

# Tom EVERITT

## PERSONAL INFORMATION

---

NATIONALITY: Swedish  
EMAIL: [tom4everitt@gmail.com](mailto:tom4everitt@gmail.com)  
WEBSITE: <http://www.tomeveritt.se>

## EDUCATION AND RELEVANT WORK EXPERIENCE

---

- 2015–current PHD STUDENT at **Australian National University** with supervisor Marcus Hutter.  
Thesis topic AI SAFETY.
- 2014 CO-FOUNDED [Kaus.se](http://kaus.se), a company/webservice doing causal analysis of health data.  
COURSE COORDINATOR of pre-university math course at **Stockholm University**.
- 2013–2014 TEACHING ASSISTANT at **Stockholm University** and **The Royal Institute of Technology (KTH)**,  
tutoring mathematics courses on linear algebra and analysis (details below).  
Studied additional courses in THEORETICAL COMPUTER SCIENCE and MATHEMATICS.  
Publication of two scientific articles (see below).
- 2011–2013 MASTER'S DEGREE IN MATHEMATICS, **Stockholm University**,  
including 6 months degree project at **Australia National University**, Canberra,  
with Professor Marcus Hutter.  
Thesis: "Universal Induction and Optimisation: No Free Lunch?" [PDF](#)  
GPA: 3.76/4
- 2007–2010 BACHELOR'S DEGREE IN MATHEMATICS, **Stockholm University**  
Personal focus: Mathematical Logic and Theoretical Computer Science  
Thesis: "Automated Theorem Proving" [PDF](#)  
GPA: 3.81/4  
Additional courses in PHILOSOPHY and COMPUTER SCIENCE, **Stockholm University**
- 2006–2007 First year of the COGNITIVE SCIENCE PROGRAM, **Umeå University**  
GPA: 3.9/4

## PUBLICATIONS

---

- Everitt, T. and Hutter, M.  
Universal Artificial Intelligence: Practical Agents and Fundamental Problems  
To appear in *Foundations of Trusted Autonomy*, Springer, 2017.
- Everitt, T., Filan, D., Daswani, M., and Hutter, M.  
Self-Modification of Policy and Utility Function in Rational Agents.  
In *Proceedings of Artificial General Intelligence (AGI)*, 2016. [PDF](#)  
Winner of the Kurzweil Prize for best AGI paper.
- Everitt, T. and Hutter, M.  
Avoiding Wireheading with Value Reinforcement Learning.  
In *Proceedings of Artificial General Intelligence (AGI)*, 2016. [PDF](#)
- Martin, J., Everitt, T., and Hutter, M.  
Death and Suicide in Universal Artificial Intelligence.  
In *Proceedings of Artificial General Intelligence (AGI)*, 2016.
- Everitt, T. and Hutter, M.  
Analytical Results on the BFS vs. DFS Algorithm Selection Problem. Part I Tree Search.  
In *Proceedings of the 28th Australasian Joint Conference on AI*, 2015. [PDF](#)
- Everitt, T. and Hutter, M.  
Analytical Results on the BFS vs. DFS Algorithm Selection Problem. Part II Graph Search.

In *Proceedings of the 28th Australasian Joint Conference on AI*, 2015. [PDF](#)

- Everitt, T., Leike, J., and Hutter, M.  
Sequential Extensions of Causal and Evidential Decision Theory.  
In *Proceedings of Algorithmic Decision Theory (ADT)*, 2015. [PDF](#)
- Everitt, T., Lattimore, T., and Hutter, M.  
Free Lunch for Optimisation under the Universal Distribution.  
In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC)*, 2014. [PDF](#)
- Alpcan, T., Everitt, T., and Hutter, M. (2014).  
Can we measure the difficulty of an optimization problem?  
In *Proceedings of the IEEE Information Theory Workshop (ITW)*, 2014. [PDF](#)

## PUBLIC OUTREACH AND COMMUNITY SERVICE

---

- Organiser of **MIRIx Canberra** workshops on WIREHEADING, AI SAFETY RESEARCH AGENDAS, SUPERINTELLIGENCE, MODEL-FREE AIXI, and LOGICAL INDUCTION 2015 and 2016.
- Tutorial on GÖDEL'S INCOMPLETENESS THEOREMS for **MIRIx Canberra** workshop, 2015.
- Interviewed in a movie on AI PREDICTIONS by **Cartina**, Stockholm, 2015.
- Reviewed papers for the journals **Algorithmica** and **Artificial Intelligence Journal**, and the conferences **Algorithmic Learning Theory (ALT)** 2015, **Uncertainty in Artificial Intelligence (UAI)** 2016, **Artificial General Intelligence (AGI)** 2015 and 2016, and **Australasian AI** 2015.

## CONFERENCES AND TALKS

---

### Conferences

- **Artificial General Intelligence 2016**, New York.  
Tutorial on Universal Artificial Intelligence.  
Paper presentations of self-modification and wireheading.
- **Australasian AI Conference 2015**, Canberra, December 2–4.  
Talk on ANALYTICAL RESULTS ON THE BFS VS. DFS ALGORITHM SELECTION PROBLEM.
- **Congress on Evolutionary Computation 2014**, part of IEEE World Congress on Computational Intelligence (IEEE WCCI), Beijing, July 6–11.  
Talk on UNIVERSAL INDUCTION AND OPTIMISATION: NO FREE LUNCH?
- **Artificial General Intelligence 2011**, Mountain View, California, August 3–7.
- **Philosophy and the Foundations of Mathematics 2009**: A conference dedicated to Per Martin-Löf on the occasion of his retirement, Uppsala, May 5–8.

### Other talks

- **AI and the Future**, an introductory talk about AI safety, at REGNET (ANU public policy research centre), ANU LEARNING COMMUNITIES, and ANU XSA, 2016.
- **Formal Models of AI Safety**, an overview of AI safety research progress, at EFFECTIVE ALTRUISM SYDNEY RETREAT, 2016.
- **Avoiding Wireheading**, at MIRI CSRBAI and UC BERKELY AI SAFETY GROUP, 2016.
- **Wireheading and Self-Modification**, at University of Alberta, 2016.
- **A Mathematical Theory of Cooperation**. Vetenskapens hus, March 2014.
- **Universal Induction and Optimisation: No Free Lunch?** Australian National University, January 2013 and Stockholm University, March 2013.
- **Meta-rationality: Choosing Computational Actions**. Australian National University, October 2012.
- **Automated Theorem Proving**. Stockholm University, June 2010.

## TEACHING AND SUPERVISION

---

I have tutored the following courses at Australian National University (ANU), Stockholm University (SU), and Royal Institute of Technology (KTH).

- Advanced Topics in Artificial Intelligence (ANU COMP4620/COMP8620)
- Mathematics I (SU MM2001) (3 times)
- Calculus II (KTH SF1603)
- Linear Algebra (KTH SF1604)
- Applied Linear Algebra with Numerical Methods (KTH SF1663)

I have also supervised:

- Jarryd Martin (Master thesis, ANU). Score: 92.
- Suraj Sasikumar (Master thesis, ANU). Ongoing.

## COMPUTER SKILLS

---

I am proficient in

- Python, Coffeescript, Java, R (Statistical Computing), Linux (mainly Ubuntu), LaTeX,

and have some experience with

- Mathematica, C/C++, Prolog, Scala, HTML/CSS, PHP, SQL, JavaScript/Node.js, Matlab/Octave, Scheme/Racket.