trees, before considering first attempts to extend these concepts from trees to networks. These attempts range from considering the treelike content of a network (e.g. the (multi)set of trees displayed by a network or the lowest stable ancestor tree associated with it) to directly taking into account the network structure. In this talk, I will discuss some of these approaches, analyze their advantages and drawbacks and indicate some directions for future research.

3.15 Global-Local Clustering

Norbert Zeh (Dalhousie University – Halifax, CA)

Joint work of Mark E. L. Jones, Remie Janssen, Yuki Murakami, Leo van Iersel, Norbert Zeh

Given two phylogenetic trees over the same leaf set, cluster reduction identifies a pair of internal (i.e., non-root, non-leaf) nodes, one from each tree, so that both nodes have the same set of descendant leaves and then splits the input into two pairs of subtrees. Roughly, the first pair is the pair of descendant trees of the two chosen nodes; the second pair is the pair of trees obtained by giving the two chosen nodes a new label and removing their proper descendants. By applying this partition repeatedly, one obtains a partition of the two inputs trees into a collection of tree pairs that are hopefully much smaller. This partition is useful because Baroni, Semple, and Steel (2006) and Linz and Semple (2011) have shown that a maximum agreement forest (MAF; equivalent to the subtree prune-and-regraft distance) and a maximum acyclic agreement forest (MAAF; equivalent to an optimal hybridization network) of two trees can be computed from MAFs or MAAFs of the tree pairs in this partition. This has been verified to be incredibly effective for speeding up the computation of such agreement forests in practice (e.g., see Li and Zeh, 2017). However, some inputs do not decompose into small clusters because there may not be any pair of internal nodes with the same set of descendant leaves. For such inputs, cluster reduction is completely ineffective.

With the goal of extending the applicability of cluster reduction to a wider range of inputs, we introduce the notion of global-local clustering. An agreement forest (AF) is a forest that can be obtained from both input trees by cutting an appropriate subset of edges. A maximum agreement forest (MAF) is an agreement forest that can be obtained by cutting as few edges as possible. Such a forest can be found in $O(2^k * poly(n))$ time, where k is the number of edges cut and n is the number of leaves in the input. A (g, l, k)-clustering is a pair of edge sets, one per input tree, each of size at most k and such that cutting these edges produces an AF, along with a partition of each set into g"global" edges and the remaining"local edges"; this partition must have the property that each cluster in the cluster partition of the two forests obtained by cutting the global edges contains at most l local edges. We call a (g, l, k)-clustering an optimal (g, l)-clustering if k is the number of edges that need to be cut to obtain a MAF. Given the set of global edges, the local edges in all clusters can be found in $O(2^l * poly(n))$ time. The question is whether we can find the global edges in $O(f(g, l) * poly(n))$ time, that is, in time independent of k. This would allow us to find MAFs and MAAFs quickly even if k is large because g and l may be much smaller than k. This is true even if the input does not decompose into small clusters without cutting global edges first.

We give two positive and one negative answer to variations of this question. We show that given a fixed pair (g, l), we can decide in $O(f(g, l) * poly(n))$ time whether a given input has a (g, l, k)-clustering for some k and if so, find the smallest k for which such a (g, l, k)-clustering exists. The standard approach for finding a MAF is to ask whether k edge cuts suffice to obtain an AF for increasing values of k until one obtains an affirmative answer. We show that using this approach, we cannot decide in $O(f(g, l) * poly(n))$ time whether there exists an optimal (g, l)-clustering for a given input, essentially because we do not know whether the clustering we have obtained for a given pair (g, l) is optimal until we have tried all clusterings up to $g = k$ or $l = k$; larger values of g or l may allow us to cut fewer edges overall. On the positive side, we show that if there exists a (g, l)-clustering, even a suboptimal one, then the display graph of the two input trees has a tree width that is a function of only g and l. Thus, if the goal is simply to find a MAF in $O(f(g, l) * poly(n))$ time, we can construct the display graph along with a tree decomposition of this graph, and then employ the monadic second order logic framework of Kelk et al. (2016) to obtain a MAF in $O(f'(treewidth) * poly(n)) = O(f(g, l) * poly(n))$ time.

The main open problem is to determine whether there exist practical algorithms for finding (g, l, k)-clusterings. Neither our branching algorithm for a fixed pair (g, l) nor the tree decomposition-based algorithm of Kelk et al. is practical.

References

1 M. Baroni, C. Semple, and M. Steel. *Hybrids in real time*. Systematic Biology, 55:46-56, 2006.
2 S. Linz and C. Semple. *A cluster reduction for computing the subtree distance between phylogenies*. Annals of Combinatorics, 15:465-484, 2011.
3 Z. Li and N. Zeh. *Computing maximum agreement forests without cluster partitioning is folly*. In Proceedings of the European Symposium on Algorithms, pages 56:1–56:14, 2017.
4 S. Kelk, L. van Iersel, C. Scornavacca, and M. Weller. *Phylogenetic incongruence through the lens of monadic second order logic*. Journal of Graph Algorithms and Applications, 20:189–215, 2016.

4 Working Groups

4.1 Longest Common Subsequence for similar strings

Laurent Bulteau (University Paris-Est – Marne-la-Vallée, FR), Mark E. L. Jones (CWI – Amsterdam, NL), Rolf Niedermeier (TU Berlin, DE), and Till Tantau (Universität zu Lübeck, DE)

The longest common subsequence problem (LCS) is a core string comparison problem, where one seeks a common pattern appearing in a (possibly large) set of (possibly long) strings. It has been well-studied in a large number of settings, in particular regarding parameterized algorithms. However, the complexity remains open for one of the most natural parameterizations : the maximum number of deletions in any input string. Similar questions are open for related problems: Shortest Common Supersequence (where the pattern is obtained by inserting characters rather than deletions), Center String and Median String (allowing all edit operations). The closest related results are an FPT algorithm for Closest String parameterized by the distance (allowing substitutions only) and FPT algorithm for LCS parameterized by the distance plus number of strings.

Using the concept of maximal common subsequence, we have developed a fixed-parameter algorithm solving LCS, with only the distance as parameter (using only linear time on the size of the input), thus answering our main open question. We expect that a similar method should apply to Shortest Common Supersequence. However, a different approach may be required for Center/Median string problems.

4.2 Hybridization for many trees with non-identical leaf sets

Britta Dorn (Universität Tübingen, DE), Christian Komusiewicz (Universität Marburg, DE), Catherine McCartin (Massey University, NZ), André Nichterlein (TU Berlin, DE), Mathias Weller (University Paris-Est – Marne-la-Vallée, FR), and Norbert Zeh (Dalhousie University – Halifax, CA)

A phylogenetic network is a directed acyclic graph with a single source and whose sinks are labeled bijectively with the elements of some label set. A phylogenetic tree is a phylogenetic network that is in fact a rooted tree. A network is said to display a tree if the tree can be obtained from the network by deleting a subset of its vertices and edges and suppressing nodes of in-degree 1 and out-degree 1. A hybridization network for a set of trees is a network that displays all trees in the set. It is an optimal hybridization network if it has the minimum (undirected) cyclomatic number among all hybridization networks for the same set of trees.

The parameterized complexity of constructing an optimal hybridization network for a pair of input trees is fairly well understood. For multiple input trees, the story is more complicated. There is no known practical parameterized algorithm for constructing an optimal hybridization network for more than two trees. The problem is known to be fixed-parameter tractable because there exists a quadratic kernel [1]. For specialized network construction problems such as finding tree-child networks or temporal networks, there exist practically efficient branching algorithms (cf. [2, 3]). Even these algorithms are limited in their usefulness to practitioners because of their assumption that all input trees share the same leaf set. This is rarely the case for real-world inputs obtained by constructing phylogenetic trees from genes shared by (subsets of) a set of species. Thus, we would like to construct optimal hybridization networks for multiple input trees with overlapping but non-identical leaf sets. This type of problem has received little attention so far because non-identical leaf sets pose no challenge whatsoever for pairs of trees. For more than two input trees, the problem becomes significantly harder. It is not fixed-parameter tractable when parameterized only by the hybridization number because deciding whether a given set of triplets (trees with 3 leaves) has a network that displays them and has hybridization number at most 2 is NP-hard [5].

Two somewhat natural parameterizations involve a pair of parameters. The first one considers the hybridization number k and the number of leaves l that are missing from at least one input tree. The second one considers the hybridization number and the number of input trees. Note that in the triplet example above, both the number of leaves absent from at least one input tree and the number of trees are large (at least linear in n). Our working group proved that the hybridization number problem is fixed-parameter tractable in both of these parameterizations. For the parameterization by hybridization number and missing leaves, we prove that there exists a kernel of size $O(k^2 + l)$ by extending the existence proof of a size-$O(k^2)$ kernel for the case of identical leaf sets. For the parameterization by hybridization number and number of trees, we prove that the display graph of the input trees has a tree width that is a function of k and t. Thus, the monadic second order logic framework of [4] can be used to find an optimal hybridization network for these trees in O(tree width) time and thus in $O(f(k, t))$ time. While these results do not lead to practical algorithms for constructing hybridization networks on sets of trees with non-identical leaf sets, they shed light on the computational complexity of this problem.

References

1 L. van Iersel, S. Kelk, and C. Scornavacca. *Kernelization for the hybridization number problem on multiple nonbinary trees*. Journal of Computer and System Sciences, 82:1075–1089, 2016.

2 L. van Iersel, R. Janssen, M. Jones, Y. Murakami, and N. Zeh. *A practical fixed-parameter algorithm for constructing tree-child networks from multiple binary trees*. aXiv:1907.08474 [cs.DM], 2019.

3 S. Borst. Personal communication, 2019.

4 S. Kelk, L. van Iersel, C. Scornavacca, and M. Weller. *Phylogenetic incongruence through the lens of monadic second order logic*. Journal of Graph Algorithms and Applications, 20:189–215, 2016.

5 J. Jansson, N. B. Nguyen, and W.-K. Sung. *Algorithms for combining rooted triplets into a galled phylogenetic network*. SIAM Journal on Computing, 35:1098–1121, 2006.

4.3 Hybridization Number for multiple multifurcating trees

Mark E. L. Jones (CWI – Amsterdam, NL) and Vincent Moulton (University of East Anglia – Norwich, GB)

In the `Hybridization Number` problem, we are given a set of rooted phylogenetic trees on a set of taxa X, and our aim is find a phylogenetic network that is as simple as possible while displaying all of those trees. ("As simple as possible" means having minimum reticulation number; that is, minimum number of edges that must be removed to turn the network into a tree). This problem is NP-hard and APX-hard, even when the input consists of two binary trees.

For two binary trees, the problem is fixed-parameter tractable with respect to the reticulation number. This result has been extended to non-binary trees and to instances with more than one input tree. In fact, FPT algorithms are known for `Hybridization Number` when either the number of trees or the maximum outdegree of any tree is bounded. However, the general case when there is an unbounded number of input trees with unbounded outdegree remains open.

Our group began attempted to find a general FPT algorithm for `Hybridization Number` on arbitrary number of trees with unbounded outdegree. Previous results have made use of the notion of a "generator" for a network which characterizes the overall structure in terms of the location of the reticulation nodes. A key observation is that a network with reticulation number k has $O(k)$ "sides" that limit the possible structure of trees displayed by this network. Existing techniques make use of this structure to show the correctness of some flavor of chain reduction rule (in which caterpillar-like substructure that are common to all trees can be reduced by deleting some taxa). This in combination with the standard subtree reduction rule has led to kernels with $O(dk^2)$ taxa (in the case of trees with maximum outdegree d) or $O(k(5k)^t)$ taxa (in the case of t trees with unbounded outdegree).

Ideally, we would be able to prove the correctness of a similar chain reduction rules for multiple large-outdegree trees (with the length of the chain likely increased), which would be enough to imply a kernel in a similar way. So far we have had no success in proving the correctness of such a rule. Nevertheless we believe it may be possible to exploit the generator structure of a k-reticulation network to enable more fine-grained reduction rules.

4.4 Maximum agreement subtrees

Katherine St. John (CUNY Hunter College – New York, US), Magnus Bordewich (Durham University, GB), Simone Linz (University of Auckland, NZ), Megan Owen (Lehman College – New York, US), Charles Semple (University of Canterbury – Christchurch, NZ), and Kristina Wicke (Universität Greifswald, DE)

Our working group focused on a conjecture of Martin and Thatte on lower bounds on the maximum agreement subtree distance between two trees. Steel and Székeley [3] shows that for any two trees on n leave, the agreement subtree is of size $\Omega(\log(\log n))$. Martin and Thatte [1] improves the lower bound to $\Omega(\sqrt{\log n})$ and conjectured that a lower bound of $\Omega(\sqrt{n})$ if both trees are balanced. There has been related work on expected distance between trees on the same shape is $\Omega(\sqrt{n})$ [2], suggesting that the conjecture should hold.

References

1. D. M. Martin, and B. D. Thatte. *The maximum agreement subtree problem*. Discrete Applied Mathematics 161:13-14, 1805–1817, 2013.
2. P. Misra, and S. Sullivant. *Bounds on the expected size of the maximum agreement subtree for a given tree shape*. SIAM Journal on Discrete Mathematics 33:2316–2325, 2019.
3. M. Steel, and L. A. Székely. *An improved bound on the maximum agreement subtree problem*. Applied Mathematics Letters 22:1778–1780, 2009.

4.5 Constructing phylogenetic networks from trinets

Leo van Iersel (TU Delft, NL), Vincent Moulton (University of East Anglia – Norwich, GB), Leen Stougie (CWI – Amsterdam, NL), and Nihan Tokaç (Antalya International University, TR)

Main reference Leo van Iersel, Vincent Moulton: "Trinets encode tree-child and level-2 phylogenetic networks", Journal of Mathematical Biology, Vol. 68(7), pp. 1707–1729, 2014.
URL http://dx.doi.org/10.1007/s00285-013-0683-5

There are many interesting problems related to constructing phylogenetic networks from subnetworks and in particular from trinets, which are 3-leaf subnetworks. While it has been shown that binary level-2 and tree-child networks are encoded (uniquely determined) by their trinets, there currently does not exist a polynomial-time algorithm for reconstructing a binary tree-child network from its trinets. Moreover, it is not clear whether binary level-k networks are encoded by their trinets for $3 \leq k \leq 11$. A counter example is known only for level-12.

In this working group, we have first focused on the question whether there exists a polynomial-time algorithm for reconstructing a binary temporal network from its trinets. The class of temporal networks form a subclass of the tree-child networks. A network is *tree-child* if every non-leaf vertex has a child that is not a reticulation. A network is *temporal* if it is tree-child and its vertices can be labeled by integers such that the label value remains unchanged along reticulation arcs and strictly increases along tree-arcs. We have a sketch of a polynomial-time algorithm for the temporal case.

Secondly, we considered the same question for the class of tree-child networks. For this case, it seems that there also exists a polynomial-time algorithm although with a worse running time than in the temporal case.

Finally, we considered the question whether level-3 networks are encoded by their trinets. Unfortunately, it seems that the techniques used for level-1 and level-2 do not generalize to level-3. Hence, we discussed different ways to try to prove the result for level-3.

5 Open Problems

5.1 List of open problems

Leo van Iersel (TU Delft, NL)

1. Are level-k networks, with $3 \leq k \leq 11$, uniquely determined by their trinets (3-leaf subnetworks)? And can they be reconstructed from their trinets in polynomial time?
2. Can tree-child networks be reconstructed from their trinets in polynomial time?
3. Are there FPT algorithms for constructing level-1 networks from non-dense trinet sets, with any reasonable parameter?
4. Is constructing a level-k network displaying a given dense set of triplets FPT with k as parameter?
5. Is constructing a temporal network displaying a given set of non-binary trees FPT with the number of reticulations in the network (and possibly the maximum outdegree of the input trees) as parameters?
6. Does there exist a polynomial-time algorithm for deciding whether there exists a binary tree-child network displaying a given set of (at least three) binary trees?
7. Does there exist a polynomial-time algorithm for deciding whether a given unrooted binary network can be oriented to become a rooted tree-child network? And to become a rooted stack-free network?
8. Does there exist an FPT algorithm for constructing a network with reticulation number at most k displaying a given set of non-binary trees, when the only parameter is k?
9. Does there exist an EPT algorithm (i.e. an FPT algorithm with running time $c^k poly(n)$) for constructing a network with reticulation number at most k displaying a given set of 4 binary trees?
10. Does there exist an FPT algorithm for constructing a network with reticulation number at most k displaying a given set of binary trees with non-equal leaf-sets, when the two parameters are k and the number of leaves that do not appear in all trees? (Note that all previous problems assume equal leaf-sets.)
11. For $4 \leq r \leq 7$, does there exist a constant $f(r)$ such that any set of r-state characters is compatible if and only if every size-$f(r)$ subset is compatible?

Participants

- Allan Bai
University of Canterbury – Christchurch, NZ
- Magnus Bordewich
Durham University, GB
- Laurent Bulteau
University Paris-Est – Marne-la-Vallée, FR
- Janosch Döcker
Universität Tübingen, DE
- Britta Dorn
Universität Tübingen, DE
- Anne-Sophie Himmel
TU Berlin, DE
- Katharina T. Huber
University of East Anglia – Norwich, GB
- Mark E. L. Jones
CWI – Amsterdam, NL
- Steven Kelk
Maastricht University, NL
- Christian Komusiewicz
Universität Marburg, DE
- Simone Linz
University of Auckland, NZ
- Marefatollah Mansouri
TU Wien, AT
- Catherine McCartin
Massey University, NZ
- Vincent Moulton
University of East Anglia – Norwich, GB
- André Nichterlein
TU Berlin, DE
- Rolf Niedermeier
TU Berlin, DE
- Megan Owen
Lehman College – New York, US
- Charles Semple
University of Canterbury – Christchurch, NZ
- Katherine St. John
CUNY Hunter College – New York, US
- Leen Stougie
CWIAmsterdam, NL
- Till Tantau
Universität zu Lübeck, DE
- Nihan Tokaç
Antalya International University, TR
- Alexandru Tomescu
University of Helsinki, FI
- Leo van Iersel
TU Delft, NL
- Mathias Weller
University Paris-Est – Marne-la-Vallée, FR
- Kristina Wicke
Universität Greifswald, DE
- Norbert Zeh
Dalhousie University – Halifax, CA