



British Wittgenstein Society Conference

Keynote Abstracts

Wittgenstein and AI



Wittgenstein and Assertion

A.C. Grayling (New College of the Humanities)
29 July 9:00

If it is right to see Wittgenstein as pivoting to a deflationary view of truth in the *Investigations*, what account of (the language-game of) assertion might be available to him given the constraints implied by his views in the later philosophy and the widely-accepted standard account of assertion as essentially truth-involving in a way dependent on truth's being substantive or normative or both? Considerable discussion has been devoted to Wittgenstein's reactions to Frege and Russell on assertion in the *Tractatus* and *Investigations*, and some cues can be taken from it; here I propose that a version of a theory of assertion not dependent on a realist conception of truth, and consistent with a minimalist view, can be adapted to fit with *Investigation*-constrained implications for sense, force and point as these bear on the assertion language-game. The point of the exercise is not to claim an articulation or reconstruction of Wittgenstein's view of assertion in the later philosophy, so much as to show that a certain understanding of assertion is more than merely consistent with, but is arguably required by, the *Investigation* constraints.

Wittgenstein, Turing and AI
Juliet Floyd (Boston)
30 July 9:00

Turing drew from Wittgenstein's 1939 Cambridge lectures the idea that everyday typings of concepts, our evolving "phraseology", plays a fundamental role in the application of logic. After going to Bletchley Park, Turing continued to think about the importance of notations, and in "The Reform of Mathematical Notation" (1944/45) he suggested that symbolic logic open itself up to a plurality of systems, attending to the specific uses to which notations are put, and Turing argued that we should take into account everyday language when constructing logical notations. This Wittgensteinian aspect of Turing's philosophy of logic culminated in his 1948 report to the National Physical Laboratory, "Intelligent Machinery", the founding document of AI. Here Turing envisioned, with great prescience, a future with machinery that would involve, not only the search for proof systems and notations, but also the use of "intelligent machinery" in biological searching and, in the end, what Turing called "the cultural search", a search involving humanity as a whole (and no machines). "Intelligence" Turing defined, tentatively, as the capacity to appreciate the importance of different kinds of searching – an echo of an idea to be found in *Philosophical Investigations*. Later in his "Computing Machinery and Intelligence" (1950) Turing devised the "imitation game", a "Turing Test" that has been much discussed in philosophy of mind and popular culture. Too often this Test has been wrongly conceived as furthering mechanism or behaviorism about the notions of thought and mentality: it is read in a dualist, Cartesian vein. On this (mis)reading, the social dimensions of Turing's game have been ignored. However, the Turing Test is actually a social experiment in "phraseology", a quest to elicit criteria from us – in brief, a language-game in Wittgenstein's sense. I will lay out this reading of the Test and discuss some implications for AI in our world today.

Ethics in AI: A Challenging Task

Ricardo Baeza-Yates (Institute for Experiential AI,
Northeastern University)

31 July 9:00

In the first part we cover five current specific challenges through examples: (1) discrimination (e.g., facial recognition, justice, sharing economy, language models); (2) phrenology (e.g., biometric based predictions); (3) unfair digital commerce (e.g., exposure and popularity bias); (4) stupid models (e.g., Signal, minimal adversarial AI) and (5) indiscriminate use of computing resources (e.g., large language models). These examples do have a personal bias but set the context for the second part where we address four generic challenges: (1) too many principles (e.g., principles vs. techniques), (2) cultural differences; (3) regulation and (4) our cognitive biases. We finish discussing what we can do to address these challenges in the near future.

*Wittgensteinian Considerations of the Moral
Status of Social Robots*

Paula Sweeney (University of Aberdeen)

31 July 16:15

I am interested in possible attributions of agency and consciousness to social robots, particularly as these possibilities arise in connection with robot rights. I set the background for this paper by drawing on recent literature in the ethics of AI regarding the question of consciousness and the possible re-emergence of theories of behaviourism. I then explore the questions of agency and consciousness in social robots from a Wittgensteinian perspective, asking what criteria we would have for ascribing inner processes to social robots. I will draw on the Wittgensteinian notion of a form of life and consider the relevance of the view that forms of life sit against a background of foundational beliefs that can shift slowly over time. I outline my metaphysical view of social robots—the fictional dualism theory—and argue that this view can help us to frame the greater inclusion of social robots in human society. Finally, I consider whether the inclusion of social robots in human society will impact on and enter into our existing forms of life, or whether these developments will cause altogether new forms of life to arise.