Curriculum Vitae

Michael K. Cohen

Student, DPhil Engineering Science, Oxford University
Email: michael.cohen@eng.ox.ac.uk

Education

University of Oxford	2019-
DPhil in Engineering Science	Oxford, UK
Research Advisor: Mike Osborne	

Australian National University	2017-2019
Advanced Master of Computing	Canberra, Australia
Research Advisor: Marcus Hutter	
Cumulative GPA: 7.00/7.00	

Yale University	2011-2015
B.A. (Hons.) Chemistry, *magna cum laude*	New Haven, USA
Research Advisor: Patrick Vaccaro	
Cumulative GPA: 3.90/4.00	

Awards

Australian National University's University Medal	2019

Grants

Future of Humanity Institute – DPhil Scholarship (£19,000/year + tuition)	2019-
Open Philanthropy Project – AI Scholarship ($83,530)	2017-2019

Publications

Cohen, M. K., Vellambi B., & Hutter, M. (2021) Intelligence and Unambitiousness Using Algorithmic Information Theory. *IEEE Journal of Selected Areas in Information Theory.*

Cohen, M. K., Catt, E., & Hutter, M. (2021) Curiosity Killed or Incapacitated the Cat and the Asymptotically Optimal Agent. *IEEE Journal of Selected Areas in Information Theory.*

Cohen, M. K. & Hutter, M. (2020). Pessimism About Unknown Unknowns Inspires Conservatism. In *Proc. COLT-20.*

Cohen, M. K., Vellambi, B., & Hutter, M. (2020). Asymptotically Unambitious Artificial General Intelligence. In *Proc. AAAI-20.*

Cohen, M. K., Catt, E., & Hutter, M. (2019). A Strongly Asymptotically Optimal Agent in General Environments. In *Proc. IJCAI-19.*

Cohen, M. (2017). Intra-feature Random Forest Clustering. In *International Workshop on Machine Learning, Optimization, and Big Data* (pp. 41-49). Springer, Cham.

Nemchick, D. J., Cohen, M. K., & Vaccaro, P. H. (2016). Dual hydrogen-bonding motifs in complexes formed between tropolone and formic acid. *The Journal of Chemical Physics*, 145(20), 204303.

Nemchick, D., Cohen, M., & Vaccaro, P. (2015). Dispersion-Dominated π-Stacked Complexes Constructed on a Dynamic Scaffold. In *70th International Symposium on Molecular Spectroscopy* (Vol. 1).

Under Review:
Cohen, M. K., Hutter, M., Osborne, M. A. Advanced Artificial Agents Intervene in the Provision of Reward.

Cohen, M. K., Hutter, M., Nanda, N. Fully General Online Imitation Learning.

Teaching

Guest Lecturer/TA. Autonomous Intelligent Machines and Systems. Postgraduates,   Fall 2020, 2021
Oxford University.

Invited Talks

Center for Human-Compatible AI Virtual Workshop, UC Berkeley. Advanced Artificial Agents Intervene in the
    Provision of Reward. June 2021.
Google DeepMind (Safety Team). Fully General Online Imitation Learning. February 2021.
Cambridge AI Safety Reading Group. Pessimism About Unknown Unknowns Inspires Conservatism. November
    2020.
Google DeepMind (Foundations team). Pessimism About Unknown Unknowns Inspires Conservatism.
    September 2020.
AI Ethics London. AI Safety in Bayesian Reinforcement Learning. February 2020.
Google DeepMind (Safety team). Pessimism About Unknown Unknowns Inspires Conservatism. February 2020.
Effective Altruism Oxford. Expected Behavior of Advanced Reinforcement Learners. October 2019.
Google DeepMind (Safety team). Asymptotically Unambitious AGI. October 2019.
Center for Human-Compatible AI, UC Berkeley. Curiosity Killed the Cat and the Asymptotically Optimal Agent.
    September 2019.
Center for Human-Compatible AI, UC Berkeley. Asymptotically Unambitious AGI. January 2019.

Employment

Visiting Researcher, Center for Human-Compatible AI, UC Berkeley, Berkeley, CA	2017-2018
Data Science Associate, Noodle.ai, Palo Alto, CA	2017
Computer Science Teacher, Teach for America, Lazear Charter Academy, Oakland, CA	2015-2016

Academic Service
Reviewing:

JMLR	2020, 2021
Journal of Consciousness Studies	2021
AAAI 2021	2020
Journal of AGI	2020
AGI 2020	2020