Patrick Butlin

Future of Humanity Institute, University of Oxford
patrick.butlin@gmail.com +44 7906210518

AOS: Philosophy of Mind, Philosophy of Cognitive Science (Psychology, Neuroscience & AI)
AOC: Ethics, Logic

Employment

Future of Humanity Institute, University of Oxford
2021- Research Fellow

King's College London
2020-2021 Visiting Research Fellow in Philosophy
2017-2020 Teaching Fellow in Philosophy

Centre for Philosophical Psychology, University of Antwerp
2016-2017 Postdoctoral Fellow

Hertford College, University of Oxford
2014-2016 and Jan-Jun 2017 Stipendiary Lecturer in Philosophy

Research Grant

Survival and Flourishing
Jan-Jun 2021 Support for research on AI alignment and human values ($19,200)

Advisory Position

Technology company in London, name removed for commercial reasons
2018-2020 Advisor on philosophy and cognitive science

Education

King's College London
2011-2015 PhD Philosophy
Supervisors: Prof. David Papineau, Prof. Nicholas Shea

University of Sheffield
2010-2011 MA Cognitive Studies – Distinction

Merton College, University of Oxford
2008-2010 BPhil Philosophy

2003-2007 MMathPhil Mathematics and Philosophy – First Class

Publications

'AI Alignment and Human Reward'. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics and Society (AIES '21)*, forthcoming.

'Affective Experience and Evidence for Animal Consciousness'. *Philosophical Topics*, issue on non-human consciousness, forthcoming.

'Directive Content'. *Pacific Philosophical Quarterly* 102(1): 2-26, 2021.

'Cognitive Models are Distinguished by Content, not Format'. *Philosophy of Science* 88(1): 83-102, 2021.

'Representation and the Active Consumer'. *Synthese* 197: 4533-4550, 2020 (online 2018).

'Why Hunger is not a Desire'. *Review of Philosophy and Psychology* 8(3): 617-635, 2017.

'Normal and Addictive Desires' (with David Papineau). In Nick Heather and Gabriel Segal (eds.), *Addiction and Choice: Rethinking the Relationship*. OUP, 2016.

Presentations

Refereed Presentations:

'AI Alignment and the Complex Structure of Human Values'
- CEPE/IACAP Joint Conference: The Philosophy and Ethics of AI, 7 July 2021

'Reward, Reward Signals, and Agents within Agents'
- 47th Annual Meeting of the Society for Philosophy and Psychology, 30 June 2021

'AI Alignment and Human Reward' (poster)
- AIES '21, 19-21 May 2021

'Sharing Our Conceptual Resources with Machines'
- PLM5, St. Andrews, 29 August 2019
- Kinds of Intelligence 2: Machine Minds, Cambridge, 20 June 2019

'Model-Based Reinforcement Learning and Acting for Reasons'
- Anticipatory Systems: Humans meet AI, Örebro, 10 June 2019

'Pleasure, Pain and the Evolution of Consciousness'
- European Society for Philosophy and Psychology, Rijeka, 13 September 2018

'Information-how and Directive Content'
- Joint Session, Oxford, 8 July 2018

'Liberal Representation and the Active Consumer'
- ECMN Future Minds Conference, Warwick, 16 March 2017

'Liberal Representation and Basic Action'
- 10th Logos Workshop on Naturalistic Theories of Intentionality, Barcelona, 1 December 2016

'Lewis and Anscombe on Direction of Fit'
- Early Career Mind Network, Durham, 1 September 2016

'Naturalizing Direction of Fit'
- KCL-UNC Workshop on 'The Normative in a Natural World', UNC-Chapel Hill, 3-5 May 2013

'Teleosemantics and the Direction of Fit of Desire'
- Graduate Conference in Theoretical Philosophy, University of Groningen, 18-20 April 2013

'Holton's Argument from Strength of Will'
- European Society for Philosophy and Psychology, London, 28-31 August 2012

'What's in a That-Clause?' (with Emanuel Viebahn)
- BPPA Conference, University of Edinburgh, 3-6 September 2012

'Schroeder's *Being For* and the Structure of the Propositional Attitudes'
- Oxford Philosophy Graduate Conference, 20-21 November 2010

Invited Presentations:

'Imperative Intensity'

- Direction of Fit Workshop, Antwerp, 9 November 2016

'Desire as a Natural Kind'

- A Highly Desirable Symposium, Durban, 18 June 2016

'The Direction of Fit of Desire'

- Hertford Philosophical Society, Oxford, 18 February 2016

'Why Hunger is Not a Desire'

- London-Warwick Mind Forum, UCL, 16 June 2015

'Normal and Addictive Desires' (with David Papineau)

- KCL-UNC Workshop on 'Philosophical Approaches to Desire and Pleasure', King's College, London, 19-20 May 2014

'What's in a That-Clause?' (with Emanuel Viebahn)

- Ockham Society, University of Oxford, 12 June 2012

Comments:

On Ethan Jerzak, 'Non-Classical Knowledge'

- London-Berkeley Graduate Conference, UC-Berkeley, 10-11 May 2013

On Sam Wilkinson, 'Explaining Addiction'

- KCL Graduate Conference in Philosophy of Mind and Psychology, 26 April 2013

Works under review

An article on machine understanding of human language. Under review.
An article on addiction and autonomy. Under review.

Teaching

As a Teaching Fellow at King's College London:

Module Convenor and Lecturer:

- Advanced Topics in Philosophy of Mind (Neuroscience and Biomedical Science students; Spring 2019 & 2020)
- Neuroscience and the Mind (philosophy of mind for Neuroscience and Biomedical Science students; Autumn 2018 & 2019)
- Topics in Philosophy of Psychology (BSc Psychology students; Spring 2018)
- Philosophy of Psychology (BSc Psychology students; Spring 2018)
- Philosophy of Psychology (BA and MA Philosophy students; Autumn 2017)

Other classes:

- Neuroscience and the Mind (seminars and revision lectures 2017-18)
- Philosophy for medical students (seminars)
- Ethics I (seminars)

Dissertation supervision for 7 undergraduates, 6 MA students

Tutorials and Classes at Hertford College, Oxford, 2014-2017:

- Philosophy of Mind (5 students)

- Philosophy of Cognitive Science (4 students)
- Ethics (18 students, 3 visiting students)
- Elements of Deductive Logic (12 students)
- Moral Philosophy (first year) (18 students)

As a Graduate Teaching Assistant at King's College London, 2012-2014:

Philosophy of Psychology; Ethics I; Elementary Logic; Ethics II: Contemporary Ethical Philosophy; Topics in Philosophy of Mind; Methodology; Philosophy of Logic and Language; Neuroscience and the Mind

As a Postgraduate Tutor at University of Sheffield, 2010-2011:

Elementary Logic; Mind, Brain and Personal Identity

Public Engagement

Judge, King's College London Regional, John Stuart Mill Cup 2021

'Addiction, Hijacking and Autonomy' talk for KCL Neuroscience Society, 26 November 2019

'Explaining Consciousness' talk at *Consciousness: Neuroscience v. Philosophy*, KCL Neuroscience Society, 21 March 2019

Participant, panel discussion on 'Addiction: What, Who and How?' at *Hooked* exhibition, Science Gallery London, 19 October 2018

Awards and Funding

King's College, London:

King's Undergraduate Research Fellowship (funding for undergraduate research assistant, Summer 2019)
AHRC Doctoral Award (Fees & Full Maintenance, 2011-2014)
AHRC Research Training and Support Grant for visit to Cambridge University (£1092.40, Autumn 2013)

University of Edinburgh:

Principal's Career Development Scholarship (£14000-15000 p.a., 2011-2014) – *declined*

University of North Carolina, Chapel Hill:

Richard Brooke Fellowship (Non-teaching stipend for PhD from 2008) – *declined*

Merton College, University of Oxford:

Postmastership 2005-2007 (highest academic scholarship)
Exhibition 2004 -2005 (academic scholarship)

Professional and Departmental Service

Refereeing for publication:

Biology & Philosophy; The Philosophical Quarterly; Episteme; British Journal for the Philosophy of Science; Philosophy; Mind & Language; Philosophy of Science; Synthese; Pacific Philosophical Quarterly; Theoria; *Philosophical Psychology; SAGE Publishing*

Departmental service at King's College, London:

- Study abroad tutor
- External relations lead

- Co-convenor, Advanced research seminar & First year research seminar
- Convenor, MA research seminar
- Co-convenor, Philosophy of Mind reading group
- Personal tutor and liaison tutor for Modern Languages and War Studies
- Web and social media lead

Admissions interviewer, Hertford College, 2014 & 2015

Conference organisation:

- Direction of Fit Workshop, Antwerp, November 2016
- 5th Annual UNC-KCL Workshop on 'Philosophical Approaches to Desire and Pleasure', May 2014

Co-organiser, ECMN Work in Progress Seminars

Organiser, BPhil Reading Parties 2009 & 2010

Non-Academic Employment

2007-2008 Research Analyst at ECA International, London

Referees

Prof. David Papineau
Department of Philosophy, King's College, London
Strand, London WC2R 2LS
david.papineau@kcl.ac.uk

Prof. Nicholas Shea
Institute of Philosophy, School of Advanced Study, University of London
Senate House, Malet Street, London WC1E 7HU
nicholas.shea@sas.ac.uk

Prof. Bence Nanay
Centre for Philosophical Psychology, University of Antwerp
D413, Grote Kauwenberg 18, 2000 Antwerp, Belgium
bn206@cam.ac.uk

Prof. Peter Millican
Hertford College, Oxford
Catte Street, Oxford OX1 3BW
peter.millican@hertford.ox.ac.uk

Prof. Stephen Laurence (PhD examiner)
Department of Philosophy, University of Sheffield
45 Victoria Street, Sheffield S3 7QB
s.laurence@sheffield.ac.uk

Dr. Stephen Butterfill (PhD examiner)
Department of Philosophy, University of Warwick
Coventry CV4 7AL
s.butterfill@warwick.ac.uk