

# Philosophical lessons in autism for Artificial Intelligence

John Harpur

Computer Science, NUI Maynooth

Kildare, IRELAND

Email: [jharpur@cs.nuim.ie](mailto:jharpur@cs.nuim.ie)

## Abstract:

Autism is a neurobiological developmental disorder characterised by impairments in communication, imagination and social interaction skills. Following Wing's research into the heterogeneity of the condition a spectrum of autistic disorders has been identified. Currently there is no single accepted causal explanation of autistic disorders. However, there is agreement that these disorders are organic and life long in persistence. What is the relevance of this unfortunate condition to computational science? Well if not autism *per se*, at least theorising about autism, should be of interest to artificial intelligence (AI) for one profound reason: many of the mysteries and problems affecting the development of AI are coincident with similar ones exhibited in autism. For example, what are the conditions for *effective reciprocal communication* among 'agents' at whatever level of intelligence? What would have to count for an agent to have a theory of self and world without sliding into solipsism? How can a *reasonable altruism* be brought about among agents? Can commonsense be inculcated in agents with 'socially negative' cognitions? Of course, these questions are not uniquely associated with the areas cited here. They have a profound history in philosophy and a gathering presence within AI research over the past thirty years. What is striking nevertheless is that these questions have arisen within specific paradigmatic interpretations of human condition not entirely discordant with classical notions of robotic intelligence. Given the medical characterisation of autism, it may come as a surprise to learn that theoretical explanations of its varieties of manifested consciousness and cognitive mechanisms are rooted in philosophical explorations of the self, intentionality and the mental organisation of information processing. Arguably, most of the work implies quite a thin theory of rationality. However, there is one potentially broad theory based in understanding self-and-other intersubjectivity. The heuristic here is that in teasing out the philosophical implications of the parallels between this work and AI concerns about agency, important gains in understanding the sufficiency conditions for agency and agent interaction can be made.