

# Marko Tešić

Leverhulme Centre for the Future of Intelligence  
University of Cambridge  
Level 1, 16 Mill Lane, Cambridge, CB2 1SB, UK

✉ [mt961@cam.ac.uk](mailto:mt961@cam.ac.uk)  
🌐 [markotesic.org](http://markotesic.org)  
in, 📷, 🐦

## Research Experience

- 
- Research Associate** at **Leverhulme Centre for the Future of Intelligence, University of Cambridge** *July 2023 – Present*
- Evaluating AI capabilities with a focus on Large Language Models (LLMs) and exploring their relevance to specific job tasks. Done in collaboration with the OECD.
- Royal Academy of Engineering UK IC Postdoctoral Research Fellow** at Birkbeck, University of London *Mar. 2021 – Feb. 2023*
- Explored how explanations of AI predictions influence human beliefs and decision-making.
- Researcher** on *The Bayesian Approach to Robust Argumentation Machines* project at MCMP, LMU, Munich & Birkbeck, University of London *Sep. 2021 – Feb. 2023*
- Automated argument generation and evaluation from Bayesian network models.
- Data Study Group (DSG) Principal Investigator** at the **Alan Turing Institute** *Oct. 2022 – April 2023*
- Led the scoping and execution of a data science challenge with the UK Department for Transport, supervising participants and ensuring high-quality solutions and reporting.
- DSG Researcher** at **the Alan Turing Institute and LIDA, University of Leeds** *July 5–23, 2021*
- Optimizing Morrisons supermarkets' supply chain as part of a DSG team
  - Analyzed data & trained gradient boosting tree models to predict future supplies
- Research Intern** at **BlackRock, Factor Based Strategies Group** *Oct. 2019 – Mar. 2020*
- Developed causal Bayesian models of investment factors and ESG criteria.
- Member of the Translation Team UK** on the project *Bayesian Argumentation via Delphi (BARD)* within **IARPA** at Birkbeck, University of London & UCL *Oct. 2017 – Nov. 2018*
- Created intelligence gathering-inspired situations
  - Built Bayesian network models of these situations
  - Fully designed, ran, and analyzed experiments testing people's evidential, causal, and probabilistic reasoning
- Member of the Research Team** on the project *Scientific Reasoning and Argumentation* at the **Center for Advanced Studies**, LMU, Munich *Oct. 2016 – Sep. 2017*
- Worked on explicating an inference pattern called 'Inference to the Best Explanation' (IBE) in Bayesian terms

## Education

- 
- Ph.D. in Psychology** *2020*  
Department of Psychological Sciences, Birkbeck, University of London, UK  
Thesis: *Explanation and Argument*  
Focus: causal-probabilistic reasoning, Bayesian networks, psychology of explanations  
Supervisors: **Ulrike Hahn** and **David Lagnado**
- M.A. in Logic and Philosophy of Science** *2016*  
Munich Center for Mathematical Philosophy, Ludwig Maximilian University, Munich, Germany
- B.A. in Philosophy** *2014*  
University of Belgrade, Serbia

## Publications

---

- **Marko Tešić\***, Lorenzo Pacchiardi\*, Lucy Cheke, José Hernández-Orallo (2024). **Leaving the barn door open for Clever Hans: Simple features predict LLM benchmark answers.** *Preprint*
- Matteo Gabriel Mecattaf\*, Ben Slater\*, **Marko Tešić**, Jonathan Prunty, Konstantinos Voudouris, Lucy Cheke (2024). **A little less conversation, a little more action, please: Investigating the physical common-sense of LLMs in a 3D embodied environment.** *Preprint*
- Rakshit S. Trivedi, Akbir Khan, Jesse Clifton, Lewis Hammond, Edgar A. Duéñez-Guzmán, John P. Agapiou, Jayd Matyas, Sasha Vezhnevets, Dipam Chakraborty, Yue Zhao, **Marko Tešić**, Barna Pásztor, Yunke Ao, Omar G. Younis, Jiawei Huang, Benjamin Swain, Haoyuan Qin, Mian Deng, Ziwei Deng, Utku Erdoğanaras, Natasha Jaques, Jakob Nicolaus Foerster, Vincent Conitzer, José Hernández-Orallo, Dylan Hadfield-Menell, Joel Z. Leibo (2024). **Melting Pot Contest: Charting the Future of Generalized Cooperative Intelligence.** *NeurIPS 2024 Track on Datasets and Benchmarks.*
- Rafael Fuchs, **Marko Tešić**, & Ulrike Hahn (2024). **Testing the maximum entropy approach to awareness growth in Bayesian epistemology and decision theory.** *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*
- **Marko Tešić** & Ulrike Hahn (2023). **The impact of explanations as communicative acts on belief in a claim: The role of source reliability.** *Cognition*, 240(105586).
- Ulrike Hahn & **Marko Tešić** (2023). **Argument and Explanation.** *Philosophical Transactions of the Royal Society A*, 381(2251). Theme issue on *Cognitive Artificial Intelligence*.
- **Marko Tešić** & Ulrike Hahn (2022). **Can counterfactual explanations of AI systems' predictions skew lay users' causal intuitions about the world? If so, can we correct for that?** *Patterns*, 3(12).
- Data Study Group team. (2022). Data Study Group Final Report: Morrisons. Zenodo. <https://doi.org/10.5281/zenodo.6498140>.
- **Marko Tešić** (2021). On the transferability of insights from the psychology of explanation to explainable AI. **Human Centered AI workshop at NeurIPS 2021.**
- **Marko Tešić** & Ulrike Hahn (2021). **Explanation in AI systems.** In S. Muggleton & N. Chater (Eds.), *Human-Like Machine Intelligence* (pp. 114–136). Oxford University Press.
- **Marko Tešić\***, Alice Liefgreen\*, & David Lagnado (2020). **The propensity interpretation of probability and diagnostic split in explaining away.** *Cognitive Psychology*, 121.
- Alice Liefgreen & **Marko Tešić** (2020). **Explaining away and the propensity interpretation of probability: The case of unequal priors.** In C. Dutilh Novaes, H. Jansen, J. A. van Laar, & B. Verheij (Eds.), *Reason to dissent. Proceedings of the 3rd European Conference on Argumentation, Vol. III* (pp. 385–403). College Publications.
- Nicole Cruz, Saoirse Desai, Stephen Dewitt, Ulrike Hahn, David Lagnado, Alice Liefgreen, Kirsty Phillips, Toby Pilditch & **Marko Tešić** (2020). **Widening access to Bayesian problem solving.** *Frontiers in Psychology*, 11, 660.
- **Marko Tešić** & Ulrike Hahn (2019). **Sequential diagnostic reasoning with independent causes.** In A.K. Goel, C.M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2947–2953). Montreal, QB: Cognitive Science Society.
- Alice Liefgreen\*, **Marko Tešić\***, & David Lagnado (2018). **Explaining away: Significance of priors, diagnostic reasoning, and structural complexity.** In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 2047–2052). Austin, TX: Cognitive Science Society.
- **Marko Tešić**, Ben Eva, & Stephan Hartmann (2017). **Confirmation by Explanation: A Bayesian Justification of IBE.** *Preprint*
- **Marko Tešić** (2017). **Confirmation and the generalized Nagel-Schaffner model of reduction: A Bayesian analysis.** *Synthese*, 196(3), 1097–1129. DOI: 10.1007/s11229-017-1501-1.

\* equal contribution

## Selected Presentations and Workshops

---

### Robust evaluation of Generative AI

- I presented a tutorial on evaluating large language models at the European Association for Data Science summer school on generative AI. June 20, 2024

### AAAI-24 tutorial: Measurement Layouts for Capability-Oriented AI Evaluation

- I co-organized a tutorial showcasing measurement layouts (Bayesian hierarchical models) for inferring AI capabilities. I presented my work on inferring and predicting the capabilities of large language models. February 20, 2024

## Workshop on Human Behavioral Aspects of (X)AI

- I organized a workshop bringing together cognitive scientist and ML researchers from academia, industry and government working on and with (explainable) AI. *September 23–24, 2022*

## Supervision, Teaching & Admissions Experience

<b>Managing</b> a Postdoctoral Research Associate and a Research Assistant working on an <b>Accenture</b> -supported project evaluating the core cognitive capabilities of AI relevant to workplace tasks.	<i>January 2024 – Present</i>
Leverhulme Centre for the Future of Intelligence, University of Cambridge	
<b>Postgraduate admissions</b> for the MSt and MPhil courses in <i>AI Ethics &amp; Society</i> and <i>Ethics of AI, Data &amp; Algorithms</i>	<i>Spring 2024</i>
Leverhulme Centre for the Future of Intelligence, University of Cambridge	
<b>Marking dissertations</b> for the MSt and MPhil courses in <i>AI Ethics &amp; Society</i> and <i>Ethics of AI, Data &amp; Algorithms</i>	<i>Spring 2024</i>
Leverhulme Centre for the Future of Intelligence, University of Cambridge	
<b>Research staff recruitment.</b> Shortlisting and interviewing for Postdoctoral Research Associate and Research Assistant roles.	<i>Winter 2023</i>
Leverhulme Centre for the Future of Intelligence, University of Cambridge	
<b>Visiting Lecturer</b> for the M.A. courses <i>Computational Approaches to Mind</i> and <i>Fundamental Debates in Cognitive Science</i>	<i>Jan. 2023 – Apr. 2023</i>
Department of Psychological Sciences, Birkbeck, University of London	
<b>Taught:</b> Bayesian modeling, Agent-based modeling, and Marr's levels of explanation	
<b>Visiting Lecturer</b> for the M.A. course <i>Cognitive and Economic Science of Rational Choice</i>	<i>Oct. 2020 – Dec. 2020</i>
Department of Psychology and Economics, City, University of London	
<b>Taught:</b> Rationality as logic and as probability theory, Probabilistic fallacies, and Causal reasoning and modeling	
<b>Seminar leader</b> for the M.A. courses <i>Neuroscience, Individual Differences, Social Psychology</i> , and <i>Developmental Psychology</i>	<i>Feb., Nov. 2020; Feb. 2021</i>
Department of Psychological Sciences, Birkbeck, University of London	
<b>Tutor</b> for the B.A. course <i>Logic and Discrete Structures</i>	<i>Summer 2017</i>
Computer Science Department, Ludwig Maximilians University	
<b>Teaching assistant</b> for the M.A. course <i>Central Topics in Philosophy of Science</i>	<i>Winter 2016</i>
Munich Center for Mathematical Philosophy, Ludwig Maximilians University	
<b>Tutor</b> for the B.A. course <i>Logic 1</i>	<i>Winter 2016</i>
Faculty of Philosophy, Ludwig Maximilians University	

## Honors and Awards

<b>The Alan Turing Institute Post-Doctoral Enrichment Award</b>	<i>July 2022 – Jan. 2023</i>
<b>The Royal Academy of Engineering</b> UK IC Postdoctoral Research Fellowship (£200,000)	<i>Mar. 2021 – Feb. 2023</i>
Ph.D. studentship from the Department of Psychological Sciences, Birkbeck, UoL	<i>2018 – 2020</i>
Ph.D. studentship from the BARD project	<i>2017 – 2018</i>
<b>Dositeja</b> scholarship for graduate studies	<i>2017/18; 2015/16; 2014/15</i>
<b>BAYHOST</b> scholarship for graduate studies	<i>2015/16; 2014/15</i>