

Daniel Filan – Curriculum Vitæ

daniel.filan@berkeley.edu

AREAS OF INTEREST	AI Safety (Value Alignment, Corrigibility, Transparency), Theory of Artificial Intelligence (Reinforcement Learning, Algorithmic Information Theory, Statistical Machine Learning).	
DEGREES	<i>Doctor of Philosophy in Computer Science</i> , University of California, Berkeley	2016 – present
	<ul style="list-style-type: none">Studying AI safety, supervised by Stuart Russell.	
	<i>Bachelor of Philosophy (Hons)</i> , Australian National University	2012 – 2015
	<ul style="list-style-type: none">Honours in Computer Science, undergraduate studies in Mathematics and Physics.Thesis: “Resource-bounded Complexity-based Priors for Agents”, supervised by Marcus Hutter.GPA: 7.00/7.00, 1st Class Honours.	
PUBLICATIONS	<ul style="list-style-type: none">Loss Bounds and Time Complexity for Speed Priors. With Jan Leike and Marcus Hutter. AISTATS 2016.Self-modification of Policy and Utility Function in Rational Agents. With Tom Everitt (lead author), Mayank Daswani, and Marcus Hutter. AGI 2016, recipient of Kurzweil Prize for Best Paper.	
SELECTED AWARDS	<i>University Medal</i> , Australian National University	2015
	<ul style="list-style-type: none">Prize; awarded to students who have obtained First Class Honours (or Masters Advanced Equivalent) and demonstrated exceptional academic excellence across their studies, the highest academic prize for undergraduates.	
	<i>Erin Brent Computer Science Prize</i> , Australian National University	2015
	<ul style="list-style-type: none">Monetary prize; awarded to the student who achieved the best Honours result in any of the degree programs relating to Computer Science, Software Engineering or Information Technology.	
	<i>National Merit Scholarship</i> , Australian National University	2012 – 2015
	<ul style="list-style-type: none">Annual funding; awarded to the top $\sim 0.5\%$ of school leavers.	
	<i>Hanna Neumann Prize for Second Year Mathematics</i> , Australian National University	2013
	<ul style="list-style-type: none">Monetary prize; awarded to the top student in second year mathematics courses.	
	<i>Dean’s Commendation List</i> , Australian National University	2012
	<ul style="list-style-type: none">Prize; awarded to students who achieve scores of 90 or above in all science courses in a particular year.	
INTERNSHIPS	<i>Future of Humanity Institute, Oxford University</i>	2016
	<ul style="list-style-type: none">Writing code for agentmodels.org, a website designed to explain the use of probabilistic programs to build models of agents and perform inference about them.	
UNDERGRAD RESEARCH	<i>Summer Research Scholar</i> ANU Mathematical Sciences Institute	Summer 2013–2014
	<ul style="list-style-type: none">An investigation into the theory and practice of measure-theoretic image packing.	

Undergraduate Research Projects 2013, 2014
ANU Research School of Computer Science

- Extreme state aggregation beyond MDPs: Tightness of FRL bounds.

Department of Quantum Sciences, ANU Research School of Physics and Engineering

- Proofs of impossibility theorems regarding tests of oneself being in superposition.
- An investigation into the self-gravitation of light in general relativity.

**TEACHING
EXPERIENCE**

Teaching Assistant, MATH2322 Advanced Algebra 1 Semester 2 2015
ANU Mathematical Sciences Institute

Teaching Assistant, MATH2320 Advanced Analysis 1 Semester 1 2015
ANU Mathematical Sciences Institute

Teaching Assistant, COMP2610 Information Theory Semester 2 2014
ANU Research School of Computer Science