
Workshop on Detection and Design for Cognitive Biases in People and Computing Systems

Tilman Dingler

University of Melbourne
Melbourne, Australia
tilman.dingler@unimelb.edu.au

Koichi Kise

Osaka Prefecture University
Osaka, Japan
kise@cs.osakafu-u.ac.jp

Benjamin Tag

University of Melbourne
Melbourne, Australia
benjamin.tag@unimelb.edu.au

Andreas Dengel

German Research Center for
Artificial Intelligence (DFKI)
Kaiserslautern, Germany
Andreas.Dengel@dfki.de

Evangelos Karapanos

Cyprus University of Technology
Limassol, Cyprus
evangelos.karapanos@cut.ac.cy

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

Copyright is held by the author/owner(s).

ACM ISBN 978-1-4503-6819-3/20/04.

<https://doi.org/10.1145/3334480.3375159>

Abstract

With social computing systems and algorithms having been shown to give rise to unintended consequences, one of the suspected success criteria is their ability to integrate and utilize people's inherent cognitive biases. Biases can be present in users, systems and their contents. With HCI being at the forefront of designing and developing user-facing computing systems, we bear special responsibility for increasing awareness of potential issues and working on solutions to mitigate problems arising from both intentional and unintentional effects of cognitive biases. This workshop brings together designers, developers, and thinkers across disciplines to redefine computing systems by focusing on inherent biases in people and systems and work towards a research agenda to mitigate their effects. By focusing on cognitive biases from content or system as well as from a human perspective, this workshop will sketch out blueprints for systems that contribute to advancing technology and media literacy, building critical thinking skills, and depolarization by design.

Author Keywords

Cognitive biases; System bias; Cognition-aware systems

Background

Serious concerns have been raised about algorithms and social computing systems manipulating opinion and deci-

sion making, and their suspected contribution to the creation of so-called *echo chambers* [6, 10] and spread of disinformation [12]. Only recently, researchers have started investigating the unintended consequences of such systems [2, 8]. By optimizing recommendation systems, for example, around the contents' 'stickiness' and keeping users' attention [1], these algorithms tend to reinforce people's biases and subsequently contribute to an increasing polarization [11].

The prevalence of cognitive biases make users of such systems susceptible to various types of manipulation. Prominent examples include *confirmation biases* (predominantly seeking out information that confirms existing views), *cognitive dissonance* (repudiation of information that does not fit into preconceived notions), or the so-called *Von Restorff Effect*, which states that a particular item that sticks out is more likely to be remembered than other items. Such biases can be found in both individuals and organizations and are a crucial obstacle for rational, logical discussions about polarizing topics.

A significant contributor to this problem is the algorithm-driven infrastructure of today's media landscape. Social Media (*e.g.*, Twitter or Facebook) as well as content distribution services (*e.g.*, YouTube)—financed through advertisements—are compelled to find ways to keep users engaged for as long as possible. A very effective mechanism is the implementation of learning algorithms that select and recommend additional contents based on prior user selections and predicted interests. These algorithms cater to the users' interests and simultaneously nurture inherent biases. Especially when it comes to polarizing topics, such as climate change, immigration, abortion rights or gun control, the selective distribution of information to receptive users fosters reinforcement of opinions and leads to the devel-

opment of so-called *Filter Bubbles* [10]. Users are often unaware of roaming inside such bubbles, which results in biases prevailing without the user being explicitly aware of them.

The goal of this workshop is to re-think the incentive structures and mechanisms of social computing systems with particular regard to news media and people's cognitive biases. We will discuss the design, implementation, and effects of sensor techniques and computing systems that detect and mitigate biases and subsequently invite users to reflect on their views, acquire and advance media literacy and build critical thinking skills. By focusing on cognitive biases from a content or system perspective as well as from a human perspective, we intend to sketch out blueprints for systems that contribute to a more informed public discourse and depolarization by design. The nature and scope of this workshop will bring together researchers across disciplines, including cognitive psychology, philosophy, information retrieval, and HCI. The goal is to establish a research agenda around cognitive biases, their detection, utilization, and possible fortification as a response to emerging changes in societal and public discourse contributed to by recent technological and political developments.

Themes

The following themes will set the tone for the workshop discussions and attendees' submissions:

- **Detection of Cognitive Biases in People:** which methodologies can be deployed to detect cognitive biases in people?
- **Detecting and Overcoming Biases in Systems:** how can we challenge the development of filter bubbles? Can artificial intelligence be unbiased? Which

algorithms support the development of unbiased systems? Are biases always bad, should some biases be reinforced?

- **Application Scenarios:** in which application fields are bias detection and mitigation necessary to ensure unbiased decision making or depolarized discourse? How can HCI research contribute to the deployment of ethical, fair, and unbiased systems?
- **Case Studies:** presentation of concrete cases where the prevalence and utilization of biases in systems, societies, people, media, and computing systems have been investigated.
- **Tools and Methods:** which modalities enable the quantification of biases, their directedness, and effects? How can tools help to increase self-awareness of biases at work?
- **Creating interventions:** how can targets of biased behaviour be protected (through computing systems, regulations, interventions)? How can biased behaviour be prevented, stopped, mitigated through algorithms [9], system design, interfaces, or agents [4]? Which methods can be used to counteract cognitive biases (e.g., through so-called *nudges* [3]).
- **Inoculation techniques:** the design and creation of systems and methods to help users build resistance against their own biased tendencies. This can include the identification and mitigation of fake news and populist messaging, but also ways to build empathy and critical thinking abilities.

Goals and Outcomes

The goal of the workshop is to foster a research community around the investigation of cognitive biases and how they

influence our interactions with computing systems and our susceptibility to messaging in digital media. As a field in the intersection of Computer Science and Psychology, HCI is well suited for contributing in a variety of ways. We are aiming at building a community with a research agenda around detecting, mitigating, and designing for cognitive biases.

With HCI being at the forefront of designing and developing user-facing computing systems, as designers and technologists, we also bear responsibility for increasing awareness of potential issues and working on solutions to mitigate these problems. To foster and educate fellow researchers, and thereby, raise awareness and self-awareness is consequently an essential goal of this workshop. The workshop themes and accepted submissions will be used as starting points for discussions.

The organizers will document the results of the discussions and will make the collected information available to workshop participants and the broader community by means of online repositories and a public website. Outcomes include a detailed research agenda forward and a cross-discipline community of international esteem.

Workshop Structure

Prior to the workshop, we will issue a call for participation along with a request for submission of position papers. These will be in the CHI extended abstract format and should include no more than 8 pages. Submissions will be peer-reviewed by a select program committee and curated by the organizing committee with the goal of a diverse discussion-fostering program. We plan to use easychair for managing submissions. The workshop will be announced via its website¹ where we will collate materials relevant to

¹<http://critical-media.org/cobi/>

the workshop themes a priori. We will additionally disseminate a Call for Participation via the well-known mailing lists and through the organizers' networks.

Participants

We plan for around 30 participants to be part of the workshop (including organizers). The workshop's program committee will make a selection of presenters based on submitted position papers and with an eye on inter-disciplinary viewpoints.

Presentations

The organizers will publicly call for participation, and all submissions will be peer-reviewed. Accepted submissions are invited for oral presentation at the workshop. A particular focus will be put on extended discussion times since the topic of cognitive biases can be inherently controversial. The organizers also intend to invite two keynote speakers, preferably one from the field of Computer Science (*e.g.*, Deb Roy) and another from the field of Psychology/Philosophy (*e.g.*, Stephan Lewandowsky).

Bias Detection Demonstration

The organizers will demonstrate a tool they developed to identify users' implicit preferences on polarizing topics. The tool is based on a method prominently used in social psychology research to detect the strength of a person's subconscious associations: the Implicit Association Test (IAT) [7]. We appropriated the test for HCI research by constructing brief versions of the IAT on thematic issues, including climate change, immigration attitude, political leaning, and feminism.

Workshop participants will be invited to self-test the tool and—with their explicit consent—share the results with other workshop participants. After an in-depth discussion of the

results, we plan to form groups based on participants' diverse IAT scoring to balance diverse groups.

Action Groups

Each group (approximately 5-6 members each) will receive a set of tasks that consist of

- analyzing an assigned issue of controversial nature,
- identifying the cognitive biases at play,
- discussing and conceptualizing how biases can be mitigated or utilized in this context, and
- developing a conceptual prototype as a solution proposal.

One half of the action groups will work on how to mitigate the impact of cognitive biases, and the other half will work on utilizing cognitive biases for improving human behaviour, critical thinking abilities, lifestyle, or well-being.

Organizers

Tilman Dingler is a postdoctoral researcher and associate lecturer at the School of Computing and Information Systems at the University of Melbourne. He builds and investigates cognition-aware computing systems that support information intake and learning. Tilman is initiator of the WAHM series, *i.e.*, a four-workshop series on *Ubiquitous Technologies for Augmenting the Human Mind* [5].

Benjamin Tag is a postdoctoral researcher at the School of Computing and Information Systems at the University of Melbourne. His research focuses on devising psychological test methods to induce changes in cognitive states.

Evangelos Karapanos is an Assistant Professor at the Department of Communication and Internet Studies of the Cyprus University of Technology, where he leads the Persuasive Technologies Lab. His research focuses on the design of behavior change technology. His "23 ways to Nudge" paper, presented at CHI 2019, reviewed the use of nudging in the field of HCI.

Koichi Kise is a professor at the Department of Computer Science and Intelligent Systems, Osaka Prefecture University, Japan, where he is also the head of the Intelligent Media Processing Group. His research interests are in the areas of document analysis, human document interaction, with a special focus on learning.

Andreas Dengel is the Site Head at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern and the Scientific Director of the Smart Data Knowledge Services Research Department at DFKI, as well as a Professor at the Computer Science Department of the University of Kaiserslautern. His main research interests are in the areas of machine learning, pattern recognition, immersive quantified learning, data mining, and semantic technologies.

Call for Participation

The effectiveness of micro-targeting and the success of falsified information distributed through social media channels have led to increasing concerns about biased information consumption. Algorithms accurately learn from user behaviour and feed the users' hunger for confirmation. An unintended side-effect is that these predictive algorithms ultimately reinforce users' biases. Twitter's CEO Jack Dorsey expressed concerns about the potential to foster the development of filter bubbles, i.e. to only encounter information and opinions reinforcing a person's own beliefs: "I think

Twitter does contribute to filter bubbles, and I think that's wrong of us and we need to fix it".

The "Workshop on Detection and Design for Cognitive Biases in People and Computing Systems" aims at bringing together researchers from HCI, Psychology, Philosophy, and Behavioral Economics to foster an insightful exchange of ideas, a fruitful discussion, and discoveries to increase our self-awareness, and that of our community. We invite position papers discussing the role of computing systems as potential amplifiers of cognitive biases, as well as their role in supporting mitigation of the impact of cognitive biases.

Candidates are requested to submit a position paper (4-8 pages of the ACM Extended Abstract Format) to the submission system (link will be provided). We welcome submissions covering a wide field of topics related to cognitive biases, but put special focus on methods and tools to detect cognitive biases, techniques to mitigate their negative impact or utilize their positive effects, application scenarios, case studies, and findings on the creation of intervention and inoculation techniques that counteract biases at play or help building resistance against biases.

Acknowledgements

This research is supported by JST CREST, Grant No. JP-MJCR16E1.

REFERENCES

- [1] Alexander Benlian. 2015. Web Personalization Cues and Their Differential Effects on User Assessments of Website Value. *Journal of Management Information Systems* 32, 1 (2015), 225–260. DOI: <http://dx.doi.org/10.1080/07421222.2015.1029394>

- [2] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 377, 14 pages. DOI : <http://dx.doi.org/10.1145/3173574.3173951>
- [3] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 503, 15 pages. DOI : <http://dx.doi.org/10.1145/3290605.3300733>
- [4] Tilman Dingler, Ashris Choudhury, and Vassilis Kostakos. 2018. Biased Bots: Conversational Agents to Overcome Polarization. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18)*. ACM, New York, NY, USA, 1664–1668. DOI : <http://dx.doi.org/10.1145/3267305.3274189>
- [5] Tilman Dingler, Kai Kunze, Nigel Davies, Albrecht Schmidt, Marc Langheinrich, and Niels Henze. 2014. WAHM 2014: workshop on ubiquitous technologies for augmenting the human mind. In *2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2014*. Association for Computing Machinery, Inc, 1339–1345.
- [6] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.
- [7] G Greenwald, Debbie E Mcghee, and Jordan L K Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology* 74, 6 (1998), 1464–1480.
- [8] Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck. 2018. Fair, Transparent, and Accountable Algorithmic Decision-making Processes. *Philosophy & Technology* 31, 4 (01 Dec 2018), 611–627. DOI : <http://dx.doi.org/10.1007/s13347-017-0279-x>
- [9] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 677–686. DOI : <http://dx.doi.org/10.1145/2566486.2568012>
- [10] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.
- [11] Briony Swire, Adam J. Berinsky, Stephan Lewandowsky, and Ullrich K. H. Ecker. 2017. Processing political misinformation: comprehending the Trump phenomenon. *Royal Society Open Science* 4, 3 (March 2017), 160802. DOI : <http://dx.doi.org/10.1098/rsos.160802>
- [12] C. Wardle and H. Derakhshan. 2017. Information Disorder: Towards an interdisciplinary framework for research and policy making. (September 2017). <https://bit.ly/2gQVcvT>