

Amber Ross

The University of Florida
Philosophy Department
330 Griffin-Floyd Hall
Gainesville, FL 32611-8545

amber.ross@ufl.edu
647-784-2511

ACADEMIC POSITIONS HELD

Assistant Professor	2018 – pres.
University of Florida, Philosophy Department	
Post-Doctoral Research Fellow, The Network for Sensory Research	2015 - 2018
University of Toronto, Philosophy Department	
Lecturer	2014 - 2015
University of Toronto, Philosophy Department	
Research Assistant, The Network for Sensory Research	2013 - 2014
University of Toronto, Philosophy Department	
Research Fellow, The Center for Cognitive Studies	2010 - 2012
Tufts University	

EDUCATION

Ph.D. University of North Carolina, Chapel Hill (Philosophy)	Spring, 2013
Dissertation: <i>Inconceivable Minds</i>	
M.A. Tufts University (Philosophy)	Spring, 2006
B.A. Texas Christian University (Philosophy)	Spring, 2000

AREAS OF SPECIALIZATION

AI Ethics, Philosophy of Mind

AREAS OF COMPETENCE

Metaphysics, Epistemology, Social Epistemology, Philosophy of Language

RESEARCH INTERESTS

Consciousness, Cognitive Science, AI, Non-human Self Consciousness

PUBLICATIONS

Submitted; R&R	Ross, A., AI and the expert; a blueprint for the ethical use of opaque AI (<i>R&R for AI and Society</i> , resubmitted May 2022)
2022	Ross, A., Mental Fictionalism: the costly combination of magic and the mind. In <i>Mental Fictionalism: Philosophical Explorations</i> . Eds. Tomas Demeter, T. Parent, & Adam Toon. Routledge https://philpapers.org/archive/ROSMFT.pdf

- 2021 Ross, A. & Matthen, M., Multisensory Perception in Philosophy: Editors' Introduction. *Multisensory Research*. 34. <https://philpapers.org/archive/ROSMPI-4.pdf>
- 2018 Ross, A., Destabilizing the Knowledge Argument and Modal Argument. *Inquiry: An Interdisciplinary Journal of Philosophy* <http://dx.doi.org/10.1080/0020174X.2017.1385530>
- 2016 Ross, A., Illusionism and the Epistemological Problems Facing Phenomenal Realism. *Journal of Consciousness Studies* 23 (11/12): 215-23.

WORKS IN PROGRESS

1. AI and the expert; a blueprint for the ethical use of opaque AI
(R&R for *AI and Society*, re-submitted May 2022)

The increasing demand for transparency in AI has recently come under scrutiny. The question is often posted in terms of “epistemic double standards”, and whether the standards for transparency in AI ought to be higher than, or equivalent to, our standards for ordinary human reasoners. I agree that the push for increased transparency in AI deserves closer examination, and that comparing these standards to our standards of transparency for *other* opaque systems is an appropriate starting point. I suggest that a more fruitful exploration of this question will involve a *different comparison class*. We routinely treat judgments made by *highly-trained experts in specialized fields* as fair or well-grounded even though—by the nature of expert/layperson division of epistemic labor—a expert will not be able to provide an explanation of the reasoning behind these judgments that makes sense to *most other people*. Regardless, laypeople are thought to be acting reasonably—and ethically—in deferring to the judgment of experts that concern those experts’ area of specialization. I suggest that we reframe our question regarding the appropriate standards of transparency in AI as one that asks when, why, and to what degree it would be ethical to accept *opacity* in AI. I argue that our epistemic relation to certain opaque AI models may be relevantly similar to the layperson’s epistemic relation to the expert, such that the successful expert/layperson division of epistemic labor can serve as a blueprint for the ethical use of opaque AI.

2. New complications from Continuous Machine Learning (or, why opaque AI will not go away)
(In prep)

As the literature on transparency, interpretability, and explainable AI has shown, there will always be some degree of opacity present in any *especially powerful* AI model. These are models for which machine learning was essential to achieving their level of performance. Especially powerful AI models have the capacity to operate over vast parameters, at speed and complexity of data processing methods that—in practice—out-perform any models that do not employ machine learning. XAI has not viewed this lingering opacity as an insurmountable barrier to creating satisfactory explanations of opaque deep learning models: it is assumed that, once a model’s parameters are set, we can develop explanations of those models that render them transparent enough *before* the models are deployed, such that we can claim to understand how the model works before it is in use. But many AI labs are now developing CML (continuous machine learning) models, which will continue to adjust their parameters, or learn, *while* the models are deployed. With CML, there is no sharp divide between the training stage and deployment stage. Because CML models *never* settle on a fixed set of parameters, they are inherently opaque. If CML models provide serious advantages over standard train-then-deploy AI models, we will need to decide whether and how we can ethically use inherently opaque AI models.

WORKS IN PROGRESS (CONT.)

3. Animal Selves- locating a middle ground in the notion of *self*

The most minimal type of selfhood is arguably the feature that allows an animal to engage in fundamental self-preservation, such as not eating its own body when hungry. According to some philosophers (see Dennett 1991) there is a difference in kind between this type of minimal self and the kind of selfhood that humans possess. On the view I propose, the self does not suddenly immerse with the development of language in human beings. Instead, I suggest that there are degrees of selfhood that fill in the gap between the most minimal self and the self-conscious, self-reflective, selves of human beings. To support this view, I formulate criteria for a level of self-hood and self-consciousness that lies in the middle ground: a Minimally Self-Conscious Self. In brief, a Minimally Self-Conscious Self is one that can form judgments about its own skills in a particular context (e.g., that branch is too far away for me to jump to it), but does not form generalizations about their skills that would apply regardless of context. The fact that these creatures make judgments about themselves makes them self-conscious (or self-aware) to a certain extent, but since their judgments about themselves exist only in the contexts of individual discrete events, their form of self-consciousness is impoverished compared to those who create persisting stories of themselves. It is this real-but-limited self-awareness of a minimally self-conscious self that shows us there is a place for degrees of self-hood that lie between the merely self-preserving self and the full-fledged selfhood that we possess.

SELECTED PRESENTATIONS

- “Narrative selfhood in non-human animals” Apr, 2023
To be presented at Logos Research Group, University of Barcelona
- “Ethical Opaque AI; a blueprint” Sept, 2022
To be presented at the [American Political Science Association annual meeting](#), Montreal, Canada
- “From Spot to Hal; how the study of self-consciousness in animal minds can inform the philosophy of AI” Jun, 2022
Plenary Speaker, [Mind and Matter 2022: Foundation of Information, Intelligence, and Consciousness](#). University of Helsinki, Finland
- “Why opaque AI will not go away; a case for ethical AI without complete transparency” Dec, 2021
Philosophies of Nature, Technology and Artificial Intelligence Workshop, hosted by the journal *Ethics and the Environment*, University of Georgia
- “The nature of conscious thought and its relation to AI systems” Mar, 2021
Guest lecture, UF graduate course in Machine Learning, Department of Engineering (CAP6610)
- “What is Thinking?” Jan, 2021
Invited presentation for the UF *AI Institute* Faculty Seminar
- “Is Fictionalism Fit to Account for the Mental?” Oct, 2019
Budapest- *Mental Fictionalism Workshop Series (III)*- Organized by the MTA BTK Lendulet Morals and Science Research Group, Hungarian Academy of Sciences.
- “How Mary closed the explanatory gap” Mar, 2014
Invited Speaker, *The Moore Humanities Symposium on Consciousness*, TCU

PROJECTS AND GRANTS

Participant (Knowledge Area: Sociological and Technological changes Affecting Individuals, Families, and Communities) in the Hatch Multistate Project S1090: AI in Agroecosystems: Big Data and Smart Technology-Driven Sustainable Production	2022-2026
Co-PI, Challenge Grant proposal, “Educating K-12 on Bias and AI (in partnership with the Penn State Dept. of Engineering Design, Technology, and Professional Programs); the Public Interest Technology University Network (application pending)	2023-2024

FELLOWSHIPS AND AWARDS

UF Summer Research Humanities Enhancement Grant for monograph, <i>All of “Us”</i> ; a unified account of the varieties of self	Summer, 2020
Post-Doctoral Fellowship (SSHRC) <i>The Network for Sensory Research</i> , University of Toronto	2015 - 2018
Research Fellowship: Tufts University, The Center for Cognitive Studies	2010 - 2012
Horace Williams Fellowship: UNC Department of Philosophy	2004 - 2010
Visiting Scholar Fellowship: University of Reading, UK	2007 - 2008
Graduate Student Opportunity Fund Grant, UNC- Chapel Hill	2007 - 2008
Scholars of Tomorrow Fellowship: UNC- Chapel Hill	2004 - 2005

SERVICE TO THE DEPARTMENT

UF Online Professional Development Course Creation, AI Ethics	Summer 2020
Committee to develop UF Graduate Certificate in Ethics, Technology, and AI	2018 – pres.
UF Online Course Development, Ethics and Technology Unit, <i>Contemporary Moral Issues</i>	Fall, 2019
Graduate Recruitment Officer, Philosophy Department	2019 – pres.
Hiring Committee, Philosophy Department, University of Florida	
Tenure track position in Philosophy of Science	2018 - 2019
Tenure track position in AI Ethics	2020 - 2021

SERVICE TO THE UNIVERSITY

Ethicist, UF <i>Artificial Intelligence Academic Initiative Center</i>	2022-
Ethicist, S1090: AI in Agroecosystems: Big Data and Smart Technology-Driven Sustainable Production, a HATCH Multistate Project	2022-2026
“Practical AI Ethics- a guide to identifying application-specific ethical issues in AI”	Aug, 2022
Session for the 2022 Annual Meeting of the HATCH S1090 AI Multistate Project	
Senior Project Personnel, NSF Grant (proposal) UF <i>AI Institute: iFIRST: Institute for Foodchain Intelligence, Resilience, Sustainability and Traceability</i>	2020 & 2021

SERVICE TO THE UNIVERSITY (CONT)

Data Ethics Consultant, UF-FSU CTSA Hub NIH Research project, *DataStory: Using Story-Driven Strategies to Teach Data Science in Precision Health Settings.* 2020 - 2021
Faculty Advisor, Secular Student Association 2019 – pres.
Faculty Advisor, Acts of Random Kindness Project 2021 – pres.

SERVICE TO THE PROFESSION

AI Ethics panelist for the [Association of American Universities](#) Humanities Task Force Discussion on Ethics and Artificial Intelligence Jul, 2022
Committee Member, National Humanities Center multi-year project in AI and the Humanities, Ethics and AI (programs delayed due to Covid 19) Fall 2019
Co-Editor (with Mohan Matthen), Special Issue of *Journal of Multisensory Perception*, Vol. 34, Issue 3. <https://doi.org/10.1163/22134808-034001ED> March, 2021
Commentator, Central APA Feb, 2021
Commentator, Southern Society of Philosophy and Psychology Spring, 2015
Referee:
Journal of Consciousness Studies *Inquiry; an Interdisciplinary Journal of Philosophy* 2013-pres.
Journal of Philosophical Research *Journal of the American Philosophical Association*
South African Journal of Philosophy *Multisensory Research*

WORKSHOP ORGANIZATION

AI Ethics in Legal Discovery (in partnership with UF Law) Nov, 2022
[University of Toronto Workshop on Non-Human Cognition](#) May, 2018
[University of Toronto Workshop on Metacognition](#) May, 2016
[University of Toronto Workshop on Multisensory Integration](#) May, 2014

SERVICE TO THE COMMUNITY

AI Ethics Discussion Leader, TEDx UF Feb, 2020
UNC Philosophy Outreach Program:
Ethics for High School Students, Saxapawhaw Charter School 2007-2010
Philosophy and Ethics Group at North Carolina Juvenile Detention Center

UNIVERSITY TEACHING EXPERIENCE

GRADUATE COURSES TAUGHT

Philosophy of Mind: Seminar on Personal Identity and The Self University of Florida
Philosophy of Mind: Seminar on Contemporary issues in Consciousness University of Florida

UNDERGRADUATE COURSES TAUGHT

Ethics, Data, and Technology 3rd year course, University of Florida

AI Ethics and Emerging Technologies	3 rd year course, University of Florida
Introduction to Philosophy	Large Lecture, Gen-Ed, University of Florida
Metaphysics; an introduction	2 nd year course, University of Florida
Metaphysics: advanced undergraduate	3 rd year course, University of Toronto
The Philosophy of Animal Minds	Senior seminar, University of Toronto & UF
Minds, Bodies, and Persons	2 nd year course University of Toronto
Philosophy of Language	3 rd year course, University of Toronto
Probability and Inductive Logic	2 nd year course, University of Toronto
Experience and Reality	UNC-Chapel Hill
Introduction to Ethics	UNC-Chapel Hill
Philosophy of Religion	UNC-Chapel Hill
Existentialism	UNC-Chapel Hill

ONLINE COURSES CREATED OR TAUGHT

Ethics of AI , a foundational course for UF AI Micro-Credential	UF Professional Development Course
AI Ethics unit in “Contemporary Moral Issues”	2 nd year course, University of Florida
Bioethics	UNC-Chapel Hill

DISSERTATION ABSTRACT

INCONCEIVABLE MINDS

The current debate over the metaphysical nature of the mind is dominated by two major philosophical views: property dualism and physicalism. According to property dualism, mental properties are of two metaphysically distinct types. There are “phenomenal properties”, or “qualia”, the properties that constitute conscious experience, and there are the cognitive or functional properties of the mind. According to physicalism, there is one metaphysical type of mental property, though there may be a deep conceptual divide between experience and cognition. I challenge both the dualist and physicalist assumptions.

Focusing on two of the most popular anti-physicalist arguments- the Conceivability Argument and the Knowledge Argument- I argue that the property dualist’s account of the relationship between consciousness and behavior- the “explanatory irrelevance” of consciousness to sufficient explanations of our behavior and how we make judgment about our conscious experience- make the first argument untenable and the second irrelevant to the metaphysical debate.

I also present a case against the typical physicalist conceptual separation between cognition and consciousness by challenging the “Phenomenal Concept Strategy”, the most popular argument supporting such a separation. This conceptual separation supposedly allows us to conceive of “philosophical zombies,” creature physically identical to human beings but who lack conscious experience. The phenomenal concept strategy aims to explain how we can conceive of zombies while maintaining a physicalist account of the metaphysics of mind. For this strategy to succeed, the physicalist must show that we share our epistemic situation regarding consciousness with our “zombie-twins”. Zombies make claims about their own phenomenal experience, just as we do, but by definition they have none. I examine the most common physicalist interpretation of the zombies’ beliefs about their own conscious experiences and show that this leads to the creation of *inconceivable minds*- creatures whose mental features would be incompatible with the very interpretation of zombie “phenomenal” belief on which this strategy is based.

My dissertation has two overarching goals. First, to undermine the plausibility of the two most popular arguments for property dualism, and second, to encourage physicalists to reconsider both the phenomenal concept strategy and their

commitment to the genuine conceivability of philosophical zombies and, by extension, the intuitive appeal of property dualism in general.

REFERENCES

Daniel C. Dennett
Tufts University, Philosophy Department & The
Center for Cognitive Studies
daniel.dennett@tufts.edu

Mohan Matthen
Senior Canada Research Chair & Principal
Investigator, The Network for Sensory Research
University of Toronto, Philosophy Department
mohan.matthen@utoronto.ca

Ram Neta
UNC-Chapel Hill, Philosophy Department
neta@email.unc.edu

William G. Lycan
University of Connecticut, Philosophy
Department
william.lycan@uconn.edu

Bryce Huebner
Georgetown University, Philosophy
Department
lbh24@georgetown.edu