HENRY SHEVLIN

Associate Director, Leverhulme Centre for the Future of Intelligence, University of Cambridge
Tel.: +44 7974 274939 • Email: henry.shevlin@gmail.com • Website: www.henryshevlin.com

EMPLOYMENT

Associate Director (Education), Leverhulme Centre for the Future of Intelligence, University of Cambridge (May 2023-)

Programme Director, Kinds of Intelligence, Leverhulme Centre for the Future of Intelligence, University of Cambridge (April 2023-)

Course Co-Lead, AI Ethics & Society MSt., Leverhulme Centre for the Future of Intelligence, University of Cambridge (January 2021- April 2023)

Senior Researcher, Leverhulme Centre for the Future of Intelligence, University of Cambridge (January 2021- March 2023)

Researcher, Leverhulme Centre for the Future of Intelligence, University of Cambridge (May 2017- January 2021)

EDUCATION

Ph.D, Philosophy, CUNY Graduate Center, September 2016 (Honors)
Thesis: Consciousness, Perception, and Short-Term Memory (Supervisor: Peter Godfrey-Smith) Committee: Ned Block, Eric Mandelbaum, David Papineau, Jesse Prinz.

BPhil, Philosophy, Trinity College, University of Oxford, June 2009.
Thesis: Linking Phenomenal and Access Consciousness (Supervisor: Tim Bayne)

BA Classics, St Anne's College, University of Oxford, June 2006 (Double First).

AREAS OF SPECIALIZATION

AI Ethics • Animal Cognition • Philosophy of Cognitive Science • Philosophy of Mind

AREAS OF COMPETENCE

Moral and Political Philosophy • Philosophy of Science • Epistemology • Metaphysics

ACADEMIC PUBLICATIONS

1. Shevlin. "Relating to Machines." *Oxford University Press Handbook of Generative AI* (Eds. Andreas Engel, Sarah Hammer, Philipp Hacker).

2. Shevlin (2024) "All too human? Identifying and mitigating ethical risks of Social AI." *Law, Ethics, & Technology.*

3. Shevlin (2024). "Imagination, Creativity, and Non-human animals." *Oxford University Press Handbook of Philosophy of Imagination and Creativity* (eds. Amy Kind & Julia Langkau)

4. Shevlin (2024). "Consciousness, Machines, and Moral Status." *Humans and Smart Machines as Partners in Thought* (ed. Anna Strasser).

5. Srivastava et al. (including Shevlin) (2024). "Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models." *Transactions on Machine Learning Research.*

6. Vervoort, Shevlin, Melnikov, & Alodjants (2022). "Deep Learning Applied to Scientific Discovery: A Hot Interface with Philosophy of Science." *Journal for General Philosophy of Science.*

7. Shevlin (2021). "Rethinking creative intelligence: comparative psychology and the concept of creativity." *European Journal for Philosophy of Science.*

8. Shevlin (2021). "General intelligence: an ecumenical heuristic for artificial consciousness research?"

Journal of Artificial Intelligence & Consciousness.

9. Shevlin (2021). "Non-human consciousness and the specificity problem." *Mind & Language.*

10. Shevlin (2021). "How could we know when a robot was a moral patient?" *Cambridge Quarterly of Healthcare Ethics.*

11. Shevlin (2020). "Which animals matter? Comparing psychological approaches to moral status in non-human systems." *Philosophical Topics.*

12. Shevlin (2020). "Current controversies in the cognitive science of short-term memory." *Current Opinion in Behavioral Sciences.*

13. Crosby & Shevlin (2020). "Defining Artificial Intelligence: a reply to Wang." *Journal of Artificial General Intelligence.*

14. Shevlin & Friesen (2020). "Pain, Placebo, and Cognitive Penetration". *Mind & Language.*

15. "Qualia and 'raw feels." (2019). In *Introduction to Philosophy of Mind*, H. Salazar (ed.).

16. Shevlin, Vold, Crosby, & Halina (2019). "The Limits of Machine Intelligence." *EMBO Reports.*

17. Shevlin & Halina (2019). "Apply rich psychological terms in AI with care". *Nature Machine Intelligence*, 1(4), 165–167.

18. Bhatnagar et al. including Shevlin (2018). "Mapping Intelligence: Requirements and Possibilities". In Muller, V. C., ed., *Philosophy and Theory of Artificial Intelligence* 2017.

19. Shevlin (2017). "Conceptual Short-Term Memory: A Missing Part of the Mind?" *Journal of Consciousness Studies*, 24(7–8), 163–188.

20. Shevlin (2017). "The Lower Bounds of Desire". *Journal of Consciousness Studies*, 24(5–6), 251–258.

SELECT PUBLIC WRITING

"What if Artificial Intelligence saves the planet?" *The New European* (cover article)

"The artist is dead, AI killed them" (2022). *iai.tv.*

"GPT-3: A Digital Remix of Humanity" (2020). *Daily Nous.*

"AI reflections in 2019" (2020). *Nature Machine Intelligence.*

"A Lack of Understanding: Storytelling for Robots" (2019). *Litro Magazine.*

"Brutality Is Common in Video Games, but Not Sexual Violence. Why?" (2018). *Aeon.*

GRANTS, PRIZES, & FELLOWSHIPS

2022: Personal gift of $50,000 by Google for project "LLMs, Empathy & Mentalising".

2022: MSt in AI Ethics & Society awarded 'Best Course in AI' at CogX Awards.

2017: Winner, President's Excellence Award for Distinguished Teaching, Baruch College.

2016: Joint winner, Fifth Annual Essay Prize at the Centre for Philosophical Psychology.

2015: Joint winner, University of Oxford Future of Humanity Institute Thesis Prize.

2015 – 2016: Dissertation Fellowship, CUNY Science Studies Committee

2015: Joint winner, University of Oxford Future of Humanity Institute Thesis Prize

2014: CUNY Doctoral Students Research Grant

2010 – 2015: Enhanced Chancellor's Fellowship & Mellon Scholar, CUNY Graduate Center

2007 – 2010: Cecil Lubbock & Graduate Scholarship, Trinity College, Oxford

ADMINISTRATION & LEADERSHIP

Unit Editor (2024-26), OUP Intersections AI & Society: AI & Relationships

Guest editor for *Philosophical Topics,* Spring 2020. "Ascriptions of consciousness."

Editorial board member, International Journal of AI and Consciousness.

Lead organiser of April 2022 Ascriptions of Consciousness Workshop, Cambridge.

Joint lead organiser of 2021-22 Kinds of Intelligence Workshop Series, Cambridge.

Joint lead organiser of four-day conference "Kinds of Intelligence" (June 2019), Cambridge.

Lead organiser four-day conference "Varieties of Mind" (June 2018), Cambridge.

Joint lead organiser of workshop "AI and Games" (December 2017), Cambridge.

Founder and convener of the Cambridge Kinds of Intelligence reading group (homepage).

Student Committee Member (2015-2017) and Student Committee Chair (2017-2018), ASSC.

TEACHING EXPERIENCE

Course Lead, AI Ethics & Society MSt

Designed and delivered multiple content modules and intensive teaching weeks, including:

- **Introduction to AI Ethics**
- **Cognitive Science & AI**
- **Anthropomorphism in AI**
- **Ethics of Large Language Models**
- **Ethics of Generative Media**
- **AI Safety and Existential Risk**
- **Ethics of Human-AI Relationships**
- **Bias, Justice, and Fairness in AI ethics**
- **Privacy, Liberty, and Autonomy in AI ethics**
- **Robots Rights & Artificial Moral Status**
- **Well-being and AI**

Master's Thesis Supervisions (University of Cambridge)

- **AI Ethics & Society MSt.**

  Connor Wright (2024), Mark Robinson (2024), Hasan Iqbal (2024), Kimberley Malfacini (2024), Alva Markelius (2024), Barry Dynkin (2024), François Buet-Golfouse (2023), Peter Douglas (2023), Claire Dugan (2023), Stephanie Hughes-Fitt (2023), Matthias Klaus (2023) Matthew Tutty (2023)

- **Ethics of AI, Data, and Algorithms Dissertations**

  Eleos Citrini (2024), Joe Fennell (2024), Daisy Coburn (2024), Dvija Mehta (2024).

- **History & Philosophy of Science MPhil Dissertations**

  Edward Potts (2022), Shruti Santosh (2020), Alice Hayler (2019), Angela Madeira (2018).

Cambridge University Undergraduate Supervisions

- Ethics and Political Philosophy Part 1A/1B/2
- Metaphysics Part 1A, Metaphysics and Epistemology Part 1B, Knowledge, Language & World 1B
- Philosophy of Science Part 2, Philosophy of Mind Part 2

Teaching Adjunct/Adjunct Assistant Professor, Baruch College, 2011-2016.

- Science Fiction and Philosophy (PHI3900), Fall 2016
- Ethics and Critical Thinking (PHI 1100), Fall 2015, Spring 2016, Fall 2016
- Philosophy of Sex and Love (PHI 4900), Summer 2015
- Philosophy of Psychology (PHI/PSY 3035), Spring 2015
- Thought and Reality (PHI/PSY 3030), Fall 2012, Fall 2013, Fall 2014.
- Minds and Computers (PHI 3040), Summer 2014, Summer 2013
- Experimental Philosophy (PHI 3220), Fall 2012, Spring 2014
- Computer Ethics (CIS/PHI 3270), Summer 2012
- Major Issues in Philosophy (PHI 1500), Fall 2011/2012/2013/2014, Spring 2012/2013/2014

Other Teaching Experience

- Writing Fellow at the School for Continuing and Professional Studies, CUNY (2014-2015).
- ESL teacher, Oxford School of English, Oxford (2007-2010).
- ESL Teacher, International House, Milan (2007)

PRESENTATIONS

(*=invited)

- *October 2023 – "Prospects & Pitfalls for a Science of Artificial Consciousness." University Centre for Human Values, Princeton University.
- *May 2023 – "LLMs, Social AI, and folk attributions of consciousness." University of California, Riverside
- *June 2022 – "Uncanny Believers: Chatbots, Beliefs, and Folk Psychology." Invited speaker at LMU Munich event, "AI in Science: Foundations and Applications."
- *May 2022 – "Uncanny Communicators and the Rise of Social AI." Invited speaker at Google UK Responsible AI Group.
- November 2021 – "Negative valence, felt unpleasantness, and animal welfare." Symposium Panellist at Philosophy of Science Association, Baltimore.
- *July 2021 – Invited Discussant at Engineering and reverse-engineering morality Workshop, CogSci 2021 (Virtual).
- June 2021 – "Non-human consciousness: current paradigms and debates." Association for the Scientific Study of Consciousness, Tel Aviv.
- *June 2021 – "Three approaches to theorizing artificial consciousness." School of Advanced Studies, University of Tyumen.
- *November 2020 – "AI, Philosophy, and Cognitive Science in the 21st Century." Universität Konstanz.
- *October 2020 – "Comparing psychological approaches to moral status." Queen Mary University London,.
- *August 2020 – "GPT-3: Technical and Policy Implications." Tandem Research Community, Goa.
- *August 2020 – "GPT-3 and Philosophy." Foresight Institute, San Francisco.
- *February 2020 – "General intelligence: an ecumenical heuristic for artificial consciousness research?" Science of Intelligence Research Cluster, Berlin.
- *October 2019 – "Learning From Nature to Assess Our Intelligence Needs." Minds & Tech Conference, Toulouse.

- *September 2019 – "Pathways to Artificial Moral Patiency." Novel Beings Workshop, Newcastle University.
- September 2019 – "What can theories of consciousness tell us about non-human subjects?" European Society for Philosophy and Psychology, Athens.
- *August 2019 – "Theories of consciousness and animal minds." Philosophy of Biology at Dolphin Beach.
- *June 2019 – "Varieties of creative intelligence." Creative Intelligence Workshop, University of Cambridge.
- *April 2019 – "Technology and the Future of Love." The Ontology of Love, University of Cambridge.
- *March 2019 – "Animals, AI, and non-human consciousness." Artificial Intelligence & the Posthumanities, Royal Holloway, University of London.
- March 2019 – "Consciousness, biology, and general intelligence". AAAI Spring Symposium 19, Stanford University.
- *October 2018 – "From Dolphins to DQNs: Intelligence, Consciousness, and Moral Status." Company Law and Novel Beings Workshop, Wellcome Collection.
- *July 2018 – "Kinds of Intelligence and the Ethics of AI." CDT Science Festival on Science & Ethics, Imperial College London.
- July 2018 – "Consciousness, Sentience, and Moral Status". BSPS Symposium on Foundations of Animal Sentience, BSPS 2018, University of Oxford.
- *April 2018 – "Is object categorisation a perceptual or cognitive process?" Sainsbury Laboratory, University of Cambridge.
- *May 2017 – "Conceptual Short-Term Memory: an interface for perception and cognition". Fifth Annual Conference on Advances in Cognitive Systems.
- *November 2015 – "CSTM: A Missing Piece in the Puzzle of Consciousness." Navy Center for Applied Research in Artificial Intelligence.
- July 2015 – "Conceptual Short-Term Memory and the Richness of Experience". 19th Meeting of the Association for the Scientific Study of Consciousness.
- July 2013 – "Generic Phenomenology and Partial Report Paradigms." 17th Meeting of the Association for the Scientific Study of Consciousness.
- May 2013 – "Do We Experience Color and Shape Generically?" 9th International Symposium of Cognition, Logic and Communication.
- December 2009 – "Two Interpretations of Partial Report Paradigms" Speaker, University of Warwick Mindgrad Conference.

LANGUAGES

Latin, Ancient Greek (Proficient) French, Italian (Intermediate)
Japanese, Tagalog (Upper Elementary)

REFERENCES

Marta Halina (Former Line Manager). University Associate Professor in the Philosophy of Cognitive Science, University of Cambridge, Selwyn College. Email: mh801@cam.ac.uk

Peter Godfrey-Smith (Dissertation Supervisor). Professor, School of History and Philosophy of Science, University of Sydney. Email: p.godfrey.smith@gmail.com