"Unraveling the self: from bodily self-consciousness to artificial general intelligence"

Gabriel Axel Montes

B.Sc. (Hons) (University of Florida); M.Sc. (Utrecht University)

THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Medicine

August 2019

STATEMENT OF ORIGINALITY

I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision. The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

Gabriel Axel Montes

21 August 2019

THESIS BY PUBLICATION

I hereby certify that this thesis is in the form of a series of papers. I have included as part of the thesis a written declaration from each co-author, endorsed in writing by the Faculty Assistant Dean (Research Training), attesting to my contribution to any jointly authored papers.

***

By signing below, I confirm that Gabriel Axel Montes contributed:

  (1) The large majority of the written content to the publication titled: "Distributed, Decentralized, and Democratized Artificial Intelligence"*; and*

  (2) The large majority of the written content and the synthesis of concepts to the publication titled: "Mindplexes, Non-Ordinary Consciousness, and Artificial General Intelligence".*

<u>*Co-author*</u>

Dr. Ben Goertzel   **Date**

Aug 25, 2019

<u>*Faculty Assistant Dean Research Training*</u>

Dr. Lesley MacDonald-Wicks   **Date**

27th August 2019

Table of Contents

The body of work enclosed herein comprises my doctoral dissertation. Its content was composed between 2015 and 2019. This thesis is by publication. The introductory and conclusion chapters have not been formally prepared and submitted for publication at the time of this writing. The remaining chapters have been published at the time of this writing, with the exception of one chapter, which has been accepted for publication. The chapter *Virtual Reality for Non-ordinary Consciousness* was published in *Frontiers in Robotics and AI* in 2018. The chapters *Causal Biomimesis* and *Non-ordinary Consciousness for Artificial Intelligence* were both published in the peer-refereed conference proceedings of the *Living Machines Conference* that was held at Stanford University in 2017. The chapter *Distributed, Decentralized, and Democratized Artificial Intelligence* was co-authored with Dr. Ben Goertzel and published in the journal *Technological Forecasting and Social Change*. The chapter *Mindplexes, Non-ordinary Consciousness, and Artificial General Intelligence* has been accepted for publication as a book chapter in an academic volume on perspectives on artificial intelligence.

I am honoured that you, the reader, have picked up this work. The ordering of the chapters is intentional and reflects the flow of ideas that builds the sweeping multipart body of work. This thesis is something akin to a written version of cubist art; each chapter presents a different angle on a subject, each piece being mostly different from the others whilst having some conceptual overlap. The flow of ideas evolves swiftly, traversing the territories of cognitive neuroscience, virtual reality, philosophy of mind, cognitive anthropology, and ethics of artificial intelligence. The conclusion of this thesis consists of an invitation into inspired action. You pick up where I leave off, taking what resonates and making something out of it that is your own.

I am grateful all those who have supported my process, including my supervisors and collaborators for their support, my parents for nurturing my growth into an independent thinker, and my life partner and wife for her unending love. My hope is that you do with this body of work what I do with the information with which I choose to engage: allowing oneself to be inspired to create something new with the intention of contributing to the greater understanding and greater good.

Yours, Sincerely,

Gabriel

ABSTRACT

The subjective sense of having a self is commonly assumed to be static and anchored to the physical body. This presumption strongly preconditions how neuroscientists, philosophers, and computer scientists conceive of mind and attempt to engineer 'intelligence'. The present thesis expounds on how the brain constructs reality through mechanisms of bodily self-consciousness (BSC) and the free-energy principle (FEP), and, notably, how this process can be manipulated and enhanced. This work shows how BSC can be manipulated endogenously—through self-regulation methods, e.g. meditation—and/or exogenously—through virtual reality and artificial intelligence technologies, with applications for healthy and clinical populations. I describe the human evolutionary emergence of selfhood and present the novel Causal Biomimesis hypothesis that tethers the formation of the sense of self with tool-making, objective thought, technological development, and sociality. I argue that, particularly regarding artificial intelligence (AI) technologies, the externalisation of human cognition as technological artefact and its feedback effect on humans raises an ethical concern: that we are building AI, a class of artefact that is poised to strongly affect the evolution of humanity, based on a current image of ourselves and our perceived capabilities. From this ethical consideration, I make a case for the research and incorporation of non-ordinary consciousness (NOC), i.e. via self-regulation methods, into the design and engineering of AI. I present a neurophenomenological praxis (NP) based on the principles of the FEP that serves as a framework for instantiating (1st-/2nd-person) and researching (3rd- person) NOC and as a cognitive map for designing AI capable of supporting NOC. I discuss how a decentralised approach to AI could maximise the chances of a more ethical and representative AI by capturing and integrating a wider swathe of humanity's values. Finally, I explore the farther frontiers of the implications of this thesis work: merged human-AI cognition in the form of what are herein termed "mindplexes", where the phenomenology of NOC cultivation may elucidate the workings of mind in a way that helps humanity build a more capable and ethical human-AI future.