© THE NEW INSTITUTE / Maximilian Glas

Kit Fine, Ph.D.

Silver Professor of Philosophy and Mathematics

New York University

Born in 1946 in Farnborough, United Kingdom
Studied Philosophy, Politics, and Economics at Balliol College, Oxford

PROJECT

A Numerical Model for Parity and Imprecision

Building on the work of Ruth Chang, I develop a numerical model for parity and imprecision in value and in belief. The basic idea is to replace the notion of an approximate value or an approximate difference in credence with the notion of an approximate difference in values and an approximate ratio of credences. Thus we can no longer properly speak of the value (even the approximate value) of an item or the credence (even the approximate credence) of a proposition, but should speak instead of the approximate difference in values or the approximate ratio of credences. This simple move turns out to be an extremely powerful device that allows us to go far beyond what is representable by means of approximate values or approximate credences. We also jettison the idea of representing parity or imprecision by a set of precise values or a set of precise credences. Sometimes there exists no underlying set of precise values or precise credences; but, even when they do exist, they do not constitute a useful tool for computing how parities or imprecisions should be resolved. The framework should admit of a wide range of applications, and among the applications we wish to consider is one to decision-making in AI, wherein the notion of an approximate difference or ratio can provide a useful tool for the machine to communicate with the user on how hard cases are to be resolved.

Recommended Reading

Fine, Kit. Modality and Tense: Philosophical Papers. Oxford: Clarendon Press, 2005.
—. Semantic Relationism. Malden, MA: Blackwell, 2007.
—. Vagueness: A Global Approach. New York: Oxford University Press, 2020.

TUESDAY COLLOQUIUM, 18.06.2024

The Parity Model: Putting Humans in the Loop

Value alignment, getting AI outputs to align with considered human values, is arguably the most significant open problem in AI research and development. We propose an approach to AI design rooted in a novel philosophical understanding of human values. The proposed model is an alternative to traditional forms of AI design – whether in the form of machine learning or symbolic systems – that may help us towards achieving value alignment. The numerical framework behind the model makes use of approximate differences and approximate quotients and thereby allows for the computational tractability of values outside the traditional framework of decision theory. We show how the model can then be used to facilitate communication between AI and its users when hard choices are in question.

PUBLICATIONS FROM THE FELLOWS' LIBRARY

Fine, Kit (Cham,2023)

Kit Fine on Truthmakers, Relevance, and Non-classical Logic

https://kxp.k10plus.de/DB=9.663/PPNSET?PPN=1877510092

Outstanding Contributions to Logic ; 26

https://kxp.k10plus.de/DB=9.663/PPNSET?PPN=1877510092

Fine, Kit (Dordrecht [u.a.],2023)

A semantics for the impure logic of ground

https://kxp.k10plus.de/DB=9.663/PPNSET?PPN=1853598283

Fine, Kit (London,2020)

The identity of social groups

https://kxp.k10plus.de/DB=9.663/PPNSET?PPN=1853595195