Dr. Joanna Bryson
Biography

Joanna Bryson is a Reader (tenured Associate Professor) at the University of Bath, United Kingdom, and an affiliate of Princeton University's Center for Information Technology Policy (CITP). Her academic interests include the structure and utility of intelligence, both natural and artificial. Venues for her research range from "reddit" to "Science".

She is best known for her work on AI systems and AI ethics, both of which she began during her doctoral work in the 1990s, but she and her colleagues publish broadly – in biology, anthropology, sociology, philosophy, cognitive science, and politics. Current projects include "Public Goods and Artificial Intelligence" with Alin Coman of Princeton University's Department of Psychology and Mark Riedl of Georgia Tech. The project is funded by Princeton University's Center for Human Values and includes both basic research in human sociality and experiments in technological interventions. Other current research projects are centered around understanding the causality behind the correlation between wealth inequality and political polarization, generating transparency for AI systems, and conducting research on machine prejudice deriving from human semantics.

Bryson holds degrees in psychology from the University of Chicago and the University of Edinburgh, and in artificial intelligence from the University of Edinburgh and the Massachusetts Institute of Technology (MIT). At Bath, she founded the Intelligent Systems research group (one of four in the Department of Computer Science) and heads their Artificial Models of Natural Intelligence.

Universität Bielefeld

CITEC

VISITING PROFESSOR FOR GENDER STUDIES
WINTER TERM 2017

Visiting Professor for Gender Studies at Bielefeld University

The interdisciplinary Visiting Professorship for Gender Studies strengthens gender-specific content in the research and teaching activities in Bielefeld University's faculties and institutes. This professorship aims at embedding and further expanding gender-related knowledge in the individual disciplines, and in research and teaching more generally. The Visiting Professor thus advances the goals of structurally strengthening gender research while also stimulating interdisciplinary exchange at Bielefeld University. To these ends, the Interdisciplinary Center for Gender Research (IZG) and the Master's degree programme in Gender Studies have already been successfully implemented. The Professorship is also integrated into the Rektorat's strategic plan to strengthen equal opportunity as well as gender and diversity issues within Bielefeld University. During the Winter term of 2017, Dr. Joanna Bryson will hold the Visiting Professor at the Cluster of Excellence Cognitive Interaction Technology (CITEC), contributing her expertise into the field of gender and cognitive interaction technology. During her stay at CITEC, she will give talks and also hold a seminar for Master's and PhD students.

More information is available online at:
www.cit-ec.de/en/gender-diversity
www.uni-bielefeld.de/gender/gendergastprofessur.html

Dr. Joanna Bryson
University of Bath | United Kingdom

TALK 1: WE DIDN'T PROVE PREJUDICE IS TRUE: WHY AND WHEN MACHINES HAVE HUMAN BIAS

Date: **Thursday, 16 November 2017**
Time: **10:00–12:00**
Location: **CITEC, Room 1.204**

Machine learning is a means to derive artificial intelligence by uncovering patterns in existing data. In 2017, Dr. Joanna Bryson and two colleagues demonstrated that applying machine learning to ordinary human language results in human-like semantic biases. They replicated a spectrum of known biases, as measured by the Implicit Association Test, applying a widely used, purely statistical machine-learning model trained on a standard corpus of texts from the Internet. Their results indicate that text corpora contain recoverable and accurate imprints of our historic biases, whether morally neutral, such as toward insects or flowers; problematic as toward race or gender; or even simply veridical – reflecting the status quo distribution of gender with respect to careers or first names. In her talk, Bryson will first present their results, and then discuss what their research on machine bias demonstrates concerning the origins of human biases, stereotypes, and prejudices.

TALK 2: WHY AI ETHICS IS A FEMINIST ISSUE: THE LEGAL AND MORAL LACUNA OF MACHINE RIGHTS

Date: **Wednesday, 22 November 2017**
Time: **10:00–12:00**
Location: **CITEC, Room 1.204**

Conferring legal personhood to purely synthetic entities is a very real legal possibility – one that is, in fact, currently under consideration by the European Union. Why do people assert that machines may need their own (rather than derivative) rights? In what sense can anything other than a human be a legal person, or a moral agent? Does extending ethical concerns to other entities ever diminish the ethical concern we have for humans, or certain categories of humans?

In this talk, Bryson will begin with a set of simple functionalist definitions for the following terms: intelligent, agent, moral agent, moral patient, ethics and legal person. In most cases, the definitions are not the "correct" usages of the terms, but rather present a set of concrete concepts that can be used to disentangle frequently confused conceptions of ethics and identity. She will argue that since both machines and ethics are cultural artifacts, there is no scientific fact about the moral standing of machines that needs to be discovered, but rather only normative recommendations that need to be made.

BLOCK SEMINAR: USING THE TOOLS OF THEORETICAL BIOLOGY AND AI SOCIAL SIMULATION TO UNDERSTAND HUMAN BEHAVIOUR

Date: **Monday and Tuesday, 20–21 November**
Time: **10:00–16:00**
Monday Location: **CITEC, Room 1.015**
Tuesday Location: **CITEC, Room 2.015**

Open to Master's students and PhD researchers
While science cannot tell us how we should be, it can help us to understand how things are, and what the likely consequences of our policies will be. In a series of lectures, Bryson will first describe and then demonstrate how simulation can be used to contribute to science, and how science can help to understand the tradeoffs of sociality, and its alternatives. She will illustrate the technique and make recommendations on policy for intelligent technology. Code for most of the simulations will be made available to students using NetLogo, the standard agent-based modelling platform which is available for most operating systems.

There will be four two-hour lecture blocks of 60–90 minutes, including 30–60 minutes of Questions and Answers:

- **Some Research in and Methods for Using Artificial Intelligence**
Topics will include modelling learning in animals, modelling the evolution of religions, making robot and game AI transparent to users, building synthetic emotions, and understanding consciousness.

- **Primate Social Organisation**
In authoritarian societies, troop hierarchy is clear, and conflict, while rare, is violent. In egalitarian societies, most hierarchy is less clear, and conflict is far more frequent but often bidirectional, meaning subordinates challenge superiors. Joanna Bryson will use models to explore a number of theories for explaining these differences and will also look for an explanation as to why some primates have matriarchal social orders, but most have patriarchal.

- **Why Information Can Be Free: Culture, Cooperation, and Trust**
In this seminar, Bryson will explore models that explain why and how cooperation is as ubiquitous as conflict in nature, and then review theories for why only humans have linguistic capabilities. She will also introduce recent models of trust and identity.

- **Cultural Variation in Public Goods Investment and Political Polarisation**
Bryson will discuss both published and in-progress research explaining anti-social punishment, and will then look at additional theories for explaining why political polarisation correlates with income inequality.

References:

Caliskan, Aylin, et al. "Semantics derived automatically from language corpora contain human-like biases." Science 356.6334 (2017): 183-186.

Bryson, Joanna J., et al. "Of, for, and by the people: the legal lacuna of synthetic persons." Artificial Intelligence and Law (2017): 1-19.

Bryson, Joanna J., et al. "Understanding and addressing cultural variation in costly antisocial punishment." In Applied Evolutionary Anthropology, 201-222. Springer New York, 2014.

Bryson, Joanna J., et al. "Agent-based modelling as scientific method: a case study analysing primate social behaviour." Philosophical Transactions of the Royal Society of London B: Biological Sciences 362.1485 (2007): 1685-1699.