

Transhumanist Aspirations for Accelerating Insight:
A Philosophical and Theological Critique

Sean Patrick Biggins
Blackfriars Hall, University of Oxford
Faculty of Theology and Religion
Supervisors: Dr. Andrew Pinsent and Prof. Alister McGrath

For the Degree of Doctor of Philosophy in Theology and Religion

11 October 2025

Short Abstract

Transhumanism has been increasing in importance as a subject of interdisciplinary scholarship for philosophy and theology. In this thesis, I argue that the transhumanist aspiration to cognitive enhancement is not well founded due to the movement's inadequate treatment of the related concepts of insight and understanding. Although understanding and insight have received at least some philosophical and theological exploration, little has been done to examine how these issues pertain to transhumanist aspirations. This thesis aims to address this omission. I develop this argument by presenting the analyses of cognition and its enhancement as presented by three representative transhumanists: Marvin Minsky, Ray Kurzweil, and Nick Bostrom. I present, in contrast to the transhumanist view, the philosophical perspectives of Bernard Lonergan and Michael Polanyi and the theological perspective of Thomas Aquinas. This thesis also draws from philosophical and psychological research on metaphor, narrative, and joint attention. I then analyse the transhumanist positions of Minsky, Kurzweil, and Bostrom in light of these philosophical and theological perspectives. The conclusion reached is that the transhumanist perspectives have an inadequate appreciation of insight and understanding, especially when contrasted with the alternative perspectives of Lonergan, Polanyi, and Aquinas. These inadequacies explain why three of the most articulate and philosophically sophisticated transhumanists fail. This thesis thus proposes a benchmark by which transhumanist accounts need to be assessed and raises ground for doubting the attractiveness of transhumanism generally. The philosophical and theological perspectives discussed also point toward an alternative conception of technologically enhancing human cognition that respects the full reality of insight and understanding.

Long Abstract

Transhumanism has been increasing in importance as an area of scholarship and as topic for interdisciplinary research with theology. Transhumanism is a term that encapsulates many different movements and research initiatives, but there is a significant strand that is interested in achieving human cognitive enhancement. In seeking to enhance human cognition, transhumanists inevitably consider the faculty known as understanding, along with insight, a name given to changes to superior understanding. Given also that insights are often associated with great intellectual advances, there is interest from the transhumanists in seeking technological ways to enhance this faculty and accelerate the acquisition of insights. Nevertheless, although understanding and insight have received at least some philosophical and theological exploration, little has been done to examine how these issues pertain to transhumanist aspirations. This thesis aims to address this omission.

Chapter 1 introduces the movement of transhumanism with its aspirations to the enhancement of all cognitive faculties, including that of insight. The chapter also introduces the key terms “insight” and “understanding” which will be the lens through which is shown the transhumanist perspective on the one hand and the philosophical perspective on the other. A brief history of the origin of transhumanist thought and its relation to theology are discussed. The key transhumanist and philosophical interlocutors are introduced. These transhumanists are Marvin Minsky, Ray Kurzweil, and Nick Bostrom. The philosophers against whom the transhumanists will be comparatively evaluated are Bernard Lonergan, Michael Polanyi, and Thomas Aquinas, with Aquinas’ theology being used to enlarge and affirm the broadly Aristotelian-Thomist perspective presented. The overall structure of the thesis is presented, and the chapter concludes with a statement of what the thesis aims to prove.

Chapter 2 presents the perspectives on insight of the three transhumanists whose work is criticized in this thesis.

Section 2.1 examines Marvin Minsky as an early thinker who investigated transhumanist concerns before the term gained wide currency. Minsky’s conceptual framework for examining understanding and insight in terms of his hypothesis of Frames is presented. On his view, understanding and insight are terms of convenience, which will be replaced by a scientific account in terms of mechanism and algorithm. It is argued that his account of understanding and insight is ultimately in terms of representation and mechanism.

Section 2.2 examines Ray Kurzweil, Minsky’s student, and prominent proponent of transhumanism. Kurzweil’s conditions for ascribing intelligence to a computer are discussed, with emphasis on his debate with John Searle concerning the viability of the Turing Test as evidence for understanding. Kurzweil is shown to hold that apparently intelligent behaviour is sufficient to claim genuine understanding. Kurzweil’s explanatory framework for cognition is then presented, along with his view that understanding is fundamentally a matter of pattern recognition, with insight being a matter of finding new metaphors. Finally, Kurzweil’s techno-optimism is presented in terms of his expectation that reverse engineering the neocortex will

allow an expansion of intellectual powers through using implants to connect brains to extensible artificial networks of pattern processors. It is concluded that his behavioural interpretation of the Turing Test frees him to speculate about artificial means to pass the Turing test, if only more sophisticated pattern processors can be developed.

Section 2.3 examines Nick Bostrom's account of potential avenues to developing what he calls "superintelligence". The three primary lines of speculation examined in this section are: speed superintelligence; collective superintelligence; and quality superintelligence. The assumptions of these three types of superintelligence and their implications for understanding are discussed.

Chapter 2 concludes with a summation of the transhumanist perspectives on understanding and insight. It is concluded that the transhumanists are interested in enhancing the ability to solve results-based problems for intelligence rather than intellectual projects where acquiring understanding is not primarily a matter of producing results. The transhumanist perspectives examined here are examined in light of the account of understanding and insight developed in chapter 3.

Chapter 3 provides philosophical and theological perspectives on understanding and insight with support from developmental psychology and linguistics.

Section 3.1 gives an overview of the three sections of the chapter. Section 3.2 first discusses aspects of insights and insight experiences as described by Lonergan, highlighting the apparent simplicity of insights, the nature of insights as step-changes between being in a state of not understanding to being in a state of understanding, and the tendency of insights to produce what may be termed a higher viewpoint. The section then discusses how Lonergan's conceptions of insight and understanding relate to the Thomistic system of virtues. Next, this section investigates the role of embodiment in understanding as facilitated by metaphor and narrative. The accounts of metaphor given by Iain McGilchrist and Lakoff and Johnson are also examined for their claim that metaphor grounds language by linking words to embodied experience. This analysis is used to illuminate the relationship between embodiment and understanding and the importance of metaphors for insight. In further support of the claim that symbols and abstractions are understood through embodiment, evidence is presented that metaphor and embodiment are not only important for linguistic understanding but that they also contribute to understanding and insight in mathematics. After having examined the role of embodiment in understanding language in general and in understanding and gaining insights into mathematics, this section turns to an examination of narrative. Narrative is explored as a privileged means to communicate understanding due to the imaginative engagement it invites with embodied scenes and the experiences and stances of the characters it depicts.

Section 3.3 examines second-person aspects of insights. I begin with a brief summary of Polanyi's account of apprenticeship as a relation between master and apprentice that allows what he calls tacit knowledge to be shared from the master to the apprentice. I then discuss the value of interactions between peers for mutually stimulating insights. Together, these points illustrate that second-person encounters can aid the communication of insights from one

person to another and the stimulation of new insights between peers. Evidence from developmental psychology on the importance of joint attention for human intellectual development is then discussed, especially that of Peter Hobson and Michael Tomasello. Evidence from studies of children with severe autism spectrum disorder is also cited to support the claim that the capacity to enter into a state of joint attention with a second-person is an important means of conveying and catalysing insights. Whatever else may be involved, the observations above imply that shared attention may play an important role in the acceleration of shared understanding.

Section 3.4 examines scriptural and theological perspectives on understanding. The association between the divine and the illumination of the intellect through sudden understanding is discussed. Moreover, the fundamental identification of the divine with all understanding and intelligibility is highlighted. Aquinas' analysis of the scriptural concept of the Word of God is used to demonstrate that, within his theological perspective, understanding *with* the divine is of supreme importance and that this understanding is metaphorically similar to an encounter with a person. Moreover, this understanding of the divine is not a matter of knowing propositions or having some mastery of representations of the divine but is an interpersonal illumination. This section then applies the examination of narrative in the preceding section to scriptural poetry and narrative, highlighted the role of embodied experiences and tacit shared understanding. Particular attention is paid to Jesus' own use of narrative in the form of parables. Jesus' use of narrative, it is argued, underpins how narratives are particularly fit for communicating theological understanding to those who do not possess it. Finally, the narratives of Paul's conversion and of the disciples on the road to Emmaus are explored. Both narratives evince that the Christian conception of insight places particular emphasis on the importance of personal encounter. Finally, Aquinas' account of the gift of understanding is discussed. Chapter 3 concludes that understanding has first- and second-person characteristics, from both philosophical and theological perspectives, with a cognitive value that goes far beyond disembodied facts.

Chapter 4 examines the transhumanist perspectives discussed in chapter 1 in light of the philosophical and theological perspectives discussed in chapter 3. In particular, this chapter evaluates the points of consonance and dissonance between the transhumanist perspectives and the philosophical and theological perspectives along five key themes. These themes are: insight as a singular ascending operation; the grounding of understanding in embodied experience; the role of metaphor in understanding; the role of narrative in understanding; and the effect of second-person catalysis on insight. This chapter concludes by highlighting the persistent gulf between transhumanist aspirations and the perspectives of theology and philosophy, notwithstanding many individual points of agreement.

Chapter 5 outlines prospects for greater harmonisation in future, with a focus on how some of the transhumanist aspirations might be realised in practice despite the problems highlighted above. The first section distinguishes direct and indirect enhancement of insight. The second section discusses the use of technology to enhance the conditions favourable to insight, whether by internal improvements to the body or by improving the external

environment accessible to embodied human cognition. The third section concerns the outsourcing of cognitive tasks in which only the result of insight matters. The fourth section examines the use of computers to uncover missing insights. The fifth section discusses the judicious use of technologies to facilitate second-person relatedness. The chapter argues that there exist at least some opportunities for the use of technology to augment acts of understanding and insight.

Chapter 6 presents my summary and conclusions, including suggestions about how the findings of this thesis might be implemented in practical terms in future.

Acknowledgements

It is my pleasure to acknowledge the persons and institutions that have helped make this thesis possible. I am grateful to the students and scholars of the University of Dallas who first introduced me to the study of the history and philosophy of science and to the scholarly debates on transhumanism. In particular, I thank Dr Christopher Mirus and Dr Charles Sullivan for their encouragement and scholarly penetration. I am also deeply indebted to the Faculty of Theology and Religion at the University of Oxford. In particular, I thank Dr Donovan Schaefer for introducing me to the distinction between post-humanist and transhumanist schools. I also owe a great debt of gratitude to my supervisors: Prof. Alister McGrath, for shepherding my early scholarship in Science and Religion and offering helpful perspectives about wider background scholarship; and Dr Andrew Pinsent for opening my eyes to the importance of second-person relations and for much patient and granular feedback over the last several years. I also thank Blackfriars Hall for its academic and personal support. I thank Dr Richard Conrad for his sage advice, pastoral guidance, and tremendous generosity during my final stage of revisions. I am grateful to Dr Clare Broome Saunders for always being ready to offer a sympathetic ear. I also extend my gratitude to Dr. John O'Connor for his advice and support. Finally, I would like to thank my parents and family for their love and encouragement during this project.

SHORT ABSTRACT	I
LONG ABSTRACT	II
ACKNOWLEDGEMENTS	VI
CHAPTER 1 INTRODUCTION	1
1.1 TRANSHUMANISM AND THE CHALLENGE OF INSIGHT	1
1.2 BACKGROUND TO TRANSHUMANISM.....	8
1.3 STRUCTURE OF THE ARGUMENT.....	16
1.4 CONCLUSION.....	30
CHAPTER 2 TRANSHUMANISM AND THE CHALLENGE OF UNDERSTANDING	33
2.1 MARVIN MINSKY: FRAME LANGUAGE, KNOWLEDGE, AND INSIGHT	33
2.1.1 Introducing Marvin Minsky	33
2.1.2 Minsky on Understanding and Frames.....	33
2.1.3 Concluding Remarks on Minsky	44
2.2 RAYMOND KURZWEIL: PATTERN RECOGNITION AND UNDERSTANDING	44
2.2.1 Introducing Ray Kurzweil	44
2.2.2 Kurzweil on Understanding and Pattern Recognition	46
2.2.3 Concluding Remarks on Kurzweil	54
2.3 NICK BOSTROM: INSIGHT AS A PROMISSORY BLACK BOX	55
2.3.1 Introducing Nick Bostrom.....	55
2.3.2 Bostrom on Understanding and Superintelligence	56
2.3.3 Concluding Remarks on Bostrom	67
2.4 SUMMING UP: TRANSHUMANISTS ON UNDERSTANDING	68
CHAPTER 3 INSIGHTS INTO INSIGHT AND SECOND-PERSON RELATEDNESS	71
3.1 INTRODUCTION: THE PERSONAL DIMENSION OF UNDERSTANDING	71
3.2 FIRST-PERSON ASPECTS OF UNDERSTANDING	74
3.2.1 Introducing Terms: Understanding and Insight.....	74
3.2.2 The General Properties of Insight: Lonergan’s Analysis	79
3.2.3 Embodiment as a Ground of Understanding: Lessons from Metaphor and Narrative	90
3.3 SECOND-PERSON ASPECTS OF UNDERSTANDING	121
3.3.1 Apprenticeship and Sharing Understanding.....	121
3.3.2 Sharing Understanding and Catalysing Insight: Support from Data Science.....	123

CHAPTER 1 Introduction

1.1 Transhumanism and the Challenge of Insight

Transhumanism is a movement that aspires to overcome human limits through the transformation of the human organism. A key aspiration for transhumanists is overcoming human cognitive limits. As a proximate goal, they advocate what they call biological enhancements such as gene therapies to increase intelligence and nootropics to improve cognitive performance. Many transhumanists have greater ambitions, however, and wish to overcome biological limitations entirely. Such transformation includes the idea of integrating computers and other machinery into the brain with the goal of expanding the computational matter available to a given human. As an ultimate goal, the transhumanists advocate ceasing to be biological entirely and to exist instead as software with access to computational capacities ostensibly greater than those of a human brain. Given that they aspire to cognitive augmentation, it follows that they have an interest in enhancing the intellectual power called understanding, and those changes in understanding that are given the special term, insight. This thesis examines the prospects of transhumanism for accelerating insight as part of its broader aspiration to enhance cognition.

One representative expression of the transhumanist ambition to augment cognition comes from Nick Bostrom, an Oxford philosopher and arguably the most influential transhumanist in the world. In his rhetorical article “Letter from Utopia”, he writes,

Your brain must grow beyond any genius of humankind, in its special faculties as well as its general intelligence, so that you may better learn, remember, and *understand*, and so that you may apprehend your own beatitude. Mind is a means: for without *insight* you will get bogged down or lose your way, and your journey will fail. Mind is also an end: for it is

in the spacetime of awareness that Utopia will exist. May the measure of your mind be vast and expanding.¹

What is striking about this passage is that, according to Bostrom, part of the work of accelerating intelligence involves enhancing the faculty of *understanding*. In fact, he also uses another word that is closely related to understanding, namely *insight*, without which, he states, transhumanist individuals will lose their way. In assessing transhumanist aspirations, it is therefore clear that an assessment and possible critique must be made of the transhumanist notions of understanding and insight.

This passage is also of interest for linking the notion of understanding to that of beatitude, which is a term denoting supreme happiness in English and derived from the Latin word for blessing. The word is most famously used to describe Jesus' teachings on the Sermon on the Mount in Matthew. The related word "beatific" is used to describe the final state of the saints as they behold God. The imagery evoked by Bostrom in describing the experience of understanding to which transhumanists aspire would appear to be religious, at least metaphorically, but whether the transhumanist meanings of understanding and insight are congruent with those of philosophical and theological perspectives is a question to be investigated.²

According to Nick Bostrom transhumanism may be described as

a loosely defined movement that has developed gradually over the past two decades, and can be viewed as an outgrowth of secular humanism and the Enlightenment. It holds that current human nature is improvable through the use of applied science and other rational methods, which may make it

¹ Nick Bostrom, "Letter from Utopia," *Studies in Ethics, Law, and Technology* 2, no. 1 (2008), <https://doi.org/10.2202/1941-6008.1025>, emphasis added.

² Other authors have treated of the differences between the Christian and transhumanist notions of human beings and the prospects for bettering them and their state, examining the significance and limitations of enhancement technologies from the perspective of Christian soteriology and virtue ethics, but work remains to be done on the differing conceptions of human cognitive powers in transhumanist and Christian theological accounts. Michael S. Burdett, "Contextualizing a Christian Perspective on Transcendence and Human Enhancement," in *Transhumanism and Transcendence: Christian Hope in an Age of Technological Enhancement*, ed. Ronald Cole-Turner (Georgetown University Press, 2011).

possible to increase human health-span, extend our intellectual and physical capacities, and give us increased control over our own mental states and moods.³

In this description, science and technology are identified as the privileged means by which humans will evolve beyond their limitations while remaining within those categories of achievement that human beings would continue to care about as human. In other words, transhumanism is typically defined in terms of extending human capabilities according to a spectrum of ever greater attainment. Hence, for example, transhumanists use terms like *human*, *transhuman*, and *posthuman*. However, transhumanists still situate themselves within the humanist tradition. In other words, transhumanism is still in contrast to what is today called *posthumanism*, a world of discourse that rejects measuring success on an anthropological basis altogether, and which is much harder to define or evaluate.⁴

A further elaboration of transhumanist goals is that,

Many transhumanists wish to follow life paths which would, sooner or later, require growing into posthuman persons: they yearn to reach intellectual heights as far above any current human genius as humans are above other primates; to be resistant to disease and impervious to aging; to have unlimited youth and vigor; to exercise control over their own desires, moods, and mental states; to be able to avoid feeling tired, hateful, or irritated about petty things; to have an increased capacity for pleasure, love, artistic appreciation, and serenity; to experience novel states of consciousness that current human brains cannot access.⁵

³ Nick Bostrom, "In Defense of Posthuman Dignity," *Bioethics* 19, no. 3 (2005): 202–3, <https://doi.org/10.1111/j.1467-8519.2005.00437.x>.

⁴ cf. Cary Wolfe, *What Is Posthumanism?*, Posthumanities 8 (Minneapolis: University of Minnesota Press, 2010), xi–xviii, <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=557541>.

⁵ "Transhumanist FAQ," Humanity+, accessed September 20, 2022, <https://www.humanityplus.org/transhumanist-faq>.

In other words, transhumanists desire to overcome human limitations in the sense of changing into different kinds of organisms, termed *posthumans*, possessing superhuman capacities, especially intellectual capacities.⁶

The aspiration to rise above human intellectual limits is of particular importance and is one of the most repeated themes in transhumanist literature. According to one oft-repeated definition, also cited in the Transhumanist FAQ:

Transhumanism is a class of philosophies of life that seek the continuation and acceleration of the evolution of intelligent life beyond its currently human form and human limitations by means of science and technology, guided by life-promoting principles and values.⁷

According to this passage commonality in different strains of transhumanism is a principled use of science and technology to overcome human form and limitations to accelerate the evolution of intelligence. Further, by implication technology will provide the means of achieving the aforementioned augmentation of understanding and consequent acceleration of insight.

Yet what is meant by the term “understanding”? The act of understanding is familiar and pervasive in human life but describing in philosophical terms what it means to understand can prove surprisingly difficult. Until 2021, even the *Stanford Encyclopedia of Philosophy* had no article on “understanding”, a fact that may reflect a broader neglect of this topic in the broader philosophical community at present. The new article by Stephen Grimm does, however, underline at least one key point, namely that understanding seems to be different to knowledge.

⁶ As a point of clarification, it should be noted that the terms transhumanism and posthumanism denote distinct schools of thought that are sometimes even in opposition, in spite of the fact that many transhumanists use the term “posthuman”. Transhumanism is typically defined in terms of going beyond, i.e., transcending human limitations, whereas posthumanism, the definition of which remains more fluid, typically seeks to redefine such boundaries altogether. This thesis is not concerned with the posthuman rejection of human categories or efforts to redefine the boundaries between the human and the inhuman. This thesis is rather interested in the effort to expand the limits of specifically human intellectual capacity. Hence the focus of this thesis is on transhumanism rather than posthumanism. cf. Cary Wolfe, *What Is Posthumanism?*, Posthumanities 8 (Minneapolis: University of Minnesota Press, 2010), <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=557541>.

⁷ Max More, 1990, quoted in Humanity+, “Transhumanist FAQ.”

In particular: Where knowledge can be atomistic, understanding is holistic. Knowledge may be isolated, but understanding is interconnected. Knowledge may be typified by assent to propositions, but understanding is paradigmatically described as a metaphorical grasping or seeing.⁸

Even at the beginning of a long philosophical tradition of reflection of understanding, there is reason to think that understanding is linked to theological concerns. In the ancient narrative of Plato's *Myth of the Cave*, prisoners in a cavern are chained their entire lives to a wall and can only see shadows cast upon the opposite wall by objects illuminated behind them. They wrongly think that in knowing the shadows before them, they understand what is. The story of the prisoners being unchained and drawn forth to see things as they are serves Plato as a metaphor for education as a mind is taught to proceed from a superficial knowledge of things to an understanding of them. The educational journey culminates in the prisoners being drawn out of the cave into the illuminating light of the sun of wisdom, which for Plato is also an image of the divine. In other words, a possible connection between understanding and theological revelation was posited early in the history of philosophy, and the longstanding philosophical concern with their relationship may suggest that studying theology may itself be illuminating with respect to understanding.

Given, therefore, that there does appear to be a dimension of human thought called "understanding" and that transhumanists themselves offer at least some acknowledgement of its importance, any transhumanist aspiration to augment human intellectual abilities will therefore need to take account of this intellectual faculty of understanding. Taking account of

⁸ Stephen Grimm, "Understanding," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2021 (Metaphysics Research Lab, Stanford University, 2021), <https://plato.stanford.edu/archives/sum2021/entries/understanding/>.

this faculty will also overlap with a long history of its exploration and role in both philosophy and theology.

As a provisional description of insight, the OED defines insight as:

The fact of penetrating with the eyes of the understanding into the inner character or hidden nature of things; a glimpse or view beneath the surface; the faculty or power of thus seeing.⁹

The OED defines insight through the metaphor of seeing. In particular, it expresses this sight as grasping an inner character or a nature beyond the superficial. At minimum, this implies that insight is cognitively significant for seeing the nature of things.

This thesis will find many occasions to refer to both understanding and insight. Both these terms will be elaborated upon more extensively in chapter 3, but for now it will serve as a rule of thumb state that whenever this thesis refers to understanding, it refers to the faculty when it is stable, and insight refers to the faculty when it is undergoing a change, when something new is understood.

The principal focus will be on philosophical and theological perspectives on insight, but the universal presence of insight in human life means it has been examined and represented in many other fields including psychology, cognitive linguistics, and literature, so perspectives from these fields on insight will also be explored.

This thesis will present the perspectives on insight of three prominent transhumanists: Marvin Minsky, Ray Kurzweil, and Nick Bostrom. This thesis will then explore the contrasting views of insight given by the philosophies of Bernard Lonergan and Michael Polanyi and the theology of Thomas Aquinas.

⁹ *OED Online*, “insight, n.1” (Oxford University Press), accessed November 26, 2019, <https://www.oed.com/view/Entry/96831>.

The argument is philosophical anthropology broadly in the Aristotelian-Thomist tradition affirmed and enlarged by the Thomist theological tradition. Presenting the transhumanist, philosophical, and theological perspectives on insight entails giving some sense of what it means to be a human that understands and has insights.

The theological perspective is developed to support the philosophical perspective, for theology can be taken to expound truths that cannot be proven by reason but which, after accepted on faith, can be shown to be philosophically fitting.

The transhumanist and philosophical perspectives will be compared relatively to each other with a movement toward criticism insofar as points where the transhumanist account is comparatively thin will come into relief. The theological perspective thickens the philosophical perspective and sharpens these points of comparative critique where appropriate.

The principal focus will be on transhumanist, philosophical and theological perspectives on insight, but the universal presence of insight in human life means it has been examined and represented in many other fields including psychology and literature, so perspectives from these fields on insight will also be explored.

This thesis will present the perspectives on insight of three prominent transhumanists: Marvin Minsky, Ray Kurzweil, and Nick Bostrom. This thesis will then show the contrasting views of insight given by the philosophies of Bernard Lonergan and Michael Polanyi and the theology of Thomas Aquinas. This thesis will then show that the transhumanists' accounts of insight and understanding are comparatively thinner than the those of the philosophers and theologians. This inadequacy of the transhumanist account, it will be argued, puts into question their aspirations for cognitive enhancement. The philosophical and theological perspectives, however, do provide for ways to understand how to use emerging technologies to benefit human insight.

The question of this thesis is not merely of scholarly interest as to whether or not the transhumanist anthropology's conception of understanding and insight are satisfactory from philosophical and theological points of view. On a natural-law perspective, for instance, an understanding of human nature helps in assessing what does, and what does not, promote true human well-being. This study is important, then, not only for the scholarly interest in its central question but for the implications it has for the ethics of human use of technology.

1.2 Background to Transhumanism

Although the transhumanists present their philosophy as naturalistic, the religious qualities of their thought have been discussed. Ted Peters observes that transhumanists often claim that religion is a road-block to progress.¹⁰ Jason Eberl argues that while the transhumanist accusation of Luddism can be valid for some religious movements, it is not universally true. Christian traditions may critique aspects of transhumanism while still accepting technologies and interventions contributing to human flourishing.¹¹ Michael Burdett observes that transhumanism has four themes of religious character: the cessation of death, the promise of individual transcendence, the transformation and glorification of self, and the promise of hope.¹²

¹⁰ Ted Peters, "Progress and Provolution: Will Transhumanism Leave Sin Behind?," in *Transhumanism and Transcendence*, with Ronald Cole-Turner (Georgetown University Press, 2011), 71–73.

¹¹ Jason T Eberl, "Enhancing the Imago Dei: Can a Christian Be a Transhumanist?," *Christian Bioethics: Non-Ecumenical Studies in Medical Morality* 28, no. 1 (2022): 76–93, <https://doi.org/10.1093/cb/cbab016>.

¹² Michael S. Burdett, *Eschatology and the Technological Future*, Routledge Studies in Religion ; 43 (Routledge, 2015), 100–101, <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://www.taylorfrancis.com/books/9781315739397>.

Even in its early stages, transhumanism existed in a degree of dialogue with religion. The earliest recorded usage of the term transhumanism was by Julian Huxley, who was the friend of Teilhard De Chardin, a Jesuit who also influenced the development of transhumanism.

The human species can, if it wishes, transcend itself—not just sporadically, an individual here in one way, an individual there in another way, but in its entirety, as humanity. We need a name for this new belief. Perhaps *transhumanism* will serve: man remaining man, but transcending himself, by realizing new possibilities of and for his human nature.

“I believe in transhumanism”: once there are enough people who can truly say that, the human species will be on the threshold of a new kind of existence, as different as ours is from that of Pekin man. It will at last be consciously fulfilling its real destiny.¹³

In this context, Huxley referred to the desire to overcome human limitations while remaining human. For Huxley, the aspiration to improve the human organism directly was secondary to augmenting society, technology, and ideas.¹⁴ For contemporary transhumanists, the transformation of the human organism into a posthuman entity is a central concern because the inherent potential of the human organism is considered too limited.¹⁵

¹³ Julian Huxley, *New Bottles for New Wine* (Chatto and Windus, 1957), 17, emphasis original.

¹⁴ “Evolution in the human sector consists mainly in changes in the form of society, in tools and machines, in new ideas, in new ways of utilizing the old innate potentialities, instead of in the nature of those potentialities, as in the biological sector. Man’s inherited mental powers cannot have changed appreciably since the time of the Aurignacian cave-dwellers; what have changed are the ways in which those powers are used, and the social and ideological frameworks that condition their use. This is not to say what has happened to man since the Aurignacian period or since the time of ancient Greece is not evolution; it is a very astonishing bit of evolution. Nor does it mean that man’s innate mental powers could not be improved... they certainly could be improved further by deliberate eugenic measures. Meanwhile, however, it is in social organization, in machines, and in ideas that human evolution is mostly made manifest.” Huxley, *New Bottles for New Wine*, 23.

¹⁵ “The range of thoughts, feelings, experiences, and activities accessible to human organisms presumably constitute only a tiny part of what is possible. There is no reason to think that the human mode of being is any more free of limitations imposed by our biological nature than are those of other animals. In much the same way as Chimpanzees lack the cognitive wherewithal to understand what it is like to be human – the ambitions we humans have, our philosophies, the complexities of human society, or the subtleties of our relationships with one another, so we humans may lack the capacity to form a realistic intuitive understanding of what it would be like to be a radically enhanced human (a “posthuman”) and of the thoughts, concerns, aspirations, and social relations that such humans may have. Our own current mode of being, therefore, spans but a minute subspace of what is possible or permitted by the physical constraints of the universe... It is not farfetched to suppose that there are parts of this larger space that represent extremely valuable ways of living, relating, feeling, and thinking.” Nick Bostrom, “Transhumanist Values,” *Journal of Philosophical Research* 30 (2005): 4–5, https://doi.org/10.5840/jpr_2005_26.

In contrast to Julian Huxley, Teilhard de Chardin's theology of the Omega Point informed the versions of transhumanism that have the strongest eschatological valences. Teilhard proposed that advancing technology had eschatological significance. He coined the term noosphere, a play on nous and terms such as biosphere and atmosphere, to describe what he called a thinking layer to the Earth that had arisen with humanity. The increasing population and its interconnection through such technologies as the telegraph, which Teilhard likened to a nervous system for human civilisation, would cause increasing cultural complexity and density. Teilhard thought this would cause a growing monoculture (in fact, being the nascent body of Christ) that would eventually exceed the limits of the material world, catapulting humanity into the eschaton. Teilhard termed the threshold for this eschatological transformation the "omega point".

Subsequent transhumanists were influenced by the broad structure of Teilhard's eschatological narrative, although they did not retain his adherence to Catholic doctrine. Frank J. Tipler and John D. Barrow, for instance, adopted Teilhard's term "omega point" in *The Anthropic Cosmological Principle*, which speculated on a transposition of eschatology into physicalist terms. Rather than arguing in favour of the classical notion of God, Tipler suggested that something satisfying the description of the divine could be technologically created. Writing with the assumption that the Big Crunch cosmological model was correct, Tipler suggested that subjectively infinite time and thought could be achieved by creating computational structures as the universe collapsed back into a singularity. While a finite amount of time would remain prior to the collapse into singularity, the available energy would approach infinity, allowing more computation to be performed with each halving of the remaining time, resulting in an asymptotic amount of computation approaching infinity. This machine by tending toward infinite computation would, Tipler thought, be tending toward computing all that can be

computed, which would he suggested be equivalent to the divine attribute of omniscience. Additionally, Tipler supposed that it would be possible for human minds to exist as simulations within this computer (either by being uploaded at an earlier point in time or by being recreated by the machine simulating human history), which he suggested would be akin to eternal life in heaven and communion with the mind of God. Teilhard's narrative of the emergence of the person of Christ from technologically connected humanity was transposed into a physicalist narrative of individual persistence within a system of asymptotically increasing computation.

Tipler was not the only writer to explore such themes. Hans Moravec also speculated on the possibility of humanity transferring its consciousness into machines to achieve immortality and overcome the cognitive limits of human biology and speculated on similar systems to approach infinite computation in both an expanding and a contracting universe. As a side note, Moravec's proposal for maintaining computation for an indefinitely long time is not in principle viable given the current understanding of cosmology. Moravec suggested that two opposed reflectors, assuming no loss of energy, could keep reflecting light to each other while being pushed apart and using this force to drive computation. The time between each computational step would grow ever greater but would in principle be able to occur forever. However, the current understanding is that the universe's expansion is accelerating, which will result in the distance between objects eventually increasing at a rate greater than the speed of light, so there would exist a point in time after which the reflected light would never be able to reach one of the reflectors, ending this process.

Modern transhumanists no longer hope for infinite existence, subjective or absolute, but they still aim to maximize the quantity of time they can exist. Bostrom writes,

Any death prior to the heat death of the universe is premature if your life is good.¹⁶

Although finite, a sufficiently large magnitude can still be psychologically treated as infinite. Another transhumanist writer, Vernor Vinge was influenced by Hans Moravec and Frank Tipler in his influential essay “The Coming Technological Singularity”, but he does not reference Teilhard de Chardin. In this article, the transhumanist notion of the Singularity shifted from being literally that of the end state of the universe harnessed to power infinite computation to a metaphor drawing on the event horizon of black hole. As what lies beyond the event horizon is impossible to observe, so was it suggested that there would be an exponential acceleration of technological progress that would make the future of existence largely unimaginable. Teilhard’s original notion of the Omega Point was explicitly theological and influenced this pervasive transhumanist idea. As Bostrom’s parable of the “Letter from Utopia” shows, religious themes continue to be used in transhumanist narratives, and questions of the natures of the human and divine minds, and the relations between them, remain pertinent. Huxley and Teilhard’s transcendent humanity has changed into the transformed humanity of Nick Bostrom.

The narrative of the Singularity went on to influence subsequent transhumanists, among them Ray Kurzweil, who argued for the near-term timeline for the creation of intelligent machines with superhuman capabilities in *The Singularity is Near* and *The Age of Spiritual Machines*. Kurzweil’s teacher, the pioneering AI researcher Marvin Minsky, did not explicitly rely on the metaphor of the Singularity, although he expressed admiration for a different vision of the achievement of superhuman intellectual ability in Vinge’s short story “True Names”.

As Burdett has observed, although transhumanists do not regard their movement as religious, it nonetheless continues to serve many of the same psychological and social functions:

¹⁶ Bostrom, “Letter from Utopia.”

offering hope of immortality in the face of biological death, promising a superior state of being, and promising a superior state of cognition. Many of the commonalities and possibilities for convergence or divergence between transhumanism and Christianity have been explored, but more work remains to be done on the question of how theological and philosophical perspectives on insight relate to that of the transhumanists. Insight has been classically identified as a basic intellectual faculty by philosophy, and theological perspectives treat it as related to the virtues and gifts. In other words, philosophical traditions, like many transhumanists, agree that the intellectual powers are of particular importance to humanity, and theological traditions agree that an elevation of the intellect is necessary to achieve an essential component of an elevated, glorified state of human beings. Beyond this initial parallel, however, the question remains to be asked whether there is substantial convergence or divergence between transhumanist perspectives and philosophical or theological perspectives.

This thesis takes as its point of departure the question of what insight is, and how this is answered by the philosophical and theological traditions on the one hand, and the transhumanist movement on the other.

As previously stated, to ask what insight is will also require an examination of what understanding is. To this end, this thesis will focus on three principal transhumanists: Marvin Minsky, Ray Kurzweil, and Nick Bostrom.

Marvin Minsky (1927-2016) wrote on transhumanist themes before the movement was widely known and influenced its priorities and expectations. He contributed greatly to the development of programming and AI and advocated for human life-extension, cognitive enhancement, and the value of transcending biological limitations.

Marvin Minsky acknowledged the importance of the cognitive processes commonly called understanding and insight and regarded the question of how to make computers

understand as a valid project. Nonetheless, he rejected both the common belief that these terms describe a real and distinct cognitive power and the philosophical discussion of these faculties. Instead, Minsky treated insight and understanding as terms of convenience to be replaced by a mechanistic explanation. Minsky is of interest as not only an early influence on early transhumanists but on AI research, given his efforts to make sense of insight and understanding in terms of computational processes.

Ray Kurzweil was Marvin Minsky's student, and he one of the most well-known public apologists of transhumanism and forecasters of technological progress. Ray Kurzweil has record of innovations of engineering, including the first text-to-speech assistive technology for the blind, and has received endorsements from Bill Gates for predictions of future technological development. He was also for decades an engineer at Google, having been hired for the express purpose of engineering computers that could be said to understand, a goal that Kurzweil had several times claimed had been achieved. Kurzweil's usage of the term understanding, especially as applied to artificial intelligence informs his projections for technological development and his sense of what is possible for augmenting human intelligence.

The third transhumanist this thesis will discuss is Nick Bostrom, founder and director of the Future of Humanity Institute of Oxford from 2005 to its closure in 2024. Bostrom is possibly the most well-known transhumanist philosopher. His works include *Superintelligence*, a study of hypothetical pathways to creating superintelligent entities and the perils they might pose, as well as numerous articles and essays arguing for human enhancement and aspiring to posthuman existence. He philosophically and narratively considers how cognition might be enhanced, and what the existence of superintelligent beings might be like.

Marvin Minsky as part of his project to create intelligent machines articulated an account of human cognition in terms of a high-level description of representation and

algorithmic processes. This study includes an examination of what is called insight and attempt to replace it in terms of an algorithmic account of its related processes.

Ray Kurzweil, who studied under Marvin Minsky and was influenced by his ideas, defends both the idea that computers can already be said to understand. He also defends a notion of human cognition and insight in terms of pattern recognition set in the context of narratives of the radical transformation by technology of the human being and world.

Nick Bostrom identifies insight as crucial to the pursuit of transhumanist goals, and his detailed account of varieties of superintelligence is used to illuminate the transhumanist sense of what intelligence is for and the assumptions about the purpose of intelligence for human beings.

The philosophers this thesis uses to develop a philosophical perspective against which to judge the transhumanist perspective are Bernard Lonergan, Michael Polanyi, and Thomas Aquinas. The Jesuit Bernard Lonergan wrote the most extensive modern philosophical study of the phenomena of insight and its relation to everyday life, science, and philosophy. His system provides a useful groundwork for an exegesis of insight in the Aristotelian tradition.

Michael Polanyi was a polymath who made contributions to both physical chemistry and philosophy. He criticized the myth of aperspectival objectivity in scientific study, arguing instead that all knowledge and insight is perspectival, which is to say personal. Moreover, he articulated the view that there is a tacit dimension to what is understood and that this is shared between persons without being explicit.

Thomas Aquinas as a philosophical antecedent to Lonergan offers a systematic account of the human intellect that includes the classical intellectual faculty often translated as insight, *intellectus*. Additionally, Aquinas offers a theological perspective on intellectual virtues and fulfilment that can be used to sharpen the philosophical and criticisms offered in this thesis.

1.3 Structure of the Argument

The structure of the thesis is informed by its methodology. The first chapter introduces the movement of transhumanism with its aspirations to the enhancement of all cognitive faculties, including that of insight. The chapter also introduces the key terms “insight” and “understanding” which will be the lens through which is shown the transhumanist perspective on the one hand and the philosophical perspective on the other. A brief history of the origin of transhumanist thought and its relation to theology are discussed. The key transhumanist and philosophical interlocutors are introduced. The overall structure of the thesis is presented, and the chapter concludes with a statement of what the thesis aims to prove.

The second chapter presents transhumanist paradigms for insight and understanding and the cognitive enhancement these paradigms are meant to accommodate. The third chapter presents philosophical and theological paradigms for insight and understanding, drawing out several key themes concerning insight. The fourth chapter develops the themes concerning insight, observing the consonances and dissonances between the transhumanist and philosophical-theological paradigms. The fifth chapter considers applications of the philosophical and theological perspective to the question of how technology may augment human insight.

The second chapter presents the transhumanist perspective on insight and understanding, paying special attention to their use of these terms, and how they directly discuss insight and understanding in their efforts to describe it computationally, both as an explanation for human intelligence and as a design proposal for computer science, and indirectly in terms of intelligence may be enhanced to the point of superintelligence.

The first section of chapter two explores Marvin Minsky's approach to investigating insight and understanding as part of his project to reverse engineer human intelligence, create intelligent machines, and pave the way for superintelligent machines, which humans may subsequently become. Minsky proposes that the terms insight and understanding should be thought of as terms of convenience, which should be replaced by explanation in terms of computational structures. He predicts that once such structures are fully described, they will capture all the functions that presently seem basic in the acts of insight and understanding. Minsky proposes hypothetical computational structures that can handle much of what is called understanding. He calls these hypothetical structures "frames". On his account, frames capture all relevant details of objects and contexts and are functionally equivalent to understanding. Further, Minsky claims these frames can be applied across realms of thought to solve hard or novel problems and are therefore a possible explanation for what computational processes would be functionally equivalent to insights.

The second section of chapter two examines Ray Kurzweil's effort to explain understanding. The first entry point for this is Kurzweil's firm stance contra John Searle that adequate performance by a computer in a Turing Test, or comparable task, is evidence that the computer is capable of understanding. This is because Kurzweil holds that intelligence is ultimately a form of pattern recognition. This is a consequence of Kurzweil's theory of "patternism" according to which intelligence is a matter of being able to identify relations between facts. Kurzweil develops this into what he calls his pattern recognition theory of mind. Creativity is just a matter of facility in recognising previously unacknowledged patterns in data. Kurzweil identifies metaphor as an important process in which relations are found between previously disparate areas of knowledge. Such recognitions of similarity, are, Kurzweil thinks, pervasive in thought. Given his patternist outlook, Kurzweil interprets metaphor as the function

of a pattern recognition engine. For Kurzweil, the path to augmenting human intelligence is to create machines whose output in pattern recognition tasks is of high enough quality to justify ascribing understanding to them, and to subsequently expand the size and power of the pattern recognisers to achieve superhuman function, which on his terms translates to superhuman understanding and insight. These could then be made available to humans through connecting the human neocortex to the artificial cortex or copying the human mind into software and expanding its pattern recognition structures.

The third section of chapter two turns to Nick Bostrom's analysis of superintelligence and pathways to it. Like his predecessors, Bostrom considers it desirable to create machines that are superhuman in intelligence and finding a way to give humans access to such improvements. One of his proposals, pervasive in the transhumanist literature, is that a cryogenically frozen brain might be scanned and "uploaded" into a brain simulation program. The brain emulation proposal includes numerous implicit assumptions, such as that the dynamical processes of the brain can either be fully recorded or that they can be safely omitted without loss to a modelling of brain activity. Another challenge is that it is controversial that the identity of the emulation would be the same as that of the original person. It might be a successor or a copy but not satisfy the conditions of the original continued or renewed existence. If the brain could be scanned non-destructively, such that the original human could be alive while the scan is taken, then the issue of identity between the copy and the original becomes even more uncertain.

This section also highlights a potential issue with the relationship between humans and so-called "oracle AI" superintelligences. For those tasks where the outcome or bottom-line answer is not the sole concern, then the significance of the superintelligence is unclear. For many cognitive tasks, the goal is not merely producing correct outputs but having a deep

comprehension of something, as is the case when reading a philosophical treatise or a work of literature. A superior intelligence might be able to comprehend these deeply, but unless this comprehension can be communicated, then the significance of this superintelligence for human intelligence is unclear. Granted, if one assumes that these challenges can be met, then one is free to speculate on what sort of alterations or improvements that may be achieved.

Bostrom proposes three main routes to superintelligence, which all to a greater or lesser extent suppose that the brain emulation can be achieved and the challenges to it met. These three forms of superintelligence are speed superintelligence, collective superintelligence, and quality superintelligence. The core idea of speed superintelligence is that a brain emulation might be accelerated to allow intellectual work to be done at a markedly faster rate, such as composing a doctoral thesis in an afternoon. Speed superintelligence would accelerate insights by enabling an individual to have insights at their usual subjective rate while performing more thought at an objectively higher rate. While this approach would, in this scenario, successfully accelerate insight, it does not concern itself with the content or circumstances of acts of insight. It simply proposes that the cognitive process could be unaltered but made more time efficient.

Bostrom's second proposed route to superintelligence is collective superintelligence. This suggestion comes in three distinct varieties. The first version of collective superintelligence increasing the population by factors of ten, for instance by running many brain emulation instances on specialized hardware (such that one is not limited by the carrying capacity of the Earth for biological human being) would facilitate a greater quantity and rate of insights by increasing the number of thinkers including uncommon geniuses, even though the base rate of insights per capita would have remained constant.

The second version of collective superintelligence is through improved communication protocols, social technologies, and administration. The core idea is that effective

communication and collaboration increases the rate at which insights occur within a given population, though Bostrom does not discuss specific recommendations for improving communication nor what conditions and contexts of communication make it conducive to insight. This thesis will explore the topic of how communication facilitates insight in chapter four when considering the catalytic effects of second-person relatedness.

The third version of collective superintelligence is the suggestion that a sufficiently complex and integrated network of humans and machines could in aggregate satisfy the conditions of being a mind. Bostrom's suggestion, which he acknowledged as highly speculative, is that the internet might be eventually "wake up" as a sapient entity. Leaving aside the question of whether it is possible for an aggregate of humans and computers to form a distinct entity capable of cognition, several questions may still be raised about this scenario. It is not entirely clear how a partial understanding of some object by several humans, in which each grasps only a part, could together constitute a single, whole understanding of the object. Further, if the collective intelligence does come into existence as a single and distinct entity, can the constituent humans still be said to have their own independent existence and understanding, or have they been assimilated into the whole and had their own understandings obliterated? Finally, if it is possible for the collective intelligence and the constituent intelligences composing it to exist simultaneously as distinct entities, then what is the significance for human understanding? It is not clear that the understanding of the collective entity would be automatically ascribable to constituent humans (or vice versa) unless, as in the case of oracle AI discussed above, the understanding is communicated from the collective to the humans composing it.

The final version of superintelligence Bostrom explores is what he terms quality superintelligence. This proposed path to superintelligence is the most nebulous and expressed

by analogy. A quality superintelligence would have cognition qualitatively superior to that of humans in the same sense that human intelligence is qualitatively superior to that of chimpanzees. The exact nature of a qualitative superintelligence is not proposed, although Bostrom does give some hint of what such an entity might be capable of in the case of an intelligence that can imagine an object with two hundred dimensions, for instance. It is proposed that such an object would be capable of gaining understandings that elude human beings.

This thesis observes that neither speed, collective, nor quality superintelligence are helpful to accelerating insight. The idea of a quality superintelligence at least takes some account of the issue of understanding itself, but it does not ultimately go beyond narratives about cognitive abilities readers cannot imagine, such as imagining 200-dimensional objects. As such, the proposed paths to superintelligence do not illuminate either what may accelerate insight or even what insight is.

The second chapter's investigation of the three prominent transhumanists Marvin Minsky, Ray Kurzweil, and Nick Bostrom shows that all three agree that what is termed understanding, and those changes in understanding termed insight, are important for cognition and that it is at least superficially distinct from subjective stepwise reasoning. They also agree that it is an important target for intelligence enhancement, either through creating tools capable of superior understanding, or, preferably, by making it possible for humans to transfer their minds to machines so the improvements made to machines can be transferred to themselves. In their analyses of insight and understanding, they share an approach that measures success in terms of correct output, although they do make some reference to cognitive phenomena that are less straightforward to fit into this paradigm, such as imagination, inarticulate knowledge, or the knowing associated with meditative experience. Evaluating the prospects for accelerating insight within the paradigm of Minsky, Kurzweil, and Bostrom requires assessing their

paradigm's treatment of insight and understanding, but this analysis requires first considering the paradigm for insight and understanding that have been given in other philosophical and theological traditions, which is the purpose of chapter three. Chapter four will compare the transhumanist and philosophical-theological paradigms for insight.

The third chapter examines insight and understanding from philosophical and theological perspectives supplemented by psychology and linguistics. The first-personal aspects of insight, what is denoted when one says, "I see", or "I understand", are discussed first.

The first section of chapter three is prefatory and offers a distinction between philosophical and theological anthropology. The former concerns that which can be known through natural human reason, and the latter concerns that which is known through faith. As an example of this distinction, Aquinas' article concerning the complementary relation between knowledge of God's existence through natural reason and knowledge of God's existence through faith is discussed. Further, theology can demonstrate that faith is philosophically fitting, as in the case of the doctrine of the resurrection of the body being philosophically fitting because to be without the body is, in one sense, unnatural to the human soul. This clarifies the purpose of the theological perspective in this thesis as having observations on human insight distinct from but congruent with the philosophical perspective on insight and the lessons from neurobiological science, psychology, narrative, and metaphor.

The second section of chapter three explores the definitions of insight and understanding. Definitions of understanding are explored first, because, as discussed above, a change in understanding is what is denoted by the term insight. The section observes that understanding is defined in terms of grasping and seeing. These metaphors suggest that embodiment is particularly important to understanding, which research into childhood development supports. After this groundwork is established, the topic of insight is discussed as

a change in understanding. The thesis takes these terms in this order because one cannot expound on a change in a faculty without first establishing what the change occurs within. After this groundwork, Lonergan's analysis of a paradigmatic case of insight, that of Archimedes' "Eureka!" is discussed, and lessons about the general properties of an act of insight, of the moment of transition. This section also clarifies how the concept of insight used by Lonergan relates to Thomas Aquinas' complex system of the intellectual virtues in anticipation of reference to Aquinas in chapters four and five.

The second section of chapter three continues its examination of first-personal dimensions of insight by examining its grounding in embodiment through the lenses of metaphor and narrative. The section begins with reflections on the tacit or implicit dimension of knowing in child development, suggesting that acts of insight and understanding are prior to language. The relationship between insight, embodiment, and language is explored through examining Ian McGilchrist and Lakoff et al. on the role of metaphor in grounding concepts and their articulation in language and the role of metaphor for insights into new concepts. This analysis has the consequence that not only is understanding more fundamental than language use, but that all abstraction is made possible through the use of metaphor, which links embodied experience, understanding, and the concepts understood. Metaphor is shown to be important to mathematical insight as well. The analysis extends to the role of narrative in expressing and communicating insights through expressing and engaging imaginative kinaesthetic experiences and concepts. Narrative is particularly privileged for enabling an imaginative sharing of experience with the depicted characters allowing the values and affective responses comprising a character's perspective and the narrative's presentation of them to be shared.

Having concluded the analysis of the first-personal dimensions of insight, the third section of chapter three investigates the second-personal dimensions of insight. The section

begins by presenting Polanyi's view of what he calls tacit knowledge, that which is known but not articulate, in the context of sharing knowledge in an apprenticeship, in which the student through their relationship to their teacher learns to share the teacher's own attitudes and connoisseurship of the object of their attention. The importance of second-personal relations for sharing understanding gains support from the observations of MIT professor Alex Pentland on the sharing of understanding and discovery of insights in the context of social networks. This subsection provides a philosophical view of the importance of second-personal relatedness for sharing insights and empirical evidence that second-personal relatedness has a catalytic effect on insights.

The next subsection provides further evidence of the importance of second-personal relatedness for insights but drawing from psychological studies of joint attention in infant development. Joint attention may be characterised as shared attention or focus, and it manifests in infants as interacting with caregivers and following their gazes. This behaviour is crucial to young children learning what in the world is worth paying attention to and how they should orient themselves toward it, seeking to understand and align with the perspectives and stances of their caregivers. When second-personal relatedness is disrupted, as in the case of severe autism, developmental delays ensue. The psychological concepts of joint attention or shared awareness are identified as bolstering the Polanyian sense of the role of personal relations in sharing understanding.

Having considered insight philosophically, the section four of chapter three considers insights into second-personal relatedness drawn from theology. The first subsection explores the ancient metaphor of the divine as light that illuminates the intellect. The subsection considers the specifically biblical perspective in which the encounter with God is described not only as an illumination but as a face-to-face meeting, in fact a personal encounter leading to

insight. This biblical perspective is elaborated by reference to Aquinas' commentary on the Word of God and the possibility of sharing in it. From this, it may be said that the highest kind of insight achievable on the Christian perspective is that of encountering the divine personally.

The second subsection examines biblical narratives drawing on Elaine James' analysis of biblical poetry and Robert Alter's study of biblical scripture to show that biblical narrative invites imaginative and sensorial engagement with the texts both for pedagogical purposes and to draw the reader into the shared life of the community to whom the narratives belong. The subsection then considers Christ's use of parables as a pedagogic method to stimulate insight through engagement with narrative. The ultimate purpose of biblical narratives and Christ's parables is identified as to draw the soul into loving God with one's understanding. From scripture and Christ's own choice to instruct through parables the lesson is drawn that narrative is a privileged means of communicating insight apart from a mere sharing of explicit information.

The third subsection examines two paradigmatic cases of person encounter and transformation in scripture: the narrative of the disciples meeting Christ on the road to Emmaus and the conversion of St. Paul. The first case illustrates the account of pedagogy described above, in which the teacher directs the pupils' attention toward its correct objects, likely to spark insight, while instilling the correct affective responses to all they are being shown, culminating in suddenly and correctly seeing the person of Christ. The second case, that of St. Paul, demonstrates a dramatic instance of sudden insight reordering one's entire perspective and character. It has the added interest of illustrating a different kind of divine pedagogy, that of receiving understanding directly from God.

The fourth subsection discusses Aquinas and the gift of understanding, according to which the soul is given by God insight into the life of virtue as distinct from a natural exercise

of one's faculties. This insight can be described as relational, as the soul is moved to align with God. The gift of understanding can be described in second-personal terms according to the metaphor of joint attention between the soul, God, and the objects of understanding. Again, theology affirms the catalytic power of second-person relatedness.

Chapter three concludes by drawing out five key themes with respect to insight: insight as a singular ascending operation, insight as grounded in embodiment, the role of metaphor, the role of narrative, and the catalytic effect of second-person relatedness. These five key themes are used to structure the evaluation of how the transhumanists treat these five themes in their own work in comparison to that of the philosophical and theological perspective.

Chapter four evaluates the transhumanist perspectives along each of these themes. The first subsection of chapter four, considers whether the transhumanists consider insight to be a singular ascending operation, showing that they interpret insight experiences as subjectively unified but as actually a disparate collection of other mental processes that achieve a reformulation of definitions and representations. Within the transhumanist framework, there is not a kind of cognitive act that can be called insight but a set of different processes whose effects are similar such that they are naïvely labelled insight. In contrast to the philosophical and theological perspective of insight as an ascent to a higher viewpoint, the transhumanist perspective is that insight is exhausted by the reformulation of mental representations.

The second subsection of chapter four evaluates the transhumanist view of the cognitive significance of embodiment compared to the philosophical and theological view. This section shows that the transhumanists see the body as potentially offering some cognitive benefit, especially for informing the structure of mental descriptions and concepts, but that the body is nonetheless detachable from cognition in that one's particular body is replaceable, the kind of body one has can be changed, and having a body in principle is optional. A key assumption of

the transhumanists is identified, that everything pertinent to embodiment can be abstracted and simulated, and this section notes there are obstacles to perfect simulation of the world, the body, and the brain. This subsection sets in contrast to the transhumanist perspective Thomas Aquinas' on the unity of the human being as a rational animal, focusing on the role of bodily powers for the proper operation of the intellect. This section concludes by observing the contrast between, on the one hand, the transhumanist view that all aspects of embodiment must be abstractable and at some level equivalent to a perfect abstraction and, on the other hand, the philosophical view that embodiment is unified to intelligence, so to insight and understanding, such that intelligence relates to both abstraction and implicit understanding of the particular precisely through the totality of one's embodiment.

The third subsection considers the transhumanist perspective on metaphor compared to the philosophical perspective on metaphor. The transhumanist account is in agreement with the philosophical account on metaphor being pervasive in thought and the discovery of new metaphors being related to insights, but that the transhumanist and philosophical perspectives disagree significantly about what metaphor is. The transhumanist view of embodiment as fully abstractable leads to a description of metaphor in terms of abstract pattern matching. The philosophical perspective sees metaphor as the link between embodied experience and the understanding of language, indeed all representation.

The fourth subsection turns to the transhumanist perspective on the relation between narrative and insight. This examination shows the transhumanists characterise narrative as an ability to create structured representations of events or patterns through time, and that they are especially important both to structuring representations of regular experiences and contexts and to parsimonious communication by relying on mutual knowledge of narratives and their structure. The transhumanists also evince a practical understanding of the power of narrative to

shape imagination, train affections, and change minds, as Bostrom's own frequent use of parables shows. This subsection observes that, nonetheless, the transhumanist concept of narrative emphasizes narrative as an abstract pattern of information and does not consider the embodied, imaginative, and emotional engagement of persons with narrative that are explored by the philosophical and theological perspectives.

The fifth subsection considers second-person relatedness. The transhumanist view of second-person relatedness and its relevance to insight are brought out by considering first the treatment of pedagogy by Minsky and Kurzweil in contrast to the philosophical perspective and the extent of their engagement with second-personal aspects of communication in the enhancements to communication proposed by Kurzweil and Bostrom. The subsection argues that the pedagogical accounts of the transhumanists is restricted in scope, offering a model of education principally concerned with transmitting information and signalling flaws in mental models with the intention of it being amenable to later automation. In contrast, the philosophical perspective recognises the importance of personal engagement and alignment of perspectives in embodied interaction to apprenticeship. The transhumanists' proposed enhancements that illuminate their perspectives on second-person relatedness are then considered. Kurzweil proposes that intelligent machines will eventually serve as mediators between human communications to the point that humans would attend more to their intermediary machines than they would jointly attend to each other. From the philosophical perspective, Kurzweil's proposal is likely to disrupt the alignment of perspective between persons and the sharing and sparking of insights. Bostrom's different senses of collective superintelligence are revisited and shown, from the philosophical perspective, to be tangential to the question of how persons together produce insights.

The fifth subsection sharpens the evaluation of the transhumanists' perspective on second-person relatedness by considering their treatments of understanding gained in meditation and in the case of the beatific vision. Bostrom considers the theological concept of the beatific vision (as defined by Aquinas and with reference to St. Paul's description of it as seeing God face to face) as one possible inspiration for how utopian existence might be designed, but he reduces the beatific vision to a matter of knowing and loving the best possible object, thereby maximizing meaning points. This subsection presents the Thomistic view of the beatific vision, showing that beyond the possibility of achievement by any natural intellect, even a posthuman one, and consists not of meaning maximization but friendship with God. Kurzweil's and Minsky's assessments of the sense of understanding in meditation are also discussed. Kurzweil offers a vague affirmation of their value, but Minsky rejects them and senses of revelation for being without explicit content. The philosophical perspective can account for a sense of understanding that is implicit, and it can be theologically grounded in the notion of the Gift of Understanding. In summary, the transhumanist perspective on second-person relatedness and its relation to insight appears, from the philosophical and theological perspective, to be extremely limited.

Chapter five considers implications of the preceding analysis. The first section of chapter five considers the distinction between internal enhancements, those which modify the human organism directly, and external enhancements, which augment human powers without directly enhancing the human being. By referring to Aquinas on the relation between the corporeal powers and the intellect, it is acknowledged that there is some ground for speculations that internal enhancements that modify the organism, could lead to increased cognitive performance. These possibilities are not concrete, however. With respect to external enhancements, the transhumanist aspiration to creating machines capable of understanding are

doubtful for the reasons discussed above in the thesis. This subsection argues that while the transhumanist interpretation remains questionable, there nonetheless exist opportunities for using technology to bring otherwise inaccessible phenomena into the reach of embodied experience, and so into contact with human insight.

The second section of chapter five develops further non-transhumanist considerations of external enhancements of human cognition by outsourcing the work of insight when only the results of that insight matter rather than the actual act of insight itself, thereby freeing insight to be directed to worthwhile objects.

The third section of chapter five elaborates on the non-transhumanist external enhancement of using technology to find what may be termed missing insights where the discovery is valuable to understand and the machines can either accelerate the acquisition of insights or even make knowable truths that would otherwise be inaccessible to human enquiry.

The fourth section of chapter five considers the judicious use of communication technologies to facilitate second-person relatedness rather than to impede it.

1.4 Conclusion

The transhumanist perspective and the Aristotelian-Thomist perspective both acknowledge the reality of human limitations, physical and cognitive. The transhumanists aspire to eliminate both sets of limitations. In contrast, the Aristotelian-Thomist perspective

contends that those qualities that limit humans, such as embodiment, structure human cognition and allow means of overcoming these limitations without eliminating them, such as through second-personal relatedness. This thesis crucially offers an overview of transhumanist perspectives and brings them into dialogue with philosophical and theological perspectives, with the aim of showing the enriching benefits of the latter, and showing how the more holistic Aristotelian-Thomist view contributes significantly to this dialogue.

Both the transhumanist and Aristotelian-Thomist perspectives recognise the cognitive power called understanding, and those changes within it called insight, as of great cognitive significance. They differ, however, in their interpretations of understanding. The transhumanist account interprets cognition according to its physicalist framework using the structuring metaphor of the brain as a computer with the implication that insight may be, ultimately, abstracted into a set of algorithmic procedures. In contrast, the Aristotelian-Thomist perspective takes insight as the unitary act of an embodied human organism.

Comparing the relative merit of these contrasting perspectives on insight is not merely of academic interest. The question what insight is informs answers to questions of how to facilitate insight and cognitive flourishing, as well as the significance of technology, especially computation and machine learning, for human cognition. If the transhumanists are correct, then technology as tool of human beings is only a transitional stage to the emergence of posthuman machine intelligence. If the transhumanist perspective appears thin in comparison with the philosophical-theological perspective, however, then the transhumanist narrative may be cast into some doubt, and an alternative perspective on the relation of human intelligence and machines may be welcome.

This thesis demonstrates that, ultimately, the transhumanist prospects for accelerating insight are limited but that an Aristotelian-Thomist perspective both provides a richer account of what insight is and can effectively inform usage of technology to benefit insight.

CHAPTER 2 Transhumanism and the Challenge of Understanding

2.1 Marvin Minsky: Frame Language, Knowledge, and Insight

2.1.1 Introducing Marvin Minsky

Marvin Minsky (1927-2016), co-founder of the Artificial Intelligence Laboratory at MIT, wrote on transhumanist themes before the term “transhumanism” was widely known and helped to shape many of its priorities and expectations.¹⁷ He contributed greatly to the development of programming and AI and advocated for human life-extension, cognitive enhancement, and the value of transcending biological limitations.

2.1.2 Minsky on Understanding and Frames

In this section, I explore his pertinent research and speculations about understanding and insight.¹⁸ First, I examine how Minsky uses the term understanding within his project to engineer intelligence in machines and, which is the same to him, to reverse engineer intelligence in humans. Second, I explore his effort to create a model for cognition that can account for

¹⁷ Minsky was a signatory of an open letter on cryonics along with prominent transhumanists including Nick Bostrom, Julian Savulescu, and Aubrey de Grey, to name a few. “Scientists’ Open Letter on Cryonics – Biostasis,” *Biostasis* (blog), accessed January 23, 2019, <https://www.biostasis.com/scientists-open-letter-on-cryonics/>; Marvin Minsky, “Will Robots Inherit the Earth?,” *Scientific American*, October 1994, <https://web.media.mit.edu/~minsky/papers/sciam.inherit.html>; Marvin Minsky, “Afterword to Vernor Vinge’s Novel, ‘True Names’,” Marvin Minsky’s Homepage, October 1, 1984, <https://web.media.mit.edu/~minsky/papers/TrueNames.Afterword.html>. Minsky was a signatory of an open-letter on cryonics along with prominent transhumanists including Nick Bostrom, Julian Savulescu, and Aubrey de Grey, to name a few. “Scientists’ Open Letter on Cryonics – Biostasis,” *Biostasis*, n.d., accessed January 23, 2019, <https://www.biostasis.com/scientists-open-letter-on-cryonics/>; Marvin Minsky, “Will Robots Inherit the Earth?,” *Scientific American*, October 1994, <https://web.media.mit.edu/~minsky/papers/sciam.inherit.html>; Marvin Minsky, “Afterword to Vernor Vinge’s Novel, ‘True Names’,” Marvin Minsky’s Homepage, October 1, 1984, <https://web.media.mit.edu/~minsky/papers/TrueNames.Afterword.html>.

¹⁸ “In the 1970s and 1980s, several organizations sprung up for life extension, cryonics, space colonization, science fiction, media arts, and futurism. They were often isolated from one another, and while they shared similar views and values, they did not yet amount to any unified coherent worldview. One prominent voice from a standpoint with strong transhumanist elements during this era came from Marvin Minsky, an eminent artificial intelligence researcher.” Nick Bostrom, “Transhumanist FAQ,” Humanity+, accessed October 16, 2019, <https://humanityplus.org/philosophy/transhumanist-faq/>.

understanding. Finally, I present how Minsky uses the term insight and show that while his usage is sometimes atypical, he is nonetheless aware of the importance of the discovery of new understandings, which is what is usually signified by the term insight.

Marvin Minsky begins his preface to *Semantic Information Processing* with the question,

How can one make machines understand things?¹⁹

He does not answer the question immediately but offers a definition of artificial intelligence research as

the science of making machines do things that would require intelligence if done by men.²⁰

The implication is that at least one of the measures of success, for Minsky, in the field of artificial intelligence research, is that machine outputs become equivalent to what would typically require human intelligence.

Minsky does, however, comment elsewhere in more detail on the topic of intelligence. He writes in *The Society of Mind*, for example,

An author's job is using words the way other people do, not telling others how to use them. In the few places the word "intelligence" appears in this book, it merely means what people usually mean—the ability to solve hard problems.²¹

According to this passage, intelligence is ascribed to beings that "solve hard problems", the implication of which, again, is that if a being, whether natural or artificial, can solve hard problems then intelligence can reasonably be ascribed to such a being.

What is striking in the statements above is that Minsky does not refer to understanding explicitly in his comments on intelligence despite that fact that understanding and intelligence

¹⁹ Marvin Minsky, *Semantic Information Processing* (MIT Press, 1968), v.

²⁰ Minsky, *Semantic Information Processing*, v.

²¹ Marvin Minsky, *The Society of Mind*, Touchstone (New York, NY: Simon & Schuster, 1988), 71.

are commonly held to be related.²² Elsewhere, however, Minsky does examine, to some extent, what is meant by the term “understanding” and its relation to intelligence, for example, in his comment that emerging cognitive science,

shows that many phenomena traditionally described by commonsense (sic.) terms like “intelligence” or “understanding” actually involve complex assemblies of machinery.²³

He indicates here a theme of his work, namely that a rigorous account of human cognition will show that understanding actually consists of an assemblage of discrete operations and, moreover, that these operations can be, and ultimately should be, described in a machine-like way. On this account, it is part of common sense that the term “understanding” is used to refer to a particular cognitive experience, a “folk model” that expresses the subjective experience of cognising some solution, but does not articulate or engage with the complex, machine-like operations of attaining this solution.

Minsky’s approach here is consistent across his work, namely that terms like “understanding” are used as terms of convenience to cover complex assemblies of cognitive operations. He describes the prosaic example of understanding a word in the following fashion:

A simple-seeming word like “rock” engages a huge society of systems in a brain including, for example, many kinds of visual, tactile, and haptic representations — which each in turn activate other networks of relationships among other structures, processes, and memories. But none of that is directly sensed by our agencies involved with goals, plans, speech, and “consciousness” — which just because of being more

²² I note in passing that although Minsky claims that solving hard problems is a common, if vague, definition or at least indication of intelligence, a perusal of the *OED*’s entry on “intelligence” does not cite any references to an ability to solve hard problems. Representative entries in the *OED* define intelligence as “the faculty of understanding... a mental manifestation of this faculty” (entry 1), “The action or fact of mentally apprehending something; understanding, knowledge, comprehension (*of something*)” (entry 3), and an “Exchange of knowledge, information, opinion, etc.; mutual conveyance of information; communication, esp. of confidential information” (entry 7a). For further information, see “Intelligence, N.” in *OED Online* (Oxford University Press), accessed February 25, 2019, <http://www.oed.com/view/Entry/97396>. “Intelligence, n.,” in *OED Online* (Oxford University Press), accessed February 25, 2019, <http://www.oed.com/view/Entry/97396>. “Intelligence, n.,” *OED Online* (Oxford University Press), accessed February 25, 2019, <http://www.oed.com/view/Entry/97396>.

²³ Marvin Minsky, “Society of Mind: A Response to Four Reviews,” *Artificial Intelligence* 48, no. 3 (1991): 371, [https://doi.org/10.1016/0004-3702\(91\)90036-J](https://doi.org/10.1016/0004-3702(91)90036-J).

“articulate” — can only deal with summaries. Paradoxically, it is because our higher-level systems know so little of how their massive infrastructures work that we feel free, in everyday life, to use enveloping terms like “understand”. It’s just *because* such terms “make sense” that they impair all our attempts to penetrate the systems that they envelop.²⁴

In other words, on this account, “understanding”, which is used as a term of convenience that implies a unitary cognitive act, in fact embraces or covers over a massive assemblage of processes that are not directly experienced but which in fact generate the outputs that are ascribed to understanding. Specifically, for Minsky, instances of what would commonly be called understanding are actually “summaries” of the outputs of huge “societies” of systems.²⁵ These societies involve multiple types of relationships in which the representational elements trigger one another. Minsky also observes that even acts of seemingly straightforward reference to objects do involve embodiment in some sense, albeit in the form of representations.

To explain how knowledge can be structured and represented in the brain, facilitating perception, language comprehension, memory and even creativity, Minsky proposed the theoretical existence of a mental structure called a “frame”, to which I now turn.

A key theme in Minsky’s work is that experience and knowledge are structured by *frames* which, like picture frames in art, demarcate sets of appropriate details from the broader context. The following passage gives a general description of a frame’s function and structure:

When one encounters a new situation (or makes a substantial change in one’s view of the present problem) one selects from memory a structure called a frame. This is a remembered framework to be adapted to fit reality by changing details as necessary... The “top levels” of a frame are fixed and represent things that are always true about the supposed situation. The lower levels have many terminals—“slots” that must be filled by specific

²⁴ Minsky, “Society of Mind,” 389–90, emphasis original.

²⁵ His willingness to use the term himself, as when remarking in his seminal essay *A Framework for Representing Knowledge* that the simple act of “listening often involves extensive and difficult constructions –which involve the totality of complexities we call understanding” indicates that he does not think the term is vacuous. Instead, the term refers to a complex processes and structures which it is the role of psychology to describe and AI research to realise as theories of machine design. Marvin Minsky, “A Framework for Representing Knowledge,” in *The Psychology of Computer Vision*, ed. Patrick Henry Winston (McGraw-Hill Inc., 1975), 238.

instances or data. Each terminal can specify conditions its assignments must meet. (The assignments themselves are usually smaller “sub-frames.”)²⁶

To give an example of what Minsky means in this passage, there may be a frame that represents the object “dog”. Actual dogs have certain inherent or essential attributes, such as “furry”, “barking”, and “quadrupedal”; other attributes are more variable, for example, all dogs have colour, but some dogs may be brown while others may be black. For Minsky, the essential attributes are organised at the top levels of a frame, while the bottom levels have many “terminals” such as, in this example, “colour”, into which specific attributes can be recorded.²⁷

Frames are used for dealing with some situation, and a change of situation brings with it a change of frame. An appropriate frame is activated, because it is in some way optimized to select details salient to a situation, each of which corresponds to a given terminal. As frames must be of a finite size and the details they recognise must be discrete, the implicit assumption is that there must be a finite, enumerable set of details in any situation, the conjunction of which are recognised.²⁸ Further, some of these details are considered “top-level”, and their recognition is a necessary condition of the activation of the frame as a whole. The necessary presence of such details contrasts with the non-necessary presence of details recognised by the “terminals” of frames. For many frames, the input to the terminals is supplied by other frames arranged in a hierarchy.²⁹ It is worth reiterating that in Minsky’s scheme agents possess only “summaries”

²⁶ Minsky, “A Framework for Representing Knowledge,” 212.

²⁷ Minsky is not unaware that there exist exceptional cases, such as three-legged dogs due to injuries, say. Frames, he thinks, are flexible insofar as frames that do not perfectly fit a scenario can still be used if no other ones are available, and they can actually be incorporated into a larger mental structure that includes exception handling frames that respond to evidence of injury, say, and strengthen the activation of the original frame when they are also activated. Minsky, 250–51.

²⁸ Cf. Hubert L. Dreyfus, *What Computers Still Can’t Do: A Critique of Artificial Reason*, Rev Ed edition (MIT Press, 1999), 35–40.

²⁹ Frame structures are similar to the concept of classes in object-oriented programming, in which objects in a given class also inherit traits from parent classes. Hubert Dreyfus also sees some similarity between Minsky’s notion of frames and Aristotle’s notion of categories, namely, that objects are classified by essential properties and

of one another's content, so frames, being agents themselves, do not contain the content of the frames to which they are connected.

Many other examples of frames can be found in the recognition of spatial objects. For instance, the frame-system for recognising a cube contains many frames capable of recognising it from different angles. Each frame's terminals respond to parts of the percept as the sides, edges, and points of the cube. If a sufficient number of terminals are activated, the frame as a whole will in turn be activated, identifying the percept.³⁰ The activation of this frame primes the other cube frames in the system to become activated should the perspective on the cube shift.³¹ In other words, Minsky's theory of frames attempts, at least implicitly, to explain how it is that humans understand what it is that they see, namely, by getting a positive signal from a system of input classifiers.

As in understanding visual perceptions, so in understanding language. Linguistic structures, he thinks, exist in the mind as nested or related frames. He suggests that linguistic frame-systems may be classed as recognising syntactic, semantic, thematic, and narrative structures (although he does not intend this list to be exhaustive), but the general structure of a frame is sufficiently homogenous that a few examples will suffice to illustrate how Minsky regards frames as explaining linguistic understanding.³²

A syntactic frame-system has terminals that are satisfied by particular letters, and these frames are in turn subframes (that is, they activate the terminals of other frames) for word-

that the classes of which they are members may be nested within largest categories. Dreyfus, *What Computers Still Can't Do*, 231–34.

³⁰ Not all features that terminals identify are of equal value. "Key" features of a given scene or concept will be weighted more heavily than transient or changeable features. For instance, an eight-by-eight checkered board is a key feature of chess but not the traditional coloration of the squares.

³¹ Poincaré divided the process of original discovery into the four steps of preparation, incubation, revelation, and evaluation. Of present interest are Minsky's remarks on the nature of incubation and revelation. Henri Poincaré, *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method*, trans. George Bruce Halsted (2012), 487–503, <http://www.gutenberg.org/ebooks/39713#download>.

³² Minsky, *The Society of Mind*, 198, 265.

recognising frames which take letter frames as their inputs. Word frames in turn are subframes of syntactic frames, which recognise grammatical structures.

Syntactic frames in turn provide input to semantic frames, allowing for sentential meaning to be recognised. These semantic frames are in turn the subframes of higher frames for linguistic comprehension, such as thematic or narrative frames. The salient points are that Minsky's frame theory arranges frames into perceptual and conceptual hierarchies and each member of the hierarchy gets only a "summary" of the information provided to it by other frames.

Structured relations between frames, which Minsky calls frame-systems, give rise to the subtlety of human knowledge and play an integral function in problem solving. At an early stage of his theory's development, he sketched some roles that frame-systems could play in human reasoning as follows:

For non-visual kinds of frames, the differences between the frames of a system can represent actions, cause-effect relations, or changes in conceptual viewpoint.³³

On Minsky's account, frames do not merely make sense of perception, they also permit the representation of the complexity of change in the world, and, crucially for this analysis, shifts in conceptual viewpoint. This conceptual viewpoint is what concerns us in making our analysis of what constitutes "understanding" for Minsky.

The interframe structure makes it possible, according to Minsky, for the brain to respond robustly to non-rigid inputs.

When a proposed frame cannot be made to fit reality—when we cannot find terminal assignments that suitably match its terminal marker conditions—this network provides a replacement frame. These inter-frame structures

³³ Minsky, "A Framework for Representing Knowledge," 212.

make possible other ways to represent knowledge about facts, analogies, and other information useful in understanding.³⁴

Frame assignments are controlled by these interframe systems, ultimately activating a frame whose expected input values and representations of those values best suit the current situation. Minsky proposed that these structures of frames were key components of creative problem solving.

To sum up, at least part of what is meant by the term “understanding” is covered by Minsky in his concept of frames. Frames identify that a thing can be classified as a particular type of thing, the particularity of which is defined in terms of the conjunction of a set of inherent, top-level attributes and a multiplicity of “terminals” that can take on diverse values.

Minsky uses the related term “insight” in connection with the term “understanding” and addresses some of the issues associated with this term in various works, notably on creative thinking in both *Society of Mind* and in his chapter “Thinking” in *The Emotion Machine*.

In his passage on “in-sight” (sic.) in *The Society of Mind*, Minsky remarks,

Each human cranium contains hundreds of kinds of computers... No one knows how many different such organs lie in our brains. But it is almost certain that they all employ somewhat different types of programming and forms of representation... If a mind whose parts use different languages and modes of thought attempted to look inside itself, few of those agencies would be able to comprehend one another...³⁵

The significance of this suggestion, he says, is that,

When our higher-level agencies attempt to prove the fine details of the lower-level submachines that they exploit, they cannot understand what’s happening.³⁶

We lack “in-sight” in Minsky’s sense insofar as our higher faculties, including our conscious awareness, are unable to make sense of the processes or representations of lower-level

³⁴Minsky, “A Framework for Representing Knowledge,” 213.

³⁵ Minsky, *The Society of Mind*, 66.

³⁶ Minsky, *The Society of Mind*, 67.

processes. As a consequence, any subjective act of understanding will appear basic because its component cognitive acts are inaccessible to the conscious mind. To this extent, Minsky thinks that the apparent unity and simplicity of the phenomena called “insight” and “understanding” are due to a lack of awareness of or explanation of cognitive operations that are somehow lower level than consciousness. Within his view, a complete account of cognitive processes will demonstrate that apparent simplicity of understanding and insight is illusory, and that they are actually complex assemblages of mechanisms for managing and processing representations.

It should be noted that in Minsky’s writing “insight” is used in a possibly peculiar way. Minsky uses “in-sight”, to refer to awareness of supposed components of cognitive acts, especially those described as acts of understanding. Minsky is somewhat aware of the more common uses of the term “insight”, but he deprecates them as follows,

We often use words like “insight” or “intuition” to talk about understandings that seem especially immediate. But it is bad psychology to assume that what seems “obvious” is therefore simple or self-evident. Many such things are done for us by huge, silent systems in our mind, built over long forgotten years of childhood. We rarely think about the giant engines we’ve developed for understanding space, which work so quietly that they leave no traces in our consciousness.³⁷

In this passage, Minsky uses the term insight to refer to a specific kind of understanding, namely those which feel immediate and obvious. Here, Minsky explains this experience of immediacy as the result of having forgotten past conscious problem solving and simply being unaware of the systems trained to solve these problems, such as those of how to navigate in space.

Minsky is also at least somewhat sensitive to the common usage of insight to describe those experiences of sudden understanding that appear to involve discovery or creativity.

The best way to solve a problem is to already know a solution for it—and this is why commonsense (sic.) knowledge is useful. But what if the problem is one you have never seen before? How can you continue to work

³⁷ Minsky, *The Society of Mind*, 128.

when you lack some of the knowledge you need? The obvious answer: you just have to guess—but how does one know how to make a good guess? We usually do this so fluently that we have almost no sense of how we are doing it, and, if someone asks about that, we tend to attribute it to mysterious traits with names like intuition, insight, creativity, or intelligence.³⁸

In this text, Minsky describes the term insight in terms of gaining new knowledge, albeit in some way that is experientially mysterious. Moreover, his use of the word “guess” also implies that whatever insight provides remains at least somewhat suspect or vulnerable to error. In addition, drawing from previously cited texts, it is also clear that Minsky attributes to systems that are invisible to the subject the operations that give rise to insights.

This picture of what is usually referred to as insight is explored in Minsky’s chapter “Thinking” in *The Emotion Machine*, which is about the exploration of the genesis of new ideas, beyond simply the outcome of logical reasoning processes. To explore this topic, Minsky finds it helpful to examine and cite Poincaré’s accounts of his own experiences of seeking solutions for mathematical problems. These accounts highlight how, after initial and fruitless work on difficult problems, Poincaré found that solutions would suddenly come to him unbidden and almost fully formed in outline.³⁹ Minsky finds these descriptions fascinating and examined how they may be reconciled with his own account of framing and the society of mind.⁴⁰ In particular, Minsky posited that the apparent suddenness of insight resulted from a hidden assemblage of mechanism working on the problem without conscious awareness and finally presenting their results to the conscious mechanisms of the mind for further checking.

³⁸ Marvin Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Reprint Edition (Simon & Schuster, 2007), 205.

³⁹ Poincaré divided the process of original discovery into the four steps of preparation, incubation, revelation, and evaluation. Of present interest are Minsky’s remarks on the nature of incubation and revelation. Poincaré, *The Foundations of Science*, 487–503.

⁴⁰ Minsky, *The Emotion Machine*, 235–40.

In contrast to deliberative reasoning, which is conscious and involves enumerable methods, the kind of thought Poincaré describes is opaque. One only sees its fruits when insight finally arrives in what Poincaré calls revelation. Minsky acknowledges,

...we still know almost nothing about how our brains could conduct such a search [for a solution] ...⁴¹

Minsky regards insight here as a solution to a search problem. Although Minsky is here speaking of mathematical discovery, it coheres with a suggestion in *The Society of Mind* that a key quality of intelligence is efficiency in searching a solution space.⁴²

What would Minsky consider to be examples of insights, and what is their likely structure? To return to the topic of frames discussed in *Society of Mind*, Minsky suggests that the interframe structure can allow entire families of frames to be applied in new contexts.

Such cross-realm correspondences can enable us to transport entire families of problems into other realms, in which we can apply to them some already well-developed skills.⁴³

On this account, in other words, insightfulness is reduced to the use of already developed problem-solving abilities in novel contexts. Once again, the quest for and fruits of insight are reduced to the search of a solution space.

⁴¹ Minsky, *The Emotion Machine*, 238.

⁴² Minsky attempts to illuminate his definition of intelligence with what he refers to as the “puzzle principle.” For any given problem in which a solution is defined, a machine can be constructed that will through random trial eventually discover the solution. Such a machine would be able to solve simplistic problems quickly but would take an arbitrarily long time to solve any problem with a sufficiently large space of possible solutions, so, Minsky supposes, a machine capable of solving “hard problems” in a reasonable timeframe would use heuristics to reduce the space of possible solutions to make searches more manageable. Minsky takes such considerations to “help, philosophically, to replace our feeling of mystery about creativity by more specific and concrete questions about the efficiency of processes.” In other words, Minsky regards his thought experiment not as answering any arguments about the nature of creativity and its role in solving problems, but merely as dispelling affects or intuition that would take such creativity in thought as something inexplicable in mechanistic terms. Minsky, *The Society of Mind*, 73.

⁴³ Minsky, *The Society of Mind*, 299.

2.1.3 Concluding Remarks on Minsky

To sum up, at least some of which is meant by the terms “understanding” and “insight” are covered, in Minsky’s conceptions, by the work done by framing and changes of frame. But it should be observed that understanding is dealt with principally in terms of frame selection, in other words the classification of objects as particular types of things. According to his conception, although not everything that we know when we understand is consciously accessible, all of it is likely contained in some sort of data structures in which all things, their attributes, and relationships between them are directly represented. For Minsky, to illuminate the terms “understanding” and “insight” requires explanation in terms of mechanism and explicit representation. Whether or not such an approach is sufficient is a question to which I shall return in chapter four, after exploring the philosophical and theological dimensions of understanding and insight in chapter three.

The next transhumanist to which we turn is Minsky’s student and great populariser of transhumanist narratives, Ray Kurzweil.⁴⁴

2.2 Raymond Kurzweil: Pattern Recognition and Understanding

2.2.1 Introducing Ray Kurzweil

Ray Kurzweil (1948 –) is perhaps the most well-known advocate of transhumanism in the contemporary sense. Kurzweil has impressive credentials as an inventor and has earned the respect of Bill Gates—who praised Kurzweil’s prescience in predicting technological trends—and the support of Google—which employed him to develop his theory of the neocortex as

⁴⁴ Ray Kurzweil studied computer science under Marvin Minsky at MIT as a Bachelor of Science. “Ray Kurzweil - Director of Engineering, Google Inc., Alphabet Inc.,” *Boardroom Insiders Profiles*; *San Francisco*, July 10, 2017, <https://search.proquest.com/docview/1933131584/abstract/6DB48C598DA44A84PQ/9>.

primarily a pattern-processing machine. Kurzweil's theories on understanding and insight were influenced by those of Marvin Minsky, of whom he was a student, and he has a particular focus on how technology will expand the human neocortex. To sum up, Minsky largely provided the theoretical underpinnings of a transhumanist view of understanding, and Kurzweil offered a vision for how the faculty of understanding may be enhanced.⁴⁵

Kurzweil has won an appreciative audience for his public pronouncements on technological trends, in part due to his career as an inventor with multiple impressive technical and commercial successes. His inventions include the Kurzweil reader for the blind, an early text-to-speech reading device relying on visual recognition,⁴⁶ and early speech-to-text systems.

He is also an enthusiastic optimist regarding prospects for AI, anticipating the creation of "superintelligent" AI capable of recursive self-improvement, exponentially increasing its own computing power. He predicts that humans, too, will reap the benefits of technological developments leading to increases in their own intelligence.

Kurzweil's predictions about technology, based in part on work with a team of analysts, are therefore received with considerable interest. Indeed, Bill Gates endorsed his 2014 book *How to Create a Mind* with the assessment that

⁴⁵ "My project is to get the Google computers to understand natural language, not just do search and answer questions based on links and words, but actually understand the semantic content. That's feasible now." David J. Hill, "Exclusive Interview: Ray Kurzweil Discusses His First Two Months At Google," *Singularity Hub*, March 19, 2013, <https://singularityhub.com/2013/03/19/exclusive-interview-ray-kurzweil-discusses-his-first-two-months-at-google/>; Kurzweil's willingness to use the term understanding to apply to algorithms successfully operating in narrow domains is not unusual among transhumanists. Consider another influential transhumanist, Hans Moravec. He writes, "There are semi-intelligent programs that can understand simplified typewritten English about restricted subjects and make elementary deductions in the course of answering questions about these texts." Like Kurzweil, Moravec treats an ability to produce correct outputs for some set of inputs as sufficient to warrant using the term understanding. Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, Reprint edition (Harvard University Press, 1990), 9.

⁴⁶ Kurzweil's invention was the software that could convert an image of text into a text file. Third party text to speech software then rendered the file audible.

Ray Kurzweil is the best person I know at predicting the future of artificial intelligenc.⁴⁷

Kurzweil's own work on pattern processing algorithms earned him his position as a director of engineering at Google with the stated goal of trying to imbue Google's algorithms with understanding.⁴⁸ Indeed, Kurzweil has publicly said that Google's algorithms now do in fact understand language.⁴⁹ From the point of view of this thesis, Kurzweil is important, first, because of his prominence in the transhumanist field, and second, because, somewhat unusually for contemporary transhumanists, he pays some attention to the meaning and importance of understanding.

2.2.2 Kurzweil on Understanding and Pattern Recognition

This section will examine Kurzweil's uses of the term "understanding". Kurzweil does not treat systematically of the nature of understanding or of insight, but they are themes in his primary topics of interest. This thesis considers three themes that are paradigmatic of his treatment of understanding and insight in reference to his principal and overall aim, namely the goal of human enhancement. First, I look at the views Kurzweil expressed regarding the Turing Test. Second, I examine Kurzweil's treatment of what he calls the pattern recognition theory of mind insofar as it serves as a proxy for understanding. Finally, I look at Kurzweil's "techno-optimism" given that these principles are so crucial to his general projection of the path of future development. I begin by looking at Kurzweil's views on the Turing Test, views that became important during his exchange of views with John Searle.

⁴⁷ Bill Gates, quoted on dust jacket of Ray Kurzweil, *How to Create a Mind: The Secret of Human Thought Revealed* (London: Duckworth Overlook, 2013).

⁴⁸ "Ray Kurzweil - Director of Engineering, Google Inc., Alphabet Inc.," 8.

⁴⁹ Ray Kurzweil, *Get Ready for Hybrid Thinking*, accessed March 24, 2019, https://www.ted.com/talks/ray_kurzweil_get_ready_for_hybrid_thinking.

The exchange began when John Searle wrote a scathing critique of Kurzweil's work in the form of a review of Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, which presents Kurzweil's predictions of technological trends and the expansion, and even replacement, of human intelligence by machine intelligence.⁵⁰

The point at issue between Kurzweil and Searle is the validity of the Turing Test, namely whether a machine designed to generate human-like responses to natural language conversations, and that does, in fact, generate responses indistinguishable from those of a human being, should be called intelligent as a human is intelligent. According to Kurzweil,

...Turing's brilliant insight in proposing the Turing Test, that convincingly answering questions in a human language really probes all of human intelligence... A computer that is capable of accomplishing this – a computer that we will run into a few decades from now – will need to be of human complexity or greater, and will indeed understand Chinese in a deep way – because otherwise it would never be convincing in its claim to understand Chinese. So just stating the computer “does not literally understand Chinese” does not make sense. It contradicts the entire premise.⁵¹

According to this passage, understanding Chinese in a deep way is the unique and necessary condition of being convincing in a claim to understand Chinese, which implies that making this claim convincingly is sufficient to show that understanding has indeed taken place.

As a result, it is also clear Kurzweil endorses the Turing Test as sufficient to “probe all of human intelligence,” which includes understanding. In making this claim, he is explicitly opposing Searle's alternative thought experiment, the Chinese Room, in which a human being, following a book of instructions, responds convincingly to questions in Chinese, even though he does not understand the language. Searle imagines looking inside the room to observe that

⁵⁰ Ray Kurzweil, *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines* (New York; London: Texere, 2001); John R. Searle, “I Married a Computer,” *The New York Review of Books*, April 8, 1999.

⁵¹ Ray Kurzweil, “Ray Kurzweil Response to Searle Review,” October 31, 2005, https://web.archive.org/web/20051031003301/http://www.kurzweiltech.com/Searle/searle_response_3.htm.

the man who is apparently carrying on a conversation in fact perceives the symbols only as meaningless squiggles. The man simply follows instructions in a book to carry out an algorithm that produces conversationally acceptable replies to the messages entering the room. Searle claims, as one can see that the man following the instructions does not understand what is being said to him or what he is saying, that this illustrates that behavior can appear intelligent without being so.

On Kurzweil's view, apparent conversational competence or, in other words, behaviour consistent with understanding, is sufficient warrant for ascribing understanding to a thing. Yet there is a further subtlety, namely that Kurzweil does not ignore the question of the means by which a cognising being, natural or artificial, reaches a state of understanding, as illustrated by his "system reply" to Searle:

...the man [in Searle's analogy] is acting only as the central-processing unit, only a small part of the system. While the man may not see it, the understanding is distributed across the entire pattern of the program itself and the billions of notes he would have to make to follow the program. Consider that I understand English, but none of my neurons do. My understanding is represented in vast patterns of neurotransmitter strengths, synaptic clefts, and interneuronal connections.⁵²

According to Kurzweil, the system as a whole can be said to understand, since it is convincing in its claim to understand Chinese. Moreover, he draws a provocative parallel with the human brain, insofar as a person may claim to understand but none of the individual neurons that fire concomitantly with this act of understanding in fact individually understand. On this account, the situation of the brain and that of the Chinese Room are not significantly different, and the failure of the man in the Chinese Room to understand cannot be used to justify the claim that the system of the man in the Room as a whole fails to understand.

⁵² Ray Kurzweil and John R. Searle, "I Married a Computer": An Exchange," *The New York Review of Books*, May 20, 1999.

In the context of the finite machines considered in the Turing Test, it is clear that Kurzweil considers that appropriate behavior is sufficient evidence of understanding, although his interest in brain operations suggest that he is not a simple behaviorist.⁵³

Another key concept in Kurzweil's treatment of issues related to understanding is what he calls *patternism*. In his 1990 work *The Age of Intelligent Machines* Kurzweil writes,

Knowledge consists of facts, and *understanding* of the relationships between facts, and their implications. One difference between humans and computers lies in the relative strengths in their respective abilities to *understand* symbolic relationships and to learn facts... Computers have been weak in their ability to understand and process information that contains abstractions and complex webs of relationships...⁵⁴

In this passage, Kurzweil draws attention to two aspects of knowledge, the first being the facts, which he describes as the basic units of information about the world gained from sense impressions, and which are combined by allowable logical (and therefore mathematically describable) transformations.⁵⁵

The second aspect of knowledge is understanding, which he claims relates to the relationships between facts and their implications. Kurzweil underlines that computers had thus

⁵³ It should be noted that there are many additional and subtle questions about the Turing Test. For example, as a thought-experiment proposed by Ned Block, one could imagine a machine with a practically infinite and simple look-up table with an answer pre-written for every conceivable question, fine-tuned to the conversation to date. Such a machine could pass the Turing Test but its lack of any engagement with the content of the questions would make Block and others reluctant to ascribe understanding. Ned Block, "Psychologism and Behaviorism," *The Philosophical Review* 90, no. 1 (1981): 5–43, <https://doi.org/10.2307/2184371>; In response to Block, Daniel Dennett has argued that this hypothetical machine, which is simple but has a practically infinite look-up table, would in fact be practically impossible anyway. Any machine that could be built would have to engage with the content of the questions and hence be immune to Block's objection. Daniel C. Dennett, "Can Machines Think?," in *Language and Meaning in Cognitive Science: Cognitive Issues and Semantic Theory*, Artificial Intelligence and Cognitive Science 4 (London: Garland Publishing, Inc., 1998), xii+300 Kurzweil follows Dennett on this point.

⁵⁴ Ray Kurzweil, *The Age of Intelligent Machines* (Cambridge, Mass; London: MIT Press, 1990), 17, emphasis added.

⁵⁵ It should be noted that Ray Kurzweil, *The Age of Intelligent Machines* (MIT Press, 1990), 34. interprets Wittgenstein's *Tractatus* as lending support to this notion of facts, "In Wittgenstein's world there are certain elementary facts, there are propositions about relations between elementary facts, and there are certain allowable transformation on such propositions that yield composite propositions. His concept of human thought is that we receive sense impressions, which comprise elementary facts. We can then transform these elementary facts and derive relationships among them according to the allowable logical processes. Any thought outside this scheme is either false or nonsensical."

far been weak in their ability to grasp the complex webs of relationships that, he claims, are constitutive of what is called understanding. As an example, while a computer performing a game of chess might, in principle, be able to play a perfect game by calculating every possible move, such a machine would require more time than has currently elapsed in the lifetime of the universe. Human beings reduce the explicit computation times for such problems by identifying strategically promising patterns on the chess board, which

...requires insight and is essentially a pattern-recognition judgment.⁵⁶

According to Kurzweil, to understand is to grasp webs of relations, and this is fundamentally a matter of pattern recognition. Furthermore, it is adroitness at pattern recognition judgements that involves insight.

Indeed, Kurzweil states explicitly that

Pattern recognition [is] ... the primary basis of human intelligence.⁵⁷

This fact, coupled with the neuroscientific evidence that the neocortex is organised into layers of cortical columns dealing with patterns of information in a hierarchical fashion,⁵⁸ is the basis of Kurzweil's aspiration to focus on the "pattern recognition theory of mind (PRTM)"⁵⁹ as a means to achieving strong AI and human cognitive enhancement.⁶⁰

⁵⁶ Ray Kurzweil, "Deep Fritz Draws: Are Humans Getting Smarter, or Are Computers Getting Stupider?" «Kurzweil», Kurzweilai, accessed November 16, 2022, <https://www.kurzweilai.net/deep-fritz-draws-are-humans-getting-smarter-or-are-computers-getting-stupider-2>.

⁵⁷ Kurzweil, "Deep Fritz Draws."

⁵⁸ According to Kurzweil, *How to Create a Mind*, 35., "...the neocortex is responsible for our ability to deal with patterns of information and to do so in a hierarchical fashion. Animals without a neocortex (basically nonmammals) are largely incapable of understanding hierarchies. ... The neocortex is responsible for sensory perception, recognition of everything from visual objects to abstract concepts, controlling movement, reasoning from spatial orientation to rational thought, and language—basically what we regard as "thinking." ... Elsewhere, Kurzweil, *The Age of Intelligent Machines*, 41., states "The pattern recognition theory of mind ... is based on the recognition of patterns by pattern recognition modules in the neocortex. These patterns (and the modules) are organized in hierarchies." On these accounts, it is clear that much of what Kurzweil calls "thinking" involves structuring the perceived world into hierarchies.

⁵⁹ Kurzweil, *How to Create a Mind*, 34.

⁶⁰ As one consequence, Kurzweil anticipates that massive parallel processing will outstrip the computation achievements of Turing or Von Neumann machines for many of the pattern-recognition problems he is interested in addressing.

Kurzweil offers an example of how the mind makes sense of the world hierarchically as follows. The letter “A” contains two equal diagonal line segments intersecting at one point and forming an acute angle and a horizontal line segment with its endpoints each on one of the diagonal’s midpoints. One’s perceptual system provides the image of these three-line segments to the pattern-recognising network. Separate pattern recognisers fire for each of these inputs, and higher-level pattern recognisers takes these outputs as their inputs and give an output corresponding to “A”. If “A” is part of a word, then other pattern recognisers specialising in identifying whole words of English will also fire in response to the outputs from the recognisers for “A” and the other letters that are present. When reading the word “FACE” for instance, pattern recognisers fire which receive input from both the recognisers for the letters composing the word and the recognisers for the perceptual features of faces.⁶¹ Kurzweil’s aspiration is that these operations of recognising hierarchically-order patterns could be implemented artificially.⁶²

Kurzweil further claims that pattern processing can also explain the origin of scientific insights such as those gained by Darwin or Einstein. Both, he claims, were adept at discovering *metaphors*,⁶³ which he describes in his schema as follows:

⁶¹ An individual is recognised by hosts of pattern processors that give a positive output in response to any one of thousands of traits that may be observed in an individual (Kurzweil gives the example of the multiplicity of expressions a face may exhibit and the variety of angles from which it may be seen). Classes are represented by higher pattern processors that respond to common traits in individuals, which serve as abstractions in Kurzweil’s view, and these higher networks are nested to create classes of classes, and so on. Kurzweil, *How to Create a Mind*, 39–41.

⁶² Kurzweil, *How to Create a Mind*, 90–91, 113., elaborates the architecture of this process as follows. Sense impressions constitute the elements, the inputs, to the hierarchy of pattern recognisers in the brain, which results in the firing of low-level pattern recognising modules. The pattern of their firings represents these lowest level elements of the world to higher level recognisers, which in turn fire, signaling higher level modules above them. The signaling is two-way, insofar as the higher pattern recognisers receive predictive information from those above them, lowering their thresholds and those of their subordinate pattern processors. Eventually, signals reach high level pattern recognisers which are responsible for abstractions, concepts and even metaphor.

⁶³ “Darwin perceived that Charles Lyell’s insight on how very gradual changes from a trickle of water could carve out great canyons was a powerful metaphor for how a trickle of small evolutionary changes over thousands of generations could carve out great changes in the differentiation of species. Thought experiments, such as the one that Einstein used to illuminate the true meaning of the Michelson-Morely experiments, are all metaphors, in the

Finding a metaphor is the process of recognizing a pattern despite differences in detail and context—an activity we undertake trivially every moment of our lives. The metaphorical leaps that we consider of significance tend to take place in the interstices of different disciplines.⁶⁴

On this account, finding a metaphor is a process by which one recognises a pattern present even when aspects of the input, such as context and detail, are changed. Kurzweil states that this process is continual and constant, as it produces both trivial instances of daily recognition and great scientific realisations.

In summary, Kurzweil credits the distinctive intellectual powers of human beings, in contrast to those of computers, to a comparative superiority in hierarchically ordered pattern recognition. On this account, one's sense impressions constitute the input to a hierarchy of pattern processors, undergoing transformations layer-by-layer, yielding a set of high-level abstract patterns, including their classes, relations, and attributes. These outputs constitute much of what Kurzweil considers is meant by the term "understanding".

Kurzweil's confidence about the ability to describe understanding in these terms, and implement the steps involved on an artificial substrate, links closely with his "techno-optimism" about sharing the process of understanding with machines and thereby also enhancing individual and collective human understanding. In his *How to Create a Mind*, he states:

Our own creativity is orders of magnitude greater than that of the mouse—and involves far more levels of abstraction—because we have a much larger neocortex, which is capable of greater levels of hierarchy. So one way to achieve greater creativity is by effectively assembling more neocortex. One approach to expand the available neocortex is through the collaboration of multiple humans. This is accomplished routinely via the communication between people gathered in a problem-solving community. Recently there have been efforts to use online collaboration tools to harness the power of real-time collaboration, which have shown success in mathematics and other fields. The next step, of course, will be to expand the neocortex itself with its nonbiological equivalent. This will be our

sense of being a 'thing regarded as representative or symbolic of something else...'” Kurzweil, *How to Create a Mind*, 114.

⁶⁴ Kurzweil, *How to Create a Mind*, 81.

ultimate act of creativity: to create the capability of being creative. A nonbiological neocortex will ultimately be faster and could rapidly search for the kinds of metaphors that inspired Darwin and Einstein. It could systematically explore all of the overlapping boundaries between our exponentially expanding frontiers of knowledge.⁶⁵

According to Kurzweil in this passage, the human is far more intellectually creative, and hence capable of insights, than a mouse by virtue of having more neocortex, which, according to PRTM, implies being able to recognise more patterns in information and represent such patterns in hierarchies. Kurzweil's language implies not only that the neocortex facilitates abstraction, but that an expanded quantity of neocortex, by increasing the availability of hierarchy, also has the potential to expand the possible degree of abstraction and the potential for insightfulness.

What is at least clear from this text and others is that Kurzweil regards the expansion of the quantity of neocortex to be a priority for enhancing insight. This expansion has happened by natural means over evolutionary timescales but may be enhanced either by the collaboration of multiple human persons, or by the effective expansion of the neocortices of individual persons, or both. Large-scale collaboration is, of course, already one of the hallmarks of what it means to be human, and Kurzweil does not have much to add on this point beyond a few remarks about comparative advantage in speed of machine-machine communication relative to human-human communication, and the desirability of developing tools to improve the latter.⁶⁶

By contrast, Kurzweil has much to say about the direct expansion of the neocortex itself, and he clearly regards this expansion as central to the next logical step for enhancing creativity. By "expansion" here he means principally the development of computational tools that emulate the human neocortex. By means of faster processors, or superior pattern-recognition algorithms,

⁶⁵ Kurzweil, *How to Create a Mind*, 116.

⁶⁶ Kurzweil predicts that humans will eventually transfer their minds onto nonbiological substrata and then likewise be capable of sharing thoughts directly. Ray Kurzweil, *The Singularity Is Near* (London: Gerald Duckworth & Co, 2010), 58, 66–68, <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=2029217>.

Kurzweil anticipates that such tools, which may properly be called AI, will be capable of searching for solutions to problems faster than is possible for humans, as well as making the kinds of links he refers to as metaphors, and manifesting enhanced creativity and insights.

Such tools may be employed in a standalone capacity, treated as a black box with external interfaces.⁶⁷ Kurzweil's favoured and more ambitious aspiration, however, is to create a direct, internal interface between the brain and the AI that extends the mind's own neocortex beyond the body and also, in effect, vastly accelerates the evolutionary trend of a growing neocortex.

2.2.3 Concluding Remarks on Kurzweil

In conclusion, Kurzweil holds a view of understanding that prioritises outputs or results, hence favouring the Turing Test over Searle's Chinese Room. This perspective also frees him to speculate about future machines capable of emulating the "pattern recognition theory of mind (PRTM)" that he ascribes to the human neocortex. According to this picture, understanding is just a matter of pattern recognition, and the term insight just indicates that the algorithm is effective at classifications. If an algorithm can perform sufficiently well on pattern recognition tasks, then their output justifies the claim that they are intelligent and possess understanding. Kurzweil describes those acquisitions of understanding typically called insights as being generated through finding metaphors, which is described as pattern recognition across differing contexts.

⁶⁷ "Our expanded neocortex will be housed [in the Cloud] too. Whether we access such expanded intelligence through direct neural connection or the way we do now—by interacting with it via our devices—is an arbitrary distinction. In my view we will all become more creative through this pervasive enhancement, whether we choose to opt in or out of direct connection to humanity's expanded intelligence. We have already outsourced much of our personal, social, historical, and cultural memory to the cloud, and we will ultimately do the same thing with our hierarchical thinking." Kurzweil, *How to Create a Mind*, 81.

Understanding is apparently essential for human cognition according to Kurzweil, so creating intelligent computers will require creating systems whose outputs justify the claim that they understand. Once such algorithms have been developed, it will be supposedly possible to arbitrarily expand the size of the algorithms, giving humans access to a kind of neocortical prosthetic. By such means, he is a techno-optimist about the prospects of partially outsourcing and hence enhancing human understanding in future.

2.3 Nick Bostrom: Insight as a Promissory Black Box

2.3.1 Introducing Nick Bostrom

Nick Bostrom (1973 –), founding director of both the Future of Humanity Institute at Oxford University and the Oxford Martin Programme of Future Technology, is arguably the most well-known transhumanist thinker. He has perhaps most rigorously expressed the aspiration to greater intellectual powers and delineated the scope for potential enhancements.⁶⁸

As cited above, Bostrom writes that insight will be needed to successfully achieve the project of posthuman existence. Bostrom indicates the importance of insight to the project of human enhancement, so what lessons may be learned about his sense of what insight is and how it relates to cognitive enhancement?

Bostrom has written a vast amount on enhanced intelligence, the key strands of which he drew together in his monograph, *Superintelligence*, published in 2014.⁶⁹ Bostrom characterises three main types of superintelligence: speed superintelligence, collective

⁶⁸ Nick Bostrom, “Transhumanist Values,” *Journal of Philosophical Research* 30 (2005): 3–14, https://doi.org/10.5840/jpr_2005_26; Nick Bostrom, “Why I Want to Be Posthuman When I Grow Up,” in *The Transhumanist Reader*, ed. Max More and Natasha Vita-More (Chichester, West Sussex, UK: Wiley-Blackwell, 2013), 28–53.

⁶⁹ Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, First edition, Ebook Central (Oxford: Oxford University Press, 2014), <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=1743975>.

superintelligence, and quality superintelligence. I shall consider each of these in turn before examining their relevance to the question of insight.

2.3.2 Bostrom on Understanding and Superintelligence

The first of Bostrom's speculations is of a "weak" form of superintelligence that consists of taking a mind functionally equivalent to that of a human mind and simply running it on faster "hardware" (to use the computer metaphor that permeates his writing on the human brain).

By running the brain on faster hardware, Bostrom does not mean using drugs or other means to stimulate an actual human brain. Rather, he suggests that a human brain might be modelled on a computer, specifically a computer platform running a hypothetical piece of software called a brain emulator, which

...models the states and functional dynamics of a brain at a relatively fine-grained level of detail.⁷⁰

In this way, he hopes it will be possible for the emulator to capture sufficient detail to allow the objective neural concomitants of subjective cognition to be emulated, producing

...the phenomenological effects of a mind.⁷¹

This process would, he claims, enable what he calls uploading, which he describes as

a hypothetical future technology that would enable a human or other animal intellect to be transferred from its original implementation in an organic brain onto a digital computer.⁷²

⁷⁰ Nick Bostrom and Anders Sandberg, *Whole Brain Emulation: A Roadmap (Technical Report 2008-3)* (The Future of Humanity Institute, 2008), www.fhi.ox.ac.uk/reports/2008-3.pdf.

⁷¹ Bostrom and Sandberg, *Whole Brain Emulation*.

⁷² Nick Bostrom and Eliezer Yudkowsky, "The Ethics of Artificial Intelligence," in *The Cambridge Handbook of Artificial Intelligence*, ed. Keith Frankish and William M. Ramsey (Cambridge: Cambridge University Press, 2014), 325.

Once the “mental software” is running on a different, more powerful platform, he speculates that it will then be possible to accelerate the objective neural concomitants of subjective thought, thereby speeding up subjective thought in objective time.

Whatever else may be said about Bostrom’s proposal or prediction, it is worth flagging up some of its implicit philosophical premises. First, this approach assumes that everything that goes on in the mind can be reduced to physical processes in the brain and that these processes can be observed and described in sufficient detail. How one would record the structure of the brain, the most complex known object in the universe, remains unclear, but there is ample speculation in the transhumanist literature. One popular suggestion, which Bostrom mentions, is that the brain might be carefully frozen, sliced into thin cross-sections, and scanned with a microscope, although of course this proposal assumes the loss of dynamism in a static brain can be tolerated.⁷³

Second, Bostrom’s approach assumes that everything in the brain can be described mathematically, in other words that material processes can be reduced to formal descriptions. On this functionalist account, thought may be realised in any material provided the right sort of pattern processing occurs, be it in flesh, silicon, or even paper.⁷⁴ Of course, there is also a hidden, subsidiary assumption, namely that these mathematical descriptions can also be encoded on the finite bit-strings that are ultimately the objects manipulated in computers.

Third, it assumes that mental processes in the mind, isolated from the body with its nerves, senses, fluids, and physical environment, will remain meaningful. Assuming that the brain may be easily separated from the body ignores the complex chemical interactions,

⁷³ Bostrom and Sandberg, *Whole Brain Emulation*.

⁷⁴For a fuller exposition of functionalism, see Daniel C. Dennett, *Kinds of Minds: Towards an Understanding of Consciousness*, Science Masters Series (Phoenix, 1997).

hormonal and otherwise, that link the brain with other organ and tissue systems.⁷⁵ The processes of the brain are sufficiently integrated with other physiological processes that to isolate the brain from them would significantly alter its processes, if they could even continue at all.⁷⁶ Bostrom is aware of these challenges and suggests that the body might be simulated along with the environment. He assumes, however, that at some level of functional description, many lower causal levels of these systems may be abstracted away. That is simply to apply to the body and the world the previous assumption that everything can be mathematically described and represented in finite bit strings.

In short, Bostrom assumes that embodiment, both in the sense of being an animal in the world and in the narrower sense of the processes of the brain, are just instantiations of more abstract functions or patterns that constitute a particular mind.

Granted all these challenges can be met, one is, of course, free to speculate about what might be achievable, and Bostrom certainly conjures up some fantastic scenarios. Imagine, for example, a superintelligence that is speeded-up by a factor of one thousand might “be able to

⁷⁵ A related view is expressed in Daniel Dennett’s summary of his involvement with the Cog research project. It was an attempt to create an algorithm that would learn to control and use input from a versatile robot to display intelligent behaviour in the real world. Dennett thought that the confidence of AI researchers who thought that a truly intelligent system could be created to interact with highly limited simulated worlds was misplaced, and that a system capable of functioning in the complex and non-specified conditions of the real world likely have sufficient power to be judged intelligent. Dennett does not fundamentally disagree, however, the notion that the function of the brain can be abstracted from the body. This experiment was simply meant to discover what a minimally complex abstraction of cognition would be required for intelligent function. “So it is something of a surprise to find this AI group conceding, in effect, that there is indeed something to the skeptics’ claim (see e.g., Dreyfus and Dreyfus, 1986) that genuine embodiment in a real world is crucial to consciousness. Not, I hasten to add, because genuine embodiment provides some special vital juice that mere virtual-world simulations cannot secrete, but for the more practical reason—or hunch—that unless you saddle yourself with all the problems of making a concrete agent take care of itself in the real world, you will tend to overlook, underestimate, or misconstrue the deepest problems of design.” An embodied agent, here defined as a robot with sensors and manipulators, might be a useful research tool for reverse engineering intelligence according to this perspective, but this is only so that the problems of design may be made explicit. Once “the deepest problems of design” have been solved, embodiment is something of an accessory to the intelligent algorithm. Daniel Clement Dennett, *Brainchildren: Essays on Designing Minds* (London: Penguin, 1998), 166, <http://archive.org/details/brainchildreiness0000denn>.

⁷⁶ Bostrom and Sandberg commit what Peter Hacker and M. R. Bennett would consider a mereological fallacy, attributing the power of thinking not to the whole organism but only to a part of it. M. R. Bennett and P. M. S. Hacker, *Philosophical Foundations of Neuroscience* (Oxford: Blackwell Publishing Ltd, 2003), 29.

read a book in a few seconds and write a PhD thesis in an afternoon”.⁷⁷ Nevertheless, the fact remains that the challenges listed above are massive and that we cannot be sure whether solving them is even possible with any conceivable degree of technological advancement.⁷⁸

Moreover, there are some subtler problems. Even if a super-fast assimilation of something very complex, such as a book by Hegel, is possible for the emulation, what is the significance of this achievement? One possibility is that the emulation carries the personal identity of the original unenhanced human being, but that claim is a bold one. In a separate but related case, Bostrom has argued that two qualitatively identical states realised in different material structures should be considered at least numerically distinct.⁷⁹ And if there is a distinction, even a numerical distinction, there cannot be an unqualified identity in the manner of the Identity of Indiscernibles. To the extent that there exist grounds for doubting that the emulation is identical with the original, then it is questionable whether the transhumanist goal of enhancing a particular individual's intellect has actually been achieved.

If, on the other hand, there is not an identity between the original subject and the emulation, then the emulation is a new cognitive being, albeit one copied in every detail at the beginning of its existence. Whatever benefits the emulation might receive from its digital existence, they would not accrue to the original subject. Might the operator of the emulation nonetheless gain some indirect cognitive benefit? In this case, however, the situation with the

⁷⁷ Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, First edition, Ebook Central (Oxford: Oxford University Press, 2014), 53, <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=1743975>.

⁷⁸ It may be that symbols can only be meaningful to a human being in the world, and that the symbols in a given computational system only have meaning relative to the world in the mind of the human subject. Dreyfus, Hubert L. *What Computers Still Can't Do: A Critique of Artificial Reason*. Rev Ed edition. Cambridge, Mass: MIT Press, 1999. Dreyfus has also questioned whether or not attempts to create so-called Heideggerian AI could ever sufficiently model the world for an AI agent to be inducted into the human world of skills, purposes, practices, and objects necessary to act intelligently by human lights. “Why Heideggerian AI Failed” in *Skillful Coping*, 2016, 272-273.

⁷⁹ Nick Bostrom, “Quantity of Experience: Brain-Duplication and Degrees of Consciousness,” *Minds and Machines* 16, no. 2 (2006): 16, <https://doi.org/10.1007/s11023-006-9036-0>.

emulator has a certain parallel with the situation with any other thinking personal subject. A second person, for example my supervisor, may well have a wonderful understanding of Hegel, but that is not my understanding of Hegel, and the second-person's understanding does not benefit me directly unless that person can teach me. Consequently, the emulation may, in fact, achieve no benefit to the subject but merely displace the intellectual challenge into a form where it is resolved but incommunicable.

In the latter case, the claims of the transhumanists may be suffering from an implicit category error. A machine can compute far faster and more accurately than I can and may generate some answer to a well-defined mathematical problem far more quickly to my benefit. But for many cognitive tasks, like assimilating Hegel, there is no well-defined bottom-line answer that may be shared easily, like, for example, the number "42".⁸⁰ Whatever cognitive achievements are not in the form of simple, bottom-line answers remain to be investigated (see chapter three). But until such achievements are investigated, mere acceleration of the speed of thought in the manner outlined above faces philosophical as well as technical challenges.

Bostrom's second speculation, collective superintelligence, is an umbrella for a variety of scenarios which divide into two main groups: the first is a set of distinct intellects that co-operate; the second is a set of intellects that make up a larger, unified intellect. Such intellects may comprise both human and machine agents. Bostrom explores potential collective enhancements in terms of increasing the total number of agents, their integration, and their

⁸⁰ Indeed, in their discussion of "oracle AI's" (machines designed to answer any question put to them and assumed to have reasoning power far beyond that of a human) Bostrom, Sandberg and Stuart assert, "There are two broad types of question one can ask an AI: predictions and problems. Generally, predictions are questions of the 'What would happen if...' type, while problems are of the 'how can we achieve this...' type..." These are cognitive tasks in which only a workable output is ultimately necessary. Similarly, ethical questions are reduced to matters in which a bottom-line answer on how one ought to act is sufficient. Stuart Armstrong, Anders Sandberg, and Nick Bostrom, "Thinking Inside the Box: Controlling and Using an Oracle AI," *Minds and Machines* 22, no. 4 (2012): 307, <https://doi.org/10.1007/s11023-012-9282-2>.

quality. The possibility of enhancing the quality of intellects is essentially the proposal to enhance many individual intellects, which is discussed under the next heading. The integration scenario is further capable of being subdivided into weak integration by virtue of improved techniques of organisation and communication, and strong integration by forging an aggregate of minds into a unitary intellect.

An implicit assumption, in the case of co-operating intelligences, is that the problem, whatever it is, can be broken into distinct parts and subcontracted to individual intelligences. Examples are well-known already in complex calculations that can be parallelised, e.g., abstract cells of models of atmospheric circulation being allocated to distinct computer processors. The difficulty, of course, and one that is recognised by Bostrom, is that not all problems can be parallelised in this way, for example calculating the sequence of digits of pi.⁸¹ And for those problems that cannot be subdivided into discrete and subcontracted parts, it is challenging to see how the collective superintelligence is going to help.

There is another issue, namely that it is unclear how a solution to a divided problem is ever understood by any one of the individual contributors. As noted above, many cognitive tasks, such as understanding Hegel, do not reduce to a simple bottom-line answer. In the case of a distributed task across a network, who or what can be said to understand the answer, if there is one? Consider, for example, the proverbial blind men describing an elephant by parts: one says it is like wall; another says it is like a sail; and another says it is like a snake: Can any be said to grasp the whole? Such questions raise other issues, such as the need to give an account of the nature of an intellectual community and the relations of the members within it. In

⁸¹ This is true not only of mathematical problems but also of artistic tasks. Bostrom illustrates the inability to substitute quantity of workers for the quality by asking rhetorically whether an office of literary hacks could produce a work by Proust, but we might equally ask whether a hundred literary luminaries could individually write as many chapters of a great novel. It is likely the characters, style, and plot would be uneven. Bostrom, *Superintelligence*, 58.

particular, an account must be given of what is meant by the phrase “we understand.” Does this locution assert that there is a mere group of entities all of whom can independently say, “I understand”? But what of the cases where a group is able to answer a question that none independently know how to solve, which are predominately the cases to which Bostrom refers?

As a more exotic variant of collective superintelligence, Bostrom entertains the possibility is that a sufficiently integrated collective may itself become a unified, intelligent agent. He asks,

Could the Internet become something more than just the backbone of a loosely integrated collective superintelligence—something more like a virtual skull housing an emerging unified superintellect?... [It is plausible] that myriad incremental improvements eventually create the basis for some more unified form of web intelligence. It seems at least conceivable that such a web-based cognitive system, supersaturated with computer power and all other resources for explosive growth save one crucial ingredient, could, when the final missing constituent is dropped into the cauldron blaze up with superintelligence.⁸²

In other words, Bostrom speculates that some future form of the current Internet may itself be said to be “superintelligent,” by implication not just a cooperation of individual human minds, but something that can be regarded as a singular intelligent agent.

Bostrom’s use of highly evocative and poetic language, such as a “blazing cauldron” risks obscuring some awkward philosophical questions. First, Bostrom seems to assume the metaphysical possibility that distinct and separate minds may become one, seemingly without destruction of the separate minds. But what is this singular agent, and can this agent say, “I understand!”? And does the understanding of this super-agent mean that any of the constituent parts understand? Is it plausible to claim that, were such a collective intelligence possible, the understanding of any one person would be directly accessible to another in such a way that one

⁸² Bostrom, *Superintelligence*, 49.

would not have to change oneself by the effortful task of acquiring insight possessed by another or by the whole. Alternatively, if the individual parts, including all the human parts, are absorbed into the collective superintelligence and, in effect, annihilated (at least temporarily), surely then there is less understanding overall than before, at least by actual human beings.

This collective, or rather emergent intelligence, whether it exists apart from humans or absorbs them, does not seem to be a useful guide to how shared understanding exists among humans.

Bostrom's third speculation is *quality superintelligence*, a somewhat vague term established in tacit recognition of the fact that speed and collective superintelligence have conceptual limitations. To justify this assessment, he cites approvingly Vernor Vinge's rhetorical question, "Imagine running a dog mind at very high speed. Would a thousand years of doggy living add up to any one human insight?"⁸³ We might also extend this question. For example, "Do the cumulative experiences of a thousand networked dogs add up to any one human insight?" In other words, quantitative boosts do not necessarily give qualitative boosts, allowing for totally new modalities of thought. Hence the need for a third category: quality superintelligence.

But are such qualitative boosts possible, and what would they be like? Bostrom tends to argue from analogy.⁸⁴ Superintelligence of this kind, he claims, would be "of quality at least as superior to that of human intelligence as the quality of human intelligence is superior to that of

⁸³ Vinge, 1993. Quoted in Bostrom and Yudkowski, "The Ethics of Artificial intelligence". 14.

⁸⁴ In a similar vein, Bostrom observes that the human appreciation for music and the resources spent in pursuit of it would be unintelligible to a species without the faculty for recognising music. He then suggests that we may be like this tune-deaf species, lacking mental powers that would open to us untold vistas of value. Bostrom's argument depends entirely on the possibility that we may stand in relation to some being with undefined mental powers as some being with only a subset of our mental powers stands in relation to us. Bostrom, 2005, "A Philosophical Quest for Our Biggest Problems", https://www.ted.com/talks/nick_bostrom_on_our_biggest_problems

elephants, dolphins, or chimpanzees.”⁸⁵ Bostrom attributes the qualitative superiority of human intelligence to the use of linguistic representation and communication, which is premised, he claims, on specialised modules or neural circuitry.⁸⁶ He makes the following imaginative leap:

... were we to *gain* some new set of modules giving an advantage comparable to that of being able to form complex linguistic representations, we would become superintelligent.⁸⁷

In other words, Bostrom infers that it may be possible for there to exist some cognitive capacities that would enable cognition qualitatively superior to that of humans analogous to the capacity for language enabling human cognition to be qualitatively superior to animal cognition.

Bostrom’s perspective on enhancing understanding may be illustrated by the kinds of problems for which he thinks cognitive enhancement might be beneficial.

There might thus be some problems that are solvable by a quality superintelligence, and perhaps by a speed superintelligence, yet which a loosely integrated collective superintelligence cannot solve (other than by first amplifying its own intelligence). We cannot clearly see what all these problems are, but we can characterize them in general terms. They would tend to be problems involving multiple complex interdependencies that do not permit of independently verifiable solution steps: problems that therefore cannot be solved in a piecemeal fashion, and that might require qualitatively new kinds of understanding or new representational frameworks that are too deep or too complicated for the current edition of mortals to discover or use effectively. Some types of artistic creation and strategic cognition might fall into this category. Some types of scientific breakthrough, perhaps, likewise. And one can speculate that the tardiness and wobbliness of humanity’s progress on many of the “eternal problems” of philosophy are due to the unsuitability of the human cortex for philosophical work. On this view, our most celebrated philosophers are

⁸⁵ Bostrom, *Superintelligence*, 57. First, the qualitative superiority of human intelligence for solving problems in contexts relevant to us do not appear based on general intellectual ability alone or on collective intelligence. Human brains are smaller, and therefore likely less computationally powerful, than many cetacean and elephant brains. Also, socially developed mammals (again elephants and cetaceans) lack our qualitative intelligence. From this, Bostrom concludes that mere computational power and connectedness of individuals does not underlie human powers of reason.

⁸⁶ Bostrom, *Superintelligence*, 57. The basic premise of the argument is not without some merit: psychological literature is replete with examples of humans with domain-specific cognitive deficits traceable to specific brain abnormalities.

⁸⁷ Bostrom, *Superintelligence*, 57.

like dogs walking on their hind legs— just barely attaining the threshold level of performance required for engaging in the activity *at all*.⁸⁸

Here Bostrom clarifies that different kinds of superintelligence might have differing degrees of relevance to the challenge of enhancing understanding. A loose kind of collective intelligence that simply increases the total number of thinkers or allows more communication would not, on his view, tend to accelerate understanding. A speed superintelligence, however, might allow one mind to grasp more quickly the whole of the complexities of a problem that cannot be subdivided. Finally, a quality superintelligence might be capable of understanding or representing problems entirely beyond a human mind. Crucially, this passage also describes problems requiring understanding as difficult to clearly see but generally describable as having complex interdependencies and not permitting solutions steps that are independently verifiable, which is to say solvable in a piecemeal fashion. Moreover, Bostrom identifies these as the kinds of problems that pertain especially to creativity, scientific breakthrough, and deep questions of philosophy. In other words, Bostrom implies that the ability to understand is to some extent distinct from procedural reasoning and that accelerating or improving the generation of new understanding would be highly valuable.

But what would the subjective experience of such qualitative enhancements be like? Bostrom is vague on this point and again has to rely on illustrative stories and arguments from analogy. In one of his earlier works, “Transhumanist Values”, Bostrom observes that,

... our own cognitive makeup may foreclose whole strata of understanding and mental activity. ...The impossibility I am referring to is more like the impossibility for us current humans to visualize a 200-dimensional hypersphere...⁸⁹

⁸⁸ Bostrom, *Superintelligence*, 58–59, emphasis original.

⁸⁹ Bostrom, “Transhumanist Values,” 6. Nick Bostrom, “Transhumanist Values,” *Journal of Philosophical Research* 30 (2005): 6, https://doi.org/10.5840/jpr_2005_26.

Bostrom is not referring here to an inability to describe mathematically a 200-dimensional hypersphere, but of the inability to visualise, to grasp such an exotic object all at once.⁹⁰ As a parallel, an inhabitant of a two-dimensional flatland might be able to describe a 3D object, such as a sphere, but not “see” it or grasp it imaginatively as a whole. In terms like “visualise,” incidentally, we may get some sense of what Bostrom means by understanding.

Bostrom also explains qualitative superintelligence by illustration and analogy.

Human working memory is able to hold no more than some four or five chunks of information at any given time. While it would be misleading to compare the size of the human memory directly with the amount of RAM in a digital computer, it is clear that the hardware advantage of digital intelligences will make it possible for them to have larger working memories. This might enable such minds to intuitively grasp complex relationships that humans can only fumblingly handle via plodding calculation.⁹¹

The general idea is that more objects might be held in attention at once, and that this would be a precondition for acts of understanding that rendered intelligible a larger set of phenomena than a human being could make sense of.

As before, there are numerous concealed assumptions in these speculations. For example, he tacitly assumes that the ability to grasp something intuitively is related to working or short-term memory, presumably analogous to the way in which computers have fast-access RAM, with slow-access, large-volume storage delegated to drives. But something that is understood as a whole is often associated with a sublimation and reduction of detail, as many individual facts are related to a unitary whole.⁹² So, for example, human beings don’t need to

⁹⁰ It should be noted in passing that the assessment may be unduly pessimistic. It is not clear that the human ability to imagine abstract objects obliquely cannot enable inquiry that eventually renders any given mathematical object not only describable but also capable of an intuitive grasp.

⁹¹ Bostrom, *Superintelligence*, 60.

⁹² By way of illustration, autistic children can suffer from an inability to cease attending to detail and instead attend to wholes. For instance, they may be overwhelmed by encountering a huge number of persons in an urban environment and be unable to view the sea of humanity before them en masse, as a crowd, and so attend to other things. In severe cases they can, quite literally, miss the forest for the trees. Even in mature, high functioning

remember every stripe of a tiger's body; it is enough to recognise a fuzzily stripy (and dangerous) animal. Being able to recall more detail may well benefit one's ability to reason discursively, by reducing one's need to transcribe steps in a lengthy argument, say, but this process may proceed in a rote fashion, without intuition. Even by this vague analogy, it is unclear that some form of enhanced short-term memory would enhance understanding in itself.⁹³

And of course, many other seemingly extremely difficult problems remain. For example, in this instance, as previously, there are questions about the unity of consciousness, what is meant by "intuitive grasping" (as Bostrom uses the phrase), and how Bostrom claims that human intelligence is superior to the higher non-human mammals. On the latter point, the question is not whether human intelligence is superior, as such, but how Bostrom understands this superiority in such a way as to furnish an effective analogy for the way in which a quality superintelligence would conceivably surpass our own. Hence, although Bostrom plausibly highlights the potential significance of quality enhancement, distinct from speed and collective enhancements, the meaning of this third category of speculative superintelligence is never really made clear.

2.3.3 Concluding Remarks on Bostrom

In conclusion, prospects for speed superintelligence and collective superintelligence seem weak in terms of their potential to benefit understanding, and quality superintelligence effectively acknowledges that the problem of understanding is not merely a matter of processing

patients, a propensity may exist for discussions of detail without any conclusions being drawn or arguments being made.

⁹³ As further corroboration, it should also be noted that some non-human animals have better short-term memory than human beings (e.g., chimpanzees recalling sequences of numbers) without this fact, in itself, boosting understanding to the human level, let alone beyond it.

speed or the number and configuration of processors (of some kind). But what understanding is, precisely, remains elusive. When outlining visions of enhancing quality, Bostrom resorts to imaginative stories, often literally asking readers to imagine the capacity to imagine realities that cannot be grasped all-at-once in ordinary life, such as, for example, the 200-dimensional hypersphere. But imagining capacities to imagine cannot realistically be deemed great progress in itself towards such goals. Hence, such visions remain something like a promissory black box, except that not only do we have little idea of what would need to go into the box to provide these new powers, but we are still hazy about what, precisely, these powers would do or be like to experience.

At any rate, one point is clear from this overview of representative transhumanist thinkers, namely that their understanding of understanding is vague. If, therefore, we are to make any realistic evaluation of transhumanist prospects for accelerating insight, it is important to go back to basics: to look in more detail at what philosophers, scientists, and theologians have said about understanding, drawing upon past wisdom and reinterpreting it in contemporary forms as needed. This is the subject of the next chapter.

2.4 Summing Up: Transhumanists on Understanding

This chapter highlights the ways in which representative transhumanists have given some thought to the question of understanding, albeit to different degrees and not always using this term. They share the view that the cognitive phenomenon known as understanding is cognitively valuable. Further, they agree that insight, gaining new understanding, is at least superficially distinct from conscious stepwise reasoning.

They also hold, however, the view that whatever is meant by understanding is capable in principle of being performed by some kind of machine-based process. Minsky describes

understanding as a complex of algorithms that instantiate frames. Kurzweil claims understanding is the operation of the pattern recognition engine of the neocortex, a kind of pattern recognition that he often associates with metaphor. Bostrom does not commit to a particular account of understanding but is hopeful that it will prove to be some system amenable to emulation. The greatest aspirations for accelerating understanding are tied to the hope that the faculty of understanding can be realised in machines, so humans may at least have access to superintelligent tools and preferably transfer their minds to superior hardware or cybernetically augment themselves to gain superintelligence directly.

A concomitant attitude that they share is an approach to understanding that measures success in terms of correct results, especially in the context of problem-solving scenarios. In some texts, there are references to aspects of cognition that do not fit easily into this paradigm, such as imagination, inarticulate knowledge or the apparent “knowing” associated with meditative experiences.

Relatedly, to the extent the transhumanists do show an interest in the operation of human intelligence, as opposed to its outputs, they conceive of it along the lines of some sort of abstract function. With respect to insight, specifically, the transhumanists think that it is just a term used to describe those architectures we do not yet know, though they are happy to speculate that these unseen systems might consist of frames or of pattern recognition machines. It is unclear, however, to what extent such an approach can describe not only the implicit aspects of human insight but the role of embodiment as well.

What at least emerges from these initial investigations is that these three important transhumanist thinkers acknowledge, to a greater or lesser extent, that understanding is an important aspect of human cognition.

Their aspirations to enhance human cognition will, therefore, need to take account of understanding and seek to enhance such understanding. My critique of these aspirations will, therefore, next need to evaluate these accounts of understanding and prospects for its enhancement. As a first step, I turn to some especially influential accounts of understanding from the broader philosophical tradition. I will proceed to discuss literature on understanding from psychology, narrative, and metaphor. This will allow the deficiencies of the perspective on insight in the transhumanist anthropology to be brought out in chapter four. Nonetheless, as chapter four shows, their anthropology is inadequate in that they treat all thought as a form of *data processing*, they detach human intelligence from the existence of human beings as *animals*, and they tend to consider human beings as data-exchanging beings so that *explicit* communication almost exhausts human shared existence.

CHAPTER 3 Insights into Insight and Second-Person Relatedness

3.1 Introduction: the personal dimension of understanding

It should be acknowledged that the transhumanists, philosophers, and theologian discussed in this thesis operate in different metaphysical frameworks. Minsky, Kurzweil, and Bostrom are physicalists. Lonergan, Polanyi, and Aquinas are used here to support an Aristotelian-Thomistic perspective. Nonetheless, exploring the treatment of insight and understanding within the different frameworks will make it possible to test which of the frameworks is better able to describe understanding and those changes in understanding called insight. This evaluation will take place in chapter four.

This chapter will examine the phenomena termed understanding and insight from multiple perspectives.

The first section will seek a clearer meaning of what is meant by understanding and insight both as used in common speech and in the more technical sense studied in psychology. With this initial sense of the terms clarified, the exploration will turn to two philosophers who have thoroughly examined the qualities of acts of understanding and their relation to the subjects who understand. Understanding will be examined as something that exists in some subject, which is to say as something that is possessed by a first-person perspective. As understanding is always from some perspective, the question will be asked whether the embodiment of a particular person is relevant to understanding as well as to more abstract forms of cognition. This exploration constitutes the first section.

The second section expands the study of understanding to describe the dynamic of two or more persons engaged in sharing understanding. The first-person aspects of understanding

in an I-it relation are also present when engaged in shared attention to which is added an I-thou relation, so this section fruitfully builds on the first.

The third section turns to explore understanding from a theological perspective, which will draw on the first-person and second-person aspects of understanding to explore the relation of the divine to human understanding.

An important distinction between the philosophical and theological perspectives should be here observed. A philosophical anthropology concerns that which can be known through natural human reason. A theological anthropology concerns that which is known through faith. Nonetheless, the two perspectives are not at odds and should, from the theological perspective, be complementary. One expression of this principle is Aquinas' statement that it is necessary for the faithful to accept prerequisites of the faith that can be known through natural reason.

It is necessary for man to accept on faith not only what lies beyond reason but also what can be known through reason. There are three reasons for this: First, in order that man might come more quickly to the cognition of God. For the science to which it belongs to prove that there is a God and other things of this sort about God is the last one to be proposed to men to learn, since it presupposes many other sciences. And in this way a man would arrive at the cognition of God only after a long stretch of his life. Second, in order that the cognition of God might be more widespread (*communior*). For there are many who are unable to be proficient in the study of science, either because of mental dullness (*propter hebetudinem ingenii*), or because of the other occupations and necessities of temporal life, or even because of laziness in learning. And they would be completely deprived of the cognition of God if the things of God were not proposed to them through the mode of faith. Third, for the sake of certitude. For in divine matters human reason is quite deficient (*multum deficiens*). An indication of this is that *the philosophers, in investigating human matters by natural reason, have fallen into many errors and have disagreed among themselves*. Therefore, in order that men might have unshakable and certain cognition of God, it was necessary that divine matters be handed down to them through the mode of faith—spoken, as it were, by God, who cannot lie.⁹⁴

⁹⁴ Thomas Aquinas, *New English Translation of St. Thomas Aquinas's Summa Theologiae*, trans. Alfred J. Freddoso (n.d.), accessed September 28, 2025, <https://www.freddoso.com/summa-translation/TOC.htm> II-II q. 2, a. 4, emphasis added.

Aquinas thus holds that it is necessary for most persons of faith to accept on faith (faith based on revelation) things that in principle can also be known through reason, owing to the difficulty in coming to know them through reason. He allows that some people may “see” (in the sense of *know*) what others have to accept on faith. He selects the example of God’s existence, which Aquinas believes may be proven philosophically.⁹⁵ This natural knowledge is difficult to acquire, however, because it presupposes that one has mastered many other sciences, so true knowledge of God’s existence would be limited to those who have the capacity and opportunity to engage in lengthy study. Knowledge of God’s existence through faith allows God’s existence to be known more quickly and by a greater number of persons than through the exercise of metaphysical acumen. Further, it complements the human mastery of the study, by offering certitude where reasoning humans are prone to disagreement.

This serves to illustrate the theological principle that that which is known through natural reason is distinct from but complementary with what is known through faith. From the theological perspective, there is a convergence between natural reason and faith, between philosophy and theology. From the theological perspective, a correct philosophical perspective will align with a theological perspective.

Moreover, Aquinas thinks that theology can show that what is known through faith can be shown to be philosophically fitting. An example is the notion of the resurrection of the body. Aquinas thinks that the immortality of the soul can be proven philosophically,⁹⁶ but that the disembodied soul will be highly limited.

As has already been explained (a. 5-7), all the powers of the soul have the soul alone as their principle. However, certain of the powers, viz., the intellect and the will, have the soul alone as their subject, and powers of this sort must remain in the soul after its body has been destroyed. By

⁹⁵ *ST* II-II q. 1, a. 5.

⁹⁶ *ST* I-I q. 75, aa. 2, 6.

contrast, certain of the powers, viz., all the powers of the sentient and nutritive parts of the soul, have the conjoined being as their subject. Now when a subject is destroyed, its accidents cannot remain. Hence, when the conjoined being is corrupted, powers of the sort in question do not remain in actuality but remain only virtually in the soul as in their principle or root. And so the claim made by some, viz., that powers of this sort remain in the soul even after the body is corrupted, is false. And the further claim that the acts of these powers remain in a separated soul is all the more false, since no action belongs to these powers except through a corporeal organ.⁹⁷

Death, though in one sense natural to humans, is in another sense unnatural because the human soul is supposed to operate with the body.⁹⁸ The theology of the resurrection of the body, therefore, may be shown to be fitting given what is known philosophically about the human being.

For the purposes of this thesis, the theological perspective will have observations on human insight distinct from but congruent with the philosophical perspective on insight and lessons from neurobiological science, psychology, narrative, and metaphor.

As a point of departure, the meaning of the terms “understanding” and “insight” will be explored.

3.2 First-Person Aspects of Understanding

3.2.1 Introducing Terms: Understanding and Insight

The use of the phrase “I understand” is commonplace in daily conversation, but what, precisely, is meant by “understanding” and the claim to understand?

A sense of the difficulty may be gained from perusing the definitions of “understanding” in the *Oxford English Dictionary (OED)*. The first definition given to which other definitions are related or else obscure is,

⁹⁷ *ST* I-I q. 77, a. 8.

⁹⁸ *ST* I-II q. 85, aa. 5, 6.; II-II q. 164, a. 1.

To comprehend; to apprehend the meaning or import of; to grasp the idea of.⁹⁹

But comprehension and apprehension are themselves difficult to define, with the *OED* describing these words largely in terms of understanding or grasping, or some similar related metaphor such as, “catching holding of”. Hence, one of the most important metaphors for understanding and associated words involves the embodied experience of taking hold of something, conveying a sense of an object’s shape, texture, heft and volume, a sense of what it is that one holds.

In the cognitive life of infants, it should be noted that this action of grasping is also closely linked to exploring the world and seeking, at least tacitly, exploring answers to the question, “What is it?”¹⁰⁰ The fact that infants reach out to explore the world does not necessarily imply they have a conscious sustained attention to the question, “What is it?” or of having an answer to that question. To borrow an example from Heidegger, one does not attend to the hammer that one uses to strike the nail since the hammer is present to us as ready-to-hand.¹⁰¹ But one’s actions reveal that one clearly understands what the object is for in that context. Only when the hammer is a novel object does one approach it by asking, explicitly or implicitly, “What is it?” But once the hammer is grasped, and its role is understood, the self-awareness of grasping and understanding will tend to recede into the background, fading from conscious awareness, though understanding clearly endures.

⁹⁹ *OED Online*, “under’stand, v.” (Oxford University Press), accessed November 19, 2018, <http://www.oed.com/view/Entry/212085>.

¹⁰⁰ The relationship between grasping activity and cognitive development in infants is well-studied. For a study that also examines the connection between grasping and caregiver engagement, see Vasudevi Reddy and Sebo Uithol, “Engagement: Looking beyond the Mirror to Understand Action Understanding,” *British Journal of Developmental Psychology* (England) 34, no. 1 (2016): 101–14, <https://doi.org/10.1111/bjdp.12106>.

¹⁰¹ This example of Heidegger’s concept of readiness-to-hand is taken from Dreyfus. Dreyfus, *What Computers Still Can’t Do*, xxxii.

The more general lesson from this example is that although the transition to new understanding may involve a change of which one is cognitively aware, this awareness is not itself bound up with the state of understanding. Indeed, a second definition of understanding in the *OED*, namely I.1.b,

is to be thoroughly acquainted or familiar with (an art or profession, etc.);
to be able to practise or deal with properly.

This notion of being thoroughly acquainted or familiar with suggests that the scope of what is understood is often tacit and outside conscious awareness, for one is not said to lose familiarity with something or ability to practise something when one is not aware of it.¹⁰²

This distinction between the transition to understanding and the state of understanding implies the possible need for a distinct concept to describe this change. Indeed, the English language has a special word for the transition to understanding, namely *insight*.

The *OED* describes insight as,

1.a. Internal sight, mental vision or perception, discernment ... 2.a. The fact of penetrating with the eyes of the understanding into the inner character or hidden nature of things; a glimpse or view beneath the surface; the faculty or power of thus seeing.¹⁰³

These definitions focus greatly on the act of seeing as a metaphor for insight, especially seeing deeply into the inner character or nature of things. The phrase “eyes of the understanding” implies also a distinct cognitive faculty that resembles physical vision.¹⁰⁴

¹⁰² “Understand, v.: Oxford English Dictionary,” accessed November 27, 2018, <http://ezproxy-prd.bodleian.ox.ac.uk:2355/view/Entry/212085?rskey=iUGDql&result=1&isAdvanced=false#eid>.

¹⁰³ *OED Online*, “insight, n.1.”

¹⁰⁴ The resemblance of insight to sight may be based on the natural prominence of sight for our perceptions. Aristotle observes “For not only with a view to action, but even when we are not going to do anything, we prefer sight to almost everything else. The reason is that this, most of all the senses, makes us know and brings to light many differences between things”, Aristotle, “Metaphysics,” in *The Complete Works of Aristotle*, ed. Jonathan Barnes, trans. W. D. Ross, Revised edition, vol. 2 (Princeton, N.J: Princeton University Press, 1984), 980a25. ; In a sense, there is a *via negativa* as well as a *via positiva* in all knowing. Knowing what something is depends equally on knowing what it is not. One cannot see a black chair without distinguishing it from the beige walls against which it rests. Further, to know what something is implies knowing what it is not. This claim gains support from the observation in the psychological literature that children begin drawing by making outlines of objects, and only

At this stage it is worth supplementing this lexical treatment of insight with a brief examination of how the term “insight” is described in the psychological literature. In psychology, insight has been studied in the form of “aha!” experiences.¹⁰⁵ The emphasis here is on the ability to spontaneously shift one’s attention to what is relevant to solve a problem, ignoring irrelevant details and recognising a coherent relationship between the tools provided in the test that can be used to adequately solve the problem.

A representative description of insight is given by Bowden et al.

The term “insight” is used to designate the clear and sudden understanding of how to solve a problem. Insight is thought to arise when a solver breaks free of unwarranted assumptions, or forms novel, task-related connections between existing concepts or skills.¹⁰⁶

Here, a distinction is drawn between insight and problem-solving. The insight is not necessarily the solution to the problem in itself, but a clear and sudden understanding by which the problem is solvable. Further, the insight allows one to break free of unwarranted assumptions, apprehending what is essential. One more correctly grasps the context and conceptualises the details, which includes the recognition that there exist details that are mistaken or that one’s overall sense of the problem and its possible solution are limited, and a new sense of the whole instead takes its place.

at a later stage of development do they advance to the stage of filling in details. The above passage from Aristotle regarding the delight humans take in sight apart from practical benefits complements the preceding definitions from the OED by observing that it is characteristic of humans to delight in knowing for its own sake. Similarly, Polanyi traces the origin of inquiry to the tension of uncertainty and the pleasure experienced when satisfying self-set standards for the quality of the solution, as found in other animals including chimpanzees. “These intellectual tastes of the animal prefigure, no doubt, the joys of discovery which our articulate powers can attain for man, but in the animal they do not remotely approach these joys in scope and elevation.” Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy*, Corrected Edition (Chicago: University Of Chicago Press, 1974), 133. Humans differ in that the animal pleasures of trick-learning, motoric-learning, and latent-learning are supplanted by intellectual appetites for sophisticated learning.

¹⁰⁵ Jessica I. Fleck et al., “Insight,” in *The Oxford Handbook of Cognitive Psychology*, 1 edition, ed. Daniel Reisberg (Oxford University Press, 2013).

¹⁰⁶ Edward M. Bowden et al., “New Approaches to Demystifying Insight,” *Trends in Cognitive Sciences* 9, no. 7 (2005): 322–28, <https://doi.org/10.1016/j.tics.2005.05.012>.

Related to the suddenness with which insight is acquired is another feature that I have noted above that is also recognised by the psychological literature, that insight occurs discretely and is distinct from problem solving by discursive reasoning,

The notion of sudden insight is related to the distinction between discrete, all-or-none, information processing and continuous or incremental processing...

Research contrasting problem solving by insight and by analytic “noninsight” search strategies has identified distinguishing cognitive and neural mechanisms...and has been identified as an important characteristic of creative thought...¹⁰⁷

Psychological researchers agree that there exists a distinction between a conclusion gained through a stepwise process of reasoning and that which occurs suddenly, in what I have called a step change. In other words, while a process of discursive reasoning allows one to reach a conclusion by a series of steps, often by applying rules, an insight yields new understanding in a single step without an application of formulaic techniques. Further, there exists support from neuroscience that undergoing an insight is a distinct neurophysiological process. Although it remains controversial, the association of insight experiences with right hemisphere activations may provide neurophysical evidence for the holistic nature of insight and its relationship to grasping contexts. Psychological research corroborates this chapter’s philosophical description of insight and supports the notion that it is a distinct kind of cognitive act apart.¹⁰⁸

¹⁰⁷ John Kounios et al., “The Origins of Insight in Resting-State Brain Activity,” *Neuropsychologia* 46, no. 1 (2008): 2, <https://doi.org/10.1016/j.neuropsychologia.2007.07.013>.

¹⁰⁸ “In fact an extensive body of research now indicates that insight, whether mathematical or verbal, the sort of problem solving that happens when we are, precisely, not concentrating on it, is associated with activation in the right hemisphere, mainly in the right anterior temporal area, specifically in the right anterior superior temporal gyrus, though where there are high levels of restructuring involved there is also activity in the right prefrontal cortex. Insight is also a perception of the previous incongruity of one’s assumptions, which links it to the right hemisphere’s capacity for detecting an anomaly.” Iain McGilchrist, *The Master and His Emissary: The Divided Brain and the Making of the Western World* (Yale University Press, 2009), 65; See also Daniel Reisberg, ed., *The Oxford Handbook of Cognitive Psychology*, 1 edition (Oxford University Press, 2013), 786–88.

These examples from the common and psychological uses of the terms “understanding” and “insight” provide an initial sense of what is meant by understanding and insight. According to common usage, understanding is a cognitive faculty that is associated with and described in terms of one’s own basic sensory and kinaesthetic experiences, namely seeing and grasping, and seems to refer to both basic components of cognition as well as to cognitive achievement marked by familiarity and potentially yielding skilful practice. Likewise, insight is described as a kind of sight in which the eyes of the understanding penetrate to the heart of things. Moreover, these cognitive and embodied aspects of the experiences of understanding and insight have been recognised as philosophically interesting, as Heidegger’s notion of readiness-to-hand shows. Further, these experiences and the descriptions of them appear apposite to early and important aspects of child cognitive development. Moreover, psychological research into experiences of insight supports the claim that insight is distinct from stepwise reasoning and involves an experientially simpler kind of apprehension.¹⁰⁹ With this initial sketch of the terms developed, a more thorough exploration can be made of first the philosophical and later the psychological research into these phenomena. The first thinkers this thesis explores are Bernard Lonergan and Michael Polanyi, whose epistemologies both make a detailed study of understanding.

3.2.2 The General Properties of Insight: Lonergan’s Analysis

In connection with the metaphor of insights as granting new sight, Lonergan describes the most notable property of insights, their fruitfulness in generating what he calls higher viewpoints,

Single insights occur either in isolation or in related fields. In the latter case, they combine, cluster, coalesce, into the mastery of a subject; they ground sets of definitions, postulates, deductions; they admit application to enormous ranges of instances. But the matter does not end there. Still

¹⁰⁹ Fleck et al., “Insight.”

further insights arise. The shortcomings of the previous position become recognized. New definitions and postulates are devised. A new and larger field of deductions is set up. Broader and more accurate applications become possible. Such a complex shift in the whole structure of insights, definitions, postulates, deductions, and applications may be referred to very briefly as the emergence of a higher viewpoint.¹¹⁰

Lonergan observes that insights ground definitions, postulates, and deductions, which is to say that insights ground the cognitive tools for explicit reasoning, but he states that insights have a further attribute. Insights expand the capacity for explicit reasoning, allow positions to become broader and more accurate, and they yield higher viewpoints. The language indicates that the superior cognitive state insights yield is best described through the metaphor of seeing, and implicitly that seeing has a personal character, since a viewpoint always requires some subject to have a perspective.

Lonergan illustrates a simple case of the emergence of a higher viewpoint with the example of a child learning about numbers, beginning with simple counting and addition. By cumulatively adding “1” to each new number, the child can grasp that there is an unbounded set of such whole numbers, a set to which mathematicians have given the name the *natural numbers*.¹¹¹ When further operations are introduced, such as adding or multiplying the whole numbers, as well as the reverse of these operations, namely subtraction and division, these operations introduce the need to expand the concept of a number. If, for example, one subtracts a number from itself, then there is a need for zero; if one subtracts from a number another number larger than itself, then there is a need for negative numbers. These operations lead towards a higher viewpoint of the concept of number that includes the natural numbers, zero, and the negative natural numbers, an expanded set to which mathematicians have given the

¹¹⁰ Bernard Lonergan, *Insight: A Study of Human Understanding, Volume 3*, 5th edition, ed. Frederick Crowe S.J and Robert Doran S.J (University of Toronto Press, 1992), 37–38.

¹¹¹ Lonergan, *Insight*, 38–42.

name *integers*. In a similar way, all the operations of division introduce the need for fractional numbers, involving a division of two integers which leads to the higher viewpoint of *rational numbers*. But further operations also hint at the need for fractional numbers which cannot be expressed as the division of any two finite integers, such as the square root of two, leading to the *irrational numbers*. Additional steps expand the concept of number to encompass *transcendental numbers*, such as pi, and also the *complex numbers*, comprising combinations of real and imaginary numbers.

At each progressively higher viewpoint of the meaning of number, what is grasped at a lower level is preserved within an ever larger and more fruitful context, and each of these viewpoints hint at further expansions.¹¹² Similar patterns can be found, for example, in the development of the physics of planetary motions, with Keplerian laws of planetary motion encompassed and explained by Newtonian physics, which was itself later encompassed and explained in the more complete account of gravitation provided by Einstein's theory of relativity. The general lesson is that the succession of higher viewpoints from insights has both an ascending and generally preservative character. As insights occur, disparate things are drawn together and seen as being connected and unified from a new perspective.¹¹³

To be disposed to have insights, especially those yielding higher viewpoints, access to sensory phenomena is not sufficient, one must have the appropriate cognitive abilities. Many subjects may perceive the same phenomena, but it may be that only in some does insight occur. This variance implies, first, that there exists some cognitive faculty distinct from perception that

¹¹² Lonergan, *Insight*, 43–50.

¹¹³ It should be noted that an alternative perspective on the development of ideas is developed by Thomas Kuhn in *The Structure of Scientific Revolutions*. Kuhn emphasizes discontinuity in moments of what he calls paradigm shift, of which he takes the shift from one world view to another. He resists the claim that one can say that Newtonian physics is subsequently known as a special case of Relativity, since fundamental assumptions of Newtonian physics must be rejected. Thomas S. Kuhn, *The Structure of Scientific Revolutions: 50th Anniversary Edition*, Fourth edition, with Ian Hacking (University of Chicago Press, 2012), 99–103.

allows one to have insights. Second, insightfulness, or the propensity to have insights, does not appear to be evenly distributed. Some individuals appear more likely to have insights than others, and that about which they have insights appears to differ. Lonergan's explanation is that one must be habituated to asking the question "why?" and be alert to the possibility of problems for one's intelligence. Such inquisitiveness is a necessary condition of insight, for the same sensible phenomena may be experienced by many but only catalyse insight in a few. I add that one's interest and purposes give direction to one's questioning. A physicist is concerned with asking different sorts of questions than a literary critic. Third, one must suitably present the questions one asks in a way that is conducive to having an insight.¹¹⁴

The attributes conducive to insight are represented in the story of Archimedes' discovery in the baths of Syracuse, which is an archetypal example of an insight experience. He was an accomplished natural philosopher to whom was given the task of determining the composition of King Hiero's crown. Many others had entered the baths and perhaps noticed the actions of physical bodies in the water, but he alone had a problem the solution of which entailed the recognition of physical principles instantiated in the common happenings of the bath. Specifically, Archimedes alone realised that the rise of the bath water when a body entered it could be used to determine how much water had been displaced, which is equivalent to the body's volume. Innumerable details which might have fixed his attention were present in the motions of the water, and he required a clear question to which some of those details, and not others, might be related. In this, he discerned that he could measure both the volume of the crown by how much it displaced water and the mass of the crown with scales. Together, they allowed him to determine what we would today call the density of the crown, enabling him to

¹¹⁴ Lonergan, *Insight*, 29.

determine whether it was pure gold or had been debased with other metals. This insight both allowed him to fulfil his assignment for Hiero and to articulate physical properties with universal application.¹¹⁵ Lonergan enumerates several properties of insight contained in this story.

The first property of insight is that it brings joy. When Archimedes realised the solution to his problem, he leapt from the bath and rushed into the streets crying, “*Eureka!*” (I have it!).¹¹⁶

A second property is that insight comes as a release of the tension of inquiry. Archimedes’ state as he pursued the question posed by Hiero was typified by this striving to satisfy the desire to understand, and its satisfaction came as a release from this striving. This tension is sustained for as long as the desire to understand remains unfulfilled, and it yields conscious and unconscious efforts that may terminate in the satisfaction of this desire. The tension of inquiry arises when a question is posed to oneself, perhaps implicitly, to which one does not yet have an answer. It is evinced by effortful seeking for answers to a vast range of enquiries, from the lifelong project of a scholar to a child striving to understand the seasonal changes that befall a tree. This release from tension is often joyful, as illustrated by Archimedes’ enthusiasm as he rushed into the streets.¹¹⁷

The third property of insight is that the release from tension comes suddenly and unexpectedly. The tension of inquiry can persist when one’s mind is occupied with other matters, as Archimedes was taken up with the task of bathing himself when insight came, and

¹¹⁵ Lonergan, *Insight*, 30.

¹¹⁶ Bernard Lonergan, *Insight: A Study of Human Understanding, Volume 3*, ed. Frederick Crowe S.J and Robert Doran S.J, 5th edition (Toronto: University of Toronto Press, 1992), 27–28.

¹¹⁷ Bernard Lonergan, *Insight: A Study of Human Understanding, Volume 3*, ed. Frederick Crowe S.J and Robert Doran S.J, 5th edition (Toronto: University of Toronto Press, 1992), 28. Lonergan, *Insight*, 28.

it can endure even when one is not awake, hence the possibility that a solution to a problem will come to one in a dream.¹¹⁸

When the insight comes one passes at once from not-understanding to understanding, so the tension of inquiry is released. The change is analogous to a step-change in physics such as a phase-change of water from solid to liquid. There is not a stable intermediate state between understanding and a failure to understand. By way of analogy, consider the famous duck-rabbit optical illusion popularized by Wittgenstein.¹¹⁹ The shift from seeing the image as a duck to seeing it as a rabbit occurs suddenly, and one's perception of the image as one of these two images is only stable in one of these states, as it were.¹²⁰ One might notice attributes of the rabbit or attributes of the duck concurrently, but this perception tends to collapse into a perception of the object as a duck or a rabbit. Though what one perceives might shift between the stable states of perceiving a duck or a rabbit, one rests in perceiving an object of some one kind at any one time. An analogous process occurs in acts of insight.

Unlike the example of Wittgenstein's duck-rabbit, however, the change in understanding is progressive. Save through injury or a weakening of the power of memory, one does not pass from understanding to a lack of understanding. One who has grasped addition does not spontaneously revert to viewing the relationship between numbers and their sums as unknown.¹²¹ Even in the case of the optical illusion of Wittgenstein's rabbit, while one can

¹¹⁸ Lonergan, *Insight*, 28; OP A. G. Sertillanges and SJ James V. Schall, *The Intellectual Life: Its Spirit, Conditions, Methods*, Reprint edition, trans. Mary Ryan (The Catholic University of America Press, 1998).

¹¹⁹ Ludwig Wittgenstein, *Philosophical Investigations* (1958), 195, http://archive.org/details/philosophicalinvestigations_201911.

¹²⁰ One might, with effort, perceive the image as a kind of chimera, a fantastic creature with two mouths, but this would be another distinct thing which the image may be seen as, and seeing it as such would also be mutually exclusive with seeing it as a duck or as a rabbit.

¹²¹ Once one has understood a true proof, say, one cannot have a new act of understanding that reverses one's understanding. One could realise that the original proof was invalid or that one or more premises were false, or its meaning can be changed if the proof is discovered to be part of some larger context, but one cannot view a proof one believes to be true and valid and have an unprompted shift from understanding it to not understanding it.

change what one perceives in the image, one cannot reverse one's understanding that the image may be correctly perceived in more than one way. The experience of gaining a new understanding is similar to that of the gestalt shift that may be elicited by Wittgenstein's optical illusions, but unlike the gestalt shift in this example, the shift to understanding is not reversible.

The fourth property of insight is that it is the "pivot" between images and concepts. Insights come with respect to particular questions, but their applicability is general. Archimedes' discoveries of the relationships between the incompressibility of water, volume, and buoyancy apply not just in the bath houses of Syracuse, but generally wherever liquids and objects that may be immersed in these liquids exist.¹²²

The fifth property of insight is that it becomes part of the "habitual texture of one's mind". Once an insight has been had, one's newfound understanding typically fades from conscious awareness. Lonergan observes that an insight into an object may require an instant of inspiration but that the subsequent understanding of the same object comes easily and with little or no effort, and this ease tends to persist. Inspiration and the joy of successful grasping occur when the step-change is fresh, as it were, and afterward one has no need of inspiration to apprehend that which is now obvious.¹²³

A sixth property is that insight links the universal and the particular, resulting in an enlargement of one's habitual understanding. The grasp of universal principles allows one to solve further definite problems by application of the principles discovered in an act of insight. Archimedes' insight allowed him to solve Hiero's problem, but it equally allowed him to analyse any comparable object in the same way. The case study of Archimedes illustrates that insight into a universal principle leads to recognising its implications for particular instances.

¹²² Lonergan, *Insight*, 30.

¹²³ Lonergan, *Insight*, 30.

The enlargement of one's habitual understanding prepares the way for subsequent insights by allowing one to anticipate new possible understandings in terms of these newly understood principles. For instance, the grasping of the definition of a circle provides one with the realisations necessary to construct an equilateral triangle and the incipient grasp of the necessary equality of its sides and angles that may be subsequently defined in a discursive proof.¹²⁴

How then does an act of insight occur? By what causal chain does the tension of inquiry eventuate in insight? The tension of inquiry results from the human impulse to ask, at least implicitly, "why is this?" or "what is this?" Children are consumed with the desire to know the answer to these questions, which becomes even more evident as they acquire speech and begin interrogating their parents.¹²⁵ Adults who have retained some measure of the child's inquisitiveness, when labours end, waken to that which they can anticipate understanding but do not yet understand.¹²⁶ They have the advantage over the child of being able to formulate their question precisely to discipline the play of their intelligence.

In both the disciplined activity of the adult and the spontaneous questioning of the child, the basic progression to understanding is the same. As one asks a question, explicitly or implicitly, the desire to know is aroused, which leads to what Lonergan calls the "tension of

¹²⁴ In a similar vein, Philip Davis and Reuben Hersh note that for the mathematician there is an inherent disjunct between the mathematicians' understanding of a proposition and the symbols they use for the propositions. "Interpreting a symbol is to associate it with some concept or mental image, to assimilate it to human consciousness. The rules for calculating should be as precise as the operation of a computing machine; the rules for interpretation cannot be any more precise than the communication of ideas among humans." The mathematicians' understanding of mathematical symbols requires the imaginative faculties and the grasping of the mathematical proposition is nonformulaic and distinct from the symbolic representation. Philip J. Davis and Reuben Hersh, *The Mathematical Experience* (Birkhauser, 1982), 125.

¹²⁵ A young child given a novel object, for instance, will grasp it in their hands while looking at it intently and then begin to manipulate it and put it into contact with other objects in the environment. By such operations, the child tries to understand its properties, uses, and differences from other objects, even before the child can ask its parents for a name and an explanation of the object.

¹²⁶ "When an animal has nothing to do, it goes to sleep. When a man has nothing to do, he may ask questions. The first moment is an awakening to one's intelligence. It is release from the dominance of biological drive and from the routines of everyday living. It is the effective emergence of wonder, of the desire to understand." Lonergan, *Insight*, 34.

inquiry”. While this tension remains, intelligence attempts to supply an adequate answer using images inventively to aid one in constructing the concepts that satisfy one’s original desire.¹²⁷ Lonergan thinks that insight begins with what he calls a “hint”. This is not a complete insight, but an intuition as to how one’s effort to understand should proceed.

These insights ground the discursive reasoning by which great intellectual systems are developed. Intellectual systems are founded on definitions and postulates that are generated from the process of asking questions, gaining insights, forming and testing concepts that express those insights, and generating imaginations that ground those concepts.

As stated above, the succession of insights allows one to achieve a higher viewpoint. If insights occur within a related field, they may become unified by a further act of understanding. As Lonergan observed, the insight that unifies other insights and creates the conditions for still further insights changes the explicit structure of associated articulate knowledge.¹²⁸

Lonergan’s sense of the terms insight and understanding encompasses many cognitive virtues that are differentiated in Aquinas’ writings. Aquinas considers there to be natural virtues that may be developed through one’s own powers and virtues that may be infused by God, and the latter can compensate for weaknesses of the former. Within Aquinas’ scheme, the natural virtue of *ars*, the rational capacity to make, would cover the ability to develop technology. The

¹²⁷ These images are not identical with the concept as can be seen from the example of a basic geometrical concept: the line. No image of a line can ever be completely faithful to a mathematical description of it, for a line is a concatenation of points that is one dimensional and lacking in all thickness. All images of a line contain this thickness, but it is nonetheless understood to represent the line that it imperfectly approximates.

¹²⁸ The drive to seek a unifying understanding is a common drive in mathematics. “Unification, the establishment of a relationship between seemingly diverse objects, is at once one of the great motivating forces and one of the great sources of aesthetic satisfaction in mathematics. It is beautifully illustrated by the formula of Euler which unifies the trigonometric functions with the ‘power’ or ‘exponential functions’.” Euler’s formula is the expression of his discovery that exponential algebra can be described in terms of trigonometry and, likewise, trigonometry can be described in terms of exponential algebra. Euler’s recognition of a unity between two seemingly distinct kinds of mathematical objects can be described as gaining a higher viewpoint in the Lonerganian sense, for Euler saw that what seemed to be two disparate mathematical objects were in fact the same. Davis and Hersh, *The Mathematical Experience*, 198–99.

theoretical sciences underlying the art of technological development are covered under what Aquinas calls *scientia*. Practical use of technology for the true human good requires the acquired virtue of *prudentia*. The practical use of technology within the journey to God requires, from a theological perspective, infused *prudentia*. The associated Gift of the Holy Spirit, Counsel (*consilium*) would also be required. The first principles of theoretical and practical (which approximately covers what is meant by ethical) thinking would be covered by *intellectus*. The philosophical sense of how everything fits together would be the acquired intellectual virtue of wisdom (*sapientia*). The theological vision of how everything fits together is *sacra doctrina* and *sapientia* (partly acquired and partly infused). The judgement by connaturality of how one's study of, development of, and use of or abstention from technology serve one's ultimate good *sub specie aeternitatis*, requires the Gift of Wisdom.¹²⁹

For instance, Aquinas speaks specifically about *intellectus*, a term usually translated as insight, as the virtue which pertains to grasping first principles and *scientia* as the virtue of having mastery of a particular branch of learning. In Lonergan's system, the term insight encompasses both.

...in a sense somewhat different from Kant's, every insight is both a priori and synthetic. It is a priori, for it goes beyond what is merely given to sense or to empirical consciousness. It is synthetic, for it adds to the merely given an explanatory unification or organization. It seems to follow that insight into insight will yield a synthetic and a priori account of the full range of synthetic a priori components in our cognitional activity.

¹²⁹ "So, then, it belongs to the wisdom that is an intellectual virtue to have right judgment about divine things by the inquiry of reason, whereas it pertains to wisdom as a gift of the Holy Spirit to have right judgment about those things by a sort of connaturality with them. For instance, in *De Divinis Nominibus*, chap. 2, Dionysius says that Herotheus is perfect in divine matters, 'not only in learning divine things, but also in experiencing them (*sed et patiens divina*).' Now this sort of connaturality or communion of affection (*compassio*) in divine things is effected by charity, which unites us to God—this according to 1 Corinthians 6:17 ('He who adheres to God is one spirit')." *ST II-II* q. 45, a. 2. In other words, the gift of wisdom is a right judgement of things from connaturality or a communion of affection. This is to experience things and affectively respond to them in the right way, as a result of being united to God in charity. It may be described metaphorically as aligning one's perspective with God in friendship.

...a unification and organization of other departments of knowledge is a philosophy. But every insight unifies and organizes. Insight into insight, then, will unify and organize the insights of mathematicians, scientists, and men of common sense.¹³⁰

On Lonergan's terms, one proceeds from an insight into the first principles through the process of acquiring further learning through a series of not just calculative steps but of insights yielding progressively higher viewpoints. In Lonergan's terms the function of acquired wisdom, *sapientia*, is fulfilled by the inquiry of Metaphysics.

...one cannot unify and organize knowing without concluding to a unification and organization of the known. But a unification and organization of what is known in mathematics, in the sciences, and by common sense is a metaphysics. Hence, in the measure that insight into insight unifies and organizes all our knowing, it will imply a metaphysics.¹³¹

For Aquinas, the virtue of wisdom covers the knowledge of how other areas of practical and theoretical knowledge fit together. For Lonergan, the insights that comprise the highest order of natural higher viewpoints will yield a unity between other forms of knowledge, and this is Metaphysics.

Aquinas defends an exceptionally complex scheme of virtues and gifts, which mirrors the unique complexity of the human psyche and of human, social life, and reflects the importance for Christians of natural-law morality alongside the vocation to live as God's children destined to share God's own bliss. The virtues of fortitude and temperance are needed to integrate the "passions of the soul" (roughly speaking, the emotions) with reason, and are relevant to an embodied creature with nervous and endocrine systems.¹³² Justice concerns human relations with others, so it is inherently social.¹³³ The infused virtues and the infused

¹³⁰ Lonergan, *Insight*, 5.

¹³¹ Lonergan, *Insight*, 5.

¹³² *ST* I-II q. 59.

¹³³ *ST* II-II q. 58, a. 2.

“theological” virtues of faith, hope, and charity,¹³⁴ are required for the journey to a divine goal and are by their nature reliant on a personal relationship with God, to whose guidance the Gifts attune persons. The whole scheme requires looking beyond the limits of individual cognitive enhancement.

From this excursus, it may be seen that while this thesis uses the terms of insight and understanding in Lonergan’s sense, there is significant conceptual overlap between Lonergan and Aquinas, and Thomas’ conceptions of the philosophical and theological virtues of insight may be used to elaborate specific aspects of understanding and insight as they are discussed by this thesis.

In this section insight and its qualities have been examined. The importance of language and the connection between language and experience has been indicated but not yet explored in detail. The next section explores the role of embodiment as grounding understanding and the connection between embodied understanding and language.

3.2.3 Embodiment as a Ground of Understanding: Lessons from Metaphor and Narrative

Even before they acquire speech, children’s play shows ample evidence of the acquisition of understanding. Before becoming able to vocalise her knowledge of a wooden block, the child demonstrates the ability to take hold of the block, to place it to support other blocks, or perhaps to use it as a tool for pushing around other objects. The child already demonstrates knowledge of the block as something upon which she can act and by which it can act, thereby showing implicit knowledge of its properties (an object lacking in solidity, for instance, would be a poor choice as the foundation for a tower of toy blocks). Polanyi identifies

¹³⁴ *ST* I-II q. 62.

such tacit knowledge (in this case evinced in successful tool-use) as prior to any explicit knowledge.

Tacit knowledge is evinced in skilful acts which one may perform prior to any ability to say what it is one does or how one achieves it. Prior to something being the object of theorising or explicit classification, it is implicitly grasped as something affording possibilities for action. How it is that humans are able to do this remains mysterious, but it appears to be relevant to the ability to answer the question “What is it?” that is the explicit utterance of the desire to know. Prior to being able to say what one knows of something one must already know a great deal that either is not or cannot be expressed. The child from the above example cannot (yet) articulate the block’s weight, material, colour, or shape (to name a few properties) but they can feel its weight, recognise its colour and shape, and feel its texture, allowing experiences of this object by which they know it in their minds and evince their qualitative understanding of it by treating it differently than they do other objects.

An illustration of this tacit knowledge may be gained from the famous thought experiment of Frank Jackson, Mary in the Black and White Room.¹³⁵ In this thought experiment, a woman by the name of Mary has spent her entire life in a room in which all objects are grayscale, and, by stipulation, she has never actually seen any colour. Yet she also studies optics and visual perception such that she becomes an expert on physical theories describing colour and perception on it. Mary knows a great deal about the theory of colour, redness, say, but when she is finally granted an opportunity to leave her black and white room, upon immediately encountering a red object, she says, “So that is what red is.” It seems that for

¹³⁵ Frank Jackson, “Epiphenomenal Qualia,” *The Philosophical Quarterly* (1950-) 32, no. 127 (1982): 130, <https://doi.org/10.2307/2960077>.

all her theoretical knowledge of colour, Mary has still learned something upon seeing the colour red for the first time.

If there exists a kind of non-propositional knowledge more fundamental than language, how should it be described? Moreover, how does it relate to language, to the simple act of grasping what a thing is, and to one's own embodied experience? Important clues may be gained from the examination of metaphors.

According to the analysis of McGilchrist in his study *The Master and His Emissary*,

...the importance of metaphor is that it *underlies all forms of understanding whatsoever*, science and philosophy no less than poetry and art.¹³⁶

The claim is that metaphor is a necessary condition of and deeply involved in acts understanding in all areas of human knowing. In other words, McGilchrist is claiming that metaphor is absolutely indispensable to all forms of understanding.

The key significance and nature of metaphor is itself suggested in the word “metaphor”.

The word metaphor implies something that carries you across an implied gap (Greek *meta-* across, *pherein* carry).¹³⁷

Metaphor carries meaning across the gap between language, which is to say, symbol systems and representations, and embodied experience. The word “metaphor” itself has origins in its Latin roots indicating that the concept of metaphor itself is fundamentally expressed in terms of metaphor, and this synaesthetic experience of bridging a gap captures the core experience of meaning between symbols and bodily experience. Metaphor is the means by which human language is possible, as that which allows signs and sounds to carry meaning. In the absence of metaphor, the expansion of human cognitive capacity through symbol systems as described by

¹³⁶ McGilchrist, *The Master and His Emissary*, 79.

¹³⁷ McGilchrist, *The Master and His Emissary*, 116.

Polanyi could not occur. Were metaphor to be lost, symbol systems would risk becoming mere patterns without meaningful content.

The carrying across from language to the experiential world occurs both at the level of context and at the level of particular things,

Speaking metaphorically, one might say that language is open to carry us across to the experiential world at the “top” and at the “bottom”.

At the “top” end, I am talking about any context - and these are not by any means to be found in poetry alone - in which words are used so as to activate a broad net of connotations, which though present to us, remains implicit, so that the meanings are appreciated as a whole, at once, to the whole of our being, conscious and unconscious, rather than being subject to the isolating effects of sequential, narrow-beam attention. As long as they remain implicit, they cannot be hijacked by the conscious mind and turned into just another series of words, a paraphrase.¹³⁸

Metaphors allow the whole of a context to be grasped without being expressly articulated through the web of connotations and shared experiences that are evoked through some experience or utterance.

Metaphor also carries across from language to experience at the bottom level of language, in the sense of allowing a link between a word and some kind of embodied experience.

...every word, in and of itself, eventually has to lead us out of the web of language, to the lived world, ultimately to something that can only be pointed to, something that relates to our embodied existence. Even words such as “virtual” or “immaterial” take us back to realities of a man’s strength (*vir-tus*), or the feel of a piece of wood (*materia*). Everything has to be expressed in terms of something else, and those something elses have to come back to the body.¹³⁹

Every word must ultimately bottom out in some sort of embodied experience, however circuitous the route. This is because humans exist in the world as bodies and the world and

¹³⁸ McGilchrist, *The Master and His Emissary*, 116.

¹³⁹ McGilchrist, *The Master and His Emissary*, 116.

oneself are perceived through bodily senses. Bodily experience forms the basis of and raw material for language about the world. Even those words which describe that which is incorporeal still refer to the body to speak of them. Anything that one wishes to express must be expressed in terms of other things and these expressions must ultimately be in terms of bodily experience.

McGilchrist makes the emphatic claim that metaphor is not a mere figure of speech whereby two objects are stated as being like one another after a process of inferential reasoning, but that metaphors precede discursive thought.

A metaphor asserts a common life that is experienced in the body of the one who makes it, and the separation is only present at the linguistic level. Our sense of the commonality of the two ideas, perceptions or entities does not lie in a *post hoc* derivation of something abstracted from each of them, which is found on subsequent comparison to be similar, or even one and the same thing; but rather on a single concrete, kinesthetic experience more fundamental than either, and *from* which they in turn are derived. Thus a clash of arguments and a clash of cymbals are not seen to have something in common only after the disembodied idea of a “clash” is abstracted from the one and from the other, and found - aha! - to be similar; it is rather that these experiences - a clash of arguments and a clash of cymbals, or, for that matter, a clash of swords or a clash of colours - are felt in our embodied selves as sharing a common nature.¹⁴⁰

The assertion of a metaphor, according to McGilchrist, is that there is what he calls a commonality of life in the individual and that the separation of things into discrete experiences and ideas is a purely linguistic action. Metaphor is an assertion of a recognised sense of commonality between ideas or of a kinaesthetic experience of their commonality. This recognition of commonality is not principally the result of a discursive process or of a discovery made about abstractions from experience. The articulation of a metaphor is not the expression of a discovery of similarity but rather an assertion that there exists a more fundamental shared

¹⁴⁰ McGilchrist, *The Master and His Emissary*, 117.

kind of experience in which the objects of the metaphor participate. Put more simply, the objects linked by metaphor are experienced as affecting the body in some similar ways. By this claim, the body is not merely a means to acquiring some information but is necessary for humans to have any understanding of the unity between things at all.

The priority of grasping the implicit to grasping the explicit is suggested by the very metaphors embedded in the notions of abstraction and explication:

Metaphorical meaning is in every sense prior to abstraction and explicitness. The words themselves tell one this: one cannot draw something away (Latin, *abs-* away, *trahere* pull), unless there is something to draw it away from. One cannot unfold something and make it explicit (Latin, *ex-* out, *plicare* fold), unless it is already folded. The roots of explicitness lie in the implicit.¹⁴¹

The words “abstraction” and “explicit”, denoting archetypal acts of creating disembodied utterance, are themselves metaphorical expressions. The very acts of creating universally applicable or symbolic representations of things themselves are linked linguistically to embodied actions. More importantly, the metaphors that comprise the terms “abstract” and “explicit” suggest that some pre-existent in the particular and implicit precedes the articulation of the abstract and explicit.

The priority of metaphor to abstraction and explicitness may be synthesised with Polanyi’s claims about the dependency of explicit knowledge on that which is grasped implicitly. Metaphor provides that link between experience and symbols that allows the development and expansion of symbol systems to form the scaffolding for sophisticated reasoning and thought. Similarly, as Lonergan said that insight is the pivot between the universal and the particular, it may be said that metaphor is the means by which acquaintanceship with a particular thing is linked to a higher perspective about it. Taking

¹⁴¹ McGilchrist, *The Master and His Emissary*, 179.

account of Lonergan and McGilchrist it is plausible that metaphor is of great importance to this pivot between the universal and the particular. Metaphor links language and embodied experience, and language is the principal means by which abstraction is achieved through creating signs that refer not only to the particular things but also to kinds to which the particular belongs and general features shared among different kinds. In other words, metaphor not only allows one to recognise that some particular things one encounters, say, the individuals of a wolf pack, are the same kinds of thing but also allows one to have the kinds of insights that allow linkages between domains such as recognising that the wolf scavenging a carcase and a fungus digesting a fallen log are performing the same kind of fundamental act, consumption of the already dead, or, more dramatically, the similarity between the behaviour of a wolf towards a human and the behaviour of humans towards one another as expressed by the idiom “man is a wolf to man”.¹⁴² In both cases, a similarity is experienced between different particulars and this similarity is expressed in ways that allows something in general to be grasped.

Language serves as an aid to allow new discoveries to be made through formal exploration of the implications of the symbol representations of the world, but the results remain meaningful through their connection to embodied experience through metaphor.

Metaphoric thinking is fundamental to our understanding of the world, because it is the *only* way in which understanding can reach outside the system of the signs of life to life itself. It is what links language to life.¹⁴³

In the absence of metaphor, symbol systems do not bear human-relevant meanings and are merely formal, empty of content.

Metaphor is not merely a means of structuring some concepts in terms of other concepts, and other concepts are not ultimately grounded in physical concepts merely in the sense that all

¹⁴² McGilchrist, *The Master and His Emissary*, 117.

¹⁴³ McGilchrist, *The Master and His Emissary*, 115.

utterances can be said to have a genealogical or structural dependency on articulations of physical experience. Rather, metaphor links experience with utterance.

Utterances make sense in a context of things that are understood, not all of which are articulated and which, to an extent, may not be fully articulatable:

...it has become fashionable to insist on the arbitrary nature of the sign – a fascinating and counterintuitive move, designed to emphasise the “freedom” of language as far as possible from the trammels of the body and of the physical world it describes. There is, however, plenty of evidence that the sounds of words are not arbitrary, but evocative, in a synaesthetic way, of the experience of the things they refer to. As has been repeatedly demonstrated, those with absolutely no knowledge of language can nonetheless correctly guess which word – which of these supposedly arbitrary signs – goes with which object, in what has become known as the “kiki/bouba” effect (“kiki” suggesting a spiky-shaped objects, where “bouba” suggests a softly rounded object). However much language may protest to the contrary, its origins lie in the body as a whole. And the existence of a close relationship between bodily gesture and verbal syntax implies that it is not just concrete nouns, the “thing-words”, but even the most apparently formal and logical elements of language, that originate in the body and emotion.¹⁴⁴

There has existed a tendency to treat language as fundamentally arbitrary in the sense that any symbol can in principle be used to fulfil the same role within the formal structure of language. This has some merit, if language is conceived of in a purely formal fashion with the syntax preserved with arbitrary symbols, as evinced by the multiplicity of possible instantiations of computing machinery. An emphasis on the arbitrariness of symbols with respect to formal structure, however, can obscure the embodied origins and continued function of language as a human creation for carrying meaning. Evidence from psychology has shown that certain vocalisations are evocative of certain kinds of objects, and that this effect is not limited to onomatopoeia. The point McGilchrist is making is that the tendency to emphasise the capacity for abstraction to the exclusion of recognising the implications of language’s origin as created

¹⁴⁴ McGilchrist, *The Master and His Emissary*, 119.

and used by embodied beings can lead to a fundamental misunderstanding of ourselves and how it is that we understand and communicate about the world.

In contrast to the view of language that treats symbols as arbitrary, McGilchrist views even simple expressions as depending upon embodiment. Even expressions as simple as a description of one's emotional state as "down" or "up" are metaphors drawn from experience.

Metaphors, even the simple ones hidden in expressions like feeling "down", derive from our experience of living as embodied creatures in the everyday world. The body is, in other words, also the necessary *context* for all human experience.¹⁴⁵

According to this view, all expressions such as directional metaphors for emotion, only make sense as a result of being a particular kind of embodied being living in the world of human experience. Embodied existence is the context of human experience, which means that human language is meaningful in virtue of being a particular kind of bodily being whose utterances refer to or rely upon experience to convey meaning.

Only through recognising the role of embodiment for language can the proper function of language be grasped, in McGilchrist's view.

Language has done its best to obscure its parentage. It has increasingly abstracted itself from its origins in the body and the experiential world. It developed its current form to enable us to refer to whatever is *not present* in experience: language helped its *re-presentation*. This had the effect of expanding its usefulness to communication and thinking for some purposes, but reducing it for others. In the process, the denotative elements that enable precision of reference and planning, have taken up residence in the left hemisphere, while other aspects of language, broadly its connotative and emotive functions, have remained in the right hemisphere. And the understanding of language at the highest level, once the bits have been put together, the making sense of an utterance in its context, taking into account whatever else is going on, including the tone, irony, sense of humour, use of metaphor, and so on, belongs once again to the right hemisphere.¹⁴⁶

¹⁴⁵ McGilchrist, *The Master and His Emissary*, 118.

¹⁴⁶ McGilchrist, *The Master and His Emissary*, 125.

Language has become increasingly abstract and removed from direct bodily experience, which is to say that many terms less clearly refer to direct bodily experiences in obvious metaphors, and, more importantly, there has been a shift toward valuing and developing increasingly abstract modes of discourse and terminology. Use of language has tended toward representation rather than presentation. The presentation of experience with its implicit content has become of secondary importance to the representation of objects of experience in explicit terms. Both explicit representation and presentation of raw experience with its implicit meaning are valuable modes of communication, according to McGilchrist, and they serve different but valuable purposes. The use of language for presentation evokes experiences. This use of language is the mode meant to communicate the implicit. The use of language for representations is used to make them explicit and manipulable. Both explicit and implicit uses of language are valuable and necessary. McGilchrist sees the left hemisphere as having a valence for explicit representation and the right hemisphere as having a valence for the implicit, but what is of key importance here is not the specialisations of brain anatomy but his observation that the implicit and high-level comprehension in which the right hemisphere specialises. The highest-level grasping of language occurs, according to McGilchrist, by putting the pieces together in a whole, including context and metaphor.

Crucially, McGilchrist's classification of the grasping of metaphor as belonging to the highest level of language comprehension implies that the link to embodied experience through metaphor is a necessary component of higher-level understanding. The use of language as a manipulable, formal system grants practical advantages in certain applications, but sophisticated cognition nevertheless involves bodily experience.

The dichotomy between metaphorical and nonmetaphorical language use can be restated:

Metaphor is the crucial aspect of language whereby it retains its connectedness to the world, and by which the “parts” of the world which language appears to identify retain their connectedness to one another. Literal language, by contrast, is the means whereby the mind loosens its contact with reality and becomes a self-consistent system of tokens.¹⁴⁷

Metaphor grounds utterance to experience, and through experience to the world. Further, metaphor is the means by which the parts of the world that are selected and named by language retain a connection to one another. In effect, McGilchrist’s claim is that metaphor is what allows the objects of thought to be grasped as part of a whole, to be understood from a unifying perspective. Literal language, on his account, attenuates the sense of the connection of these items to one another and to experience. The ability to successfully manipulate symbols becomes paramount.

Metaphor, according to McGilchrist, allows for new discoveries by allowing the linking of new ideas together.

Metaphor is not just a reflection of what has been, however, but the means whereby the truly new, rather than just the novel, may come about. When a metaphor actually lives in the mind it can generate new thoughts or understanding – it is cognitively real and active, not just a dead historical remnant of a once live metaphor, a cliché.¹⁴⁸

A cognitively living metaphor is one which is grasped but which reflects a linking of ideas or a whole that captures implicit and rich meanings. Metaphors, on this account, enable the discovery of that which is truly new rather than that which is merely novel. Mere novelty can simply be achieved through a rearrangement of old ideas or elements into some new order or sequence. The central idea is that the metaphor can be truly creative and can facilitate new discoveries. A living metaphor is a catalyst for insight.

¹⁴⁷ McGilchrist, *The Master and His Emissary*, 125–26.

¹⁴⁸ McGilchrist, *The Master and His Emissary*, 179.

The point is illustrated by the distinction between living metaphors and dead metaphors (clichés).

There is some confusion over what is meant by metaphor. Obviously, there is metaphoric content to almost everything we say – language is essentially metaphoric in nature, at the simplest level. If the chosen “metaphor” for an experiment to distinguish between the hemispheres is so bland, obvious or banal that it requires no imagination, presents no novel thought, and does not bring together disparate meanings, little is likely to be discovered. We know that unfamiliar phrases activate the right hemisphere, while familiar ones activate the left hemisphere... So when it comes to metaphor, it is in keeping with this that poetic phrases, such as “rain clouds are pregnant ghosts” are understood by the right hemisphere, while clichés, such as “babies are angels”, are processed by the left hemisphere... Specifically, it is not just the novelty inherent in metaphor that activates the right hemisphere, but the combination of novelty with the bringing together of disparate ideas that activates the right hemisphere...¹⁴⁹

Again, according to McGilchrist the core of metaphor is the bringing together of disparate ideas, and a cliché, a dead metaphor, is one which has become familiar to the point that it becomes a rote utterance that may express a meaning but is not generative of new ideas. Living metaphors are generative of new ideas and are experienced as illuminating, that is as sources of insight. Novelty may aid metaphors in being living, since novel metaphors still present a discovery that has not yet become familiar and subject to simply being referred to by common language. Metaphors may remain living, however, as long as they continue to evoke an experience of similarity that allows new ideas to be seen.

This should not imply that “dead” metaphors are without value, as much of human language that does not refer directly to embodied experience is composed of such metaphors, such as the metaphor of one’s emotional state as “down” or “up”.¹⁵⁰ McGilchrist’s point, rather, is that when one has an insight, one will be attending not only to isolated symbols or phrases

¹⁴⁹ McGilchrist, *The Master and His Emissary*, 470 n. 135.

¹⁵⁰ The ground of this metaphor in embodiment is explored by Lakoff and Johnson and is discussed below.

repeated by rote but to complex and interconnected parts of the world linked to one another in a newly discovered unity that is grasped through one's embodied experience.

Are McGilchrist's claims about the fundamental importance of metaphor for understanding and insight supported by the philosophical perspective explored in this thesis? To McGilchrist's credit, it is true that metaphor appears to often be important to the act of understanding. Consider again Lonergan's simple example of the gaining insight into the idea of a circle, which requires one to have a mental image of a circle that is nonetheless necessarily distinct from the actual, unimaginable concept of a circle. One can understand that a circle has infinite points an equal distance from a centre point, but one cannot imagine the circle having no thickness. Such an image is, to that extent, inaccurate. The mental image links the abstract concept to an embodied, visual experience. Is the mental image of the circle, then, also a metaphor? Such a mental image does fulfil the role described by McGilchrist of allowing a carrying from an abstraction to experience, and the experience described by Lonergan of arriving at insight into the meaning of the definition of a circle does rely heavily on visual experience.

In Lonergan's example, one begins by inquiring why it is that a wheel is round, not as a practical question of why round wheels are practical and square wheels are not, but as an effort to understand the nature of being round. In his example, one gains the clue that the spokes of the wheel determine its shape, and they need to be of equal length for the wheel to be round. One then realises that the spokes could be sunk unevenly into the wheel and that even were the spokes of equal length and sunk equally into the wheel, there would still be the possibility of the portions of the wheel between spokes being of irregular shape. One proceeds to imagine that there are infinite spokes of infinitesimal thickness and that this properly constrains the shape of the wheel. The key observation to make is that such a process of pursuing an insight

is thoroughly embodied. Even as one proceeds to abstraction, there is a coming into focus of what it means to be a circle and what relation this concept bears to actual round objects. To this extent, the example apparently illustrates that the kind of shared commonality between linked ideas and entities that McGilchrist identifies as being the function of metaphor. Additionally, it provides a simple illustration of the claim that abstraction depends upon embodiment not only as the material from which an abstraction is drawn but also as the means by which the abstraction is understood and expressed.

There is another point of homology between Lonergan's examination of insight and McGilchrist's claim that there is no understanding without metaphor. Lonergan describes a key aspect of insight as producing higher viewpoints that bring what were previously disparate objects of understanding into a newly cognised whole. Similarly, McGilchrist speaks of the need to understand the whole from disparate parts. In McGilchrist's account, metaphor is meant to serve the purpose of facilitating a recognition of connection between things to allow unities to be grasped.

McGilchrist's analysis does suggest some ways that other, seemingly nonphysical acts of understanding could be linked to bodily experience. Consider the example of insight Lonergan gives with the child's discovery of the natural numbers. The child realises that by continuously adding one to the preceding number, a larger number than the last can be produced and that this can in principle be continued indefinitely to produce an unbounded quantity of numbers in which each excepting the first is larger than the one before it and smaller than the one after it. Where is the metaphor in this act of insight? Presumably, the core metaphor would be something akin to the recognition of similarity between incrementing the tally of the natural numbers and adding one additional object to some set, say a pebble added to a mound. As the

mound can be increased in an unbounded fashion, so may the natural numbers be through serial addition.

Still, another item to consider is that, if there can be understanding of the world through simple experience or acquaintanceship, then it might appear that there can exist understanding without metaphor. If one picks up a rock, does one only understand it as an object one can manipulate through a perception of similarity between it and other objects? If this is so, then this would mean that only as they gain more embodied experience do infants become capable of gaining crude understanding of the physical world. Nonetheless, it is the case that any implicit understanding of bodies will still depend on the body of one's experience. There can be no understanding of even a rock without implicit awareness of its relation to one's own body and one's experience of other bodies. One cannot grasp and lift an object without implicit reference to other objects that one has manipulated, and it is this web of associations and awareness of overlapping similarities that McGilchrist believes metaphor to underly. This web of experiences and concepts further forms the context by which utterances make sense. In other words, to grasp anything is to grasp it in light of one's current and past experience, and it is by metaphor that everything is understood through something else.

Utterances make sense in a context and can have nonexplicit meaning. In fact, in the ordinary use of language much of the content communicated is contained in conversational implicature. In many cases, the attempt to dispense with all conversational implicature to state only the explicit sense of the sentence (as is sometimes treated as a laudable target by those pursuing an analytic approach to philosophy and language) can result in a great deal of the meaning being evaporated from the sentence.

The above passages have examined McGilchrist's account of metaphor and the role that it plays in human cognitive achievement. It would be helpful now to look at another influential

account of metaphor and its points of confluence with McGilchrist's. The influential exploration of metaphors by George Lakoff and Mark Johnson suggestively titled *Metaphors We Live By* provides another examination of the fundamental role played by metaphors in human reasoning and language use.

Lakoff and Johnson find that metaphors suffuse language. In the preceding paragraph, for instance, Lakoff and Johnson's work was described as an "exploration". Such a metaphor is an example of what they call a "structuring metaphor", for the act of making an academic investigation is described according to the structure of an exploration of physical territory, in which the field of inquiry is surveyed, observations are recorded, and a path is followed to a destination.

Lakoff and Johnson also claim that meanings of language are rooted in experience, but this does not imply a simplistic sort of one-to-one correspondence between statements and the world.

Since there are *systematic correlates* between our emotions (like happiness) and our sensory-motor experiences (like erect posture), these form the basis of orientational metaphorical concepts (such as HAPPY IS UP). Such metaphors allow us to conceptualize our emotions in more sharply delineated terms and also to relate them to other concepts having to do with general well-being (e.g., HEALTH, LIFE, CONTROL, etc.).¹⁵¹

Lakoff and Johnson cite the common example of the metaphorical association between the direction upward and positive affect. Lakoff and Johnson speak of this as a kind of orientational metaphor, which is based on our lived experience. We know from experience that being upright is associated with such things as health and being in a prone, horizontal position results for such ills as injury and disease. Nor does it imply that one's encounter with the world implies a wholesale pre-propositional knowledge that is simply translated or directly fitted to one's terms.

¹⁵¹ George Lakoff and Mark Johnson, *The Metaphors We Live By* (University of Chicago Press, 2003), 58.

Rather, the tacit knowledge of the world discussed by Polanyi grounds simple meanings but also facilitates the creation of more complex concepts that become attached to these otherwise simple utterances.

More sophisticated examples such as the metaphor of argument as a kind of combat draw on notions such as giving or taking ground or securing as position, which are in turn based on metaphors themselves, namely the description of land as a kind of object that may be acquired, relinquished, or protected. This reliance on a network of related metaphors that use language about embodied experience to articulate concepts, concrete or abstract, based on a recognition of their similarity to other experiences or the concepts built up from metaphors suggests that the ability to recognise similarity between experiences and concepts is fundamental to our use of language and to the formulation of concepts.

Lakoff and Johnson claim that it is likely that

most of our fundamental concepts are organized in terms of one or more spatialization metaphors¹⁵²

which is to say that fundamental concepts are structured according to a similarity recognised between the concepts and the human experience of being bodies in the world. Lakoff and Johnson further argue that

no metaphor can ever be comprehended or even adequately represented independently of its experiential basis.¹⁵³

To use again the metaphoric language of uprightness, the metaphorical link between being upright and health depends on the particular constitution of the human organism and the way it functions in the world. A sea-dwelling organism accustomed to resting horizontally on the ocean floor arguably would not find the linkage between uprightness and health intelligible. To

¹⁵² Lakoff and Johnson, *The Metaphors We Live By*, 17.

¹⁵³ Lakoff and Johnson, *The Metaphors We Live By*, 19.

quote another authority, “If a lion could speak, we would not be able to understand him.” In this oft-quoted passage, Wittgenstein asserts that were a lion to have linguistic abilities, its experience of the world and means of interacting with it are so different than human beings that we could not have a shared context in which symbols could be given a mutually understood meaning.¹⁵⁴

The particularity of human experience provides the grounding for articulate knowledge.

Perhaps the most important thing to stress about grounding is the distinction between an experience and the way we conceptualize it. We are not claiming that physical experience is in any way more basic than other kinds of experience, whether emotional, mental, cultural, or whatever. All of these may be just as basic as physical experiences. Rather, what we are claiming about grounding is that we typically conceptualize the nonphysical *in terms of* the physical—that is, we conceptualize the less clearly delineated in terms of the more clearly delineated.¹⁵⁵

Lakoff and Johnson do not think that all our cognitive states are simply reducible to bodily experiences, but this passage makes clear that they do think that our ability to form a noetic structure depends on our ability to create concepts of various levels of abstraction organised by making use of embodied experiences and concepts about this physical experience. In other words, the scope of human cognition is not confined to bodily experience, but this bodily experience provides necessary material for the inarticulate understanding of structure in nonphysical concepts and provides the basis of the language that allows articulate systems to be built up that depend upon such an inarticulate grounding and expands it.¹⁵⁶

¹⁵⁴ Lakoff and Johnson, 17–19.

¹⁵⁵ Lakoff and Johnson, *The Metaphors We Live By*, 59, emphasis original.

¹⁵⁶ Lakoff and Johnson’s implicit acknowledgment that understanding must cooperate with our experiences and is operative in the very act of having understanding may be seen in the following. “It is by means of conceptualizing that we pick out the ‘important’ aspects of an experience. And by picking out what is ‘important’ in the experience, we can categorize the experience, understand it, and remember it.” Selection of importance is necessary for conceptualization and occurs with the act of making a concept. I differ from these authors in that I contend understanding is not only an act performed after the important aspects of an experience have been recognised as such but also an act that occurs in the very step of identifying what is of importance. Lakoff and Johnson, 83.

These experiences analysed by Lakoff and Johnson as providing the matter for metaphors are notable for one other reason, they are described as gestalts:

What constitutes a “basic domain of experience”? Each such domain is a structured whole within our experience that is conceptualized as what we have called an *experiential gestalt*. Such gestalts are *experientially basic* because they characterize structured wholes within recurrent human experience... Domains of experience are organized as gestalts in terms of such natural dimensions seem to us to be *natural kinds of experience*.¹⁵⁷

Human experience may be parsed into pieces by our actions of selection (though this does not necessary imply that human experience is in fact not itself holistic, just that it may be analysed into parts), but this process terminates not in atoms but in “domains” within which a field of holistic experiences may occur. These experiences are recurrent in human experience, which allows them to provide the shared ground for metaphors that may be understood by multiple persons, allowing a sharing of concepts and the development of the shared view toward these experiences, since shared metaphors facilitate similar conceptualisation. Lakoff and Johnson make clear that these experiences are *natural* kinds in the sense that they arise from our body structure, the interactions of our bodies with the physical world, and the interactions we have with other humans in a shared culture.

The key lesson Lakoff and Johnson draw from the above examples are that we develop our structure of concepts in terms of other concepts based on a recognition of a kind of shared structure.

What may at first appear to be random, isolated metaphorical expressions—for example, *cover those points*, *buttress your argument*, *get to the core*, *dig deeper*, *attack a position*, and *shoot down*—turn out to be not random at all. Rather, they are part of whole metaphorical systems that together serve the complex purpose of characterizing the concept of an argument in all of its aspects, as we conceive them. Though such metaphors do not provide us with a single consistent concrete image, they are nonetheless coherent and do fit together when there are overlapping

¹⁵⁷ Lakoff and Johnson, *The Metaphors We Live By*, 117, emphasis original.

entailments, though not otherwise. The metaphors come out of our clearly delineated and concrete experiences and allow us to construct highly abstract and elaborate concepts, like that of an argument.¹⁵⁸

According to Lakoff and Johnson various metaphors may employ contradictory images that nonetheless together coherently capture aspects of a concept. They overlap not only by being used to expound the same concept but also by both expressing a shared commonality between the image being evoked and the concept being expressed. To buttress an argument and to marshal evidence rely on different images, one architectural and one military, but both not only capture aspects of the same concept; they also cover partially overlapping aspects of the concept. The complex use of different metaphors to illuminate the same concept allows the concept to be adequately expressed. Metaphor allows embodied experience to be put into the service of grasping that which is complex and abstract.

I further observe that the very ability to recognise a kind of shared structure or other similarity between distinct kinds of things already implies that one engages in a concurrent act of understanding with the creation of such metaphors. This is not to imply that metaphors are merely derivative from an already extant understanding. Metaphor allows a recognition of the similarity between two categorically distinct things and the development of metaphor provides a kind of scaffold for the cognition of more sophisticated concepts. Metaphor facilitates understanding.

Lakoff and Johnson's thesis that concepts are largely structured with metaphor bears a similarity to Lonergan's observation that concepts exist with accompanying images, as with the example of the image of a line in which one thing (an image of a straight mark with thickness) must be imagined for a categorically different thing (an infinite succession of points along one

¹⁵⁸ Lakoff and Johnson, *The Metaphors We Live By*, 105, emphasis original.

dimension in Euclidean space) to be conceived. In an extended sense, these representations may be called a kind of metaphor, for they, too, allow concepts to be structured by a recognition of similarity between a given concept and direct or indirect embodied experience.

As a recognition of similarity between things allows a metaphor to carry across from words to experience, so too does the discovery of a metaphor yield new understandings of erstwhile hidden structures. The discovery of a suitable image is part of the production of a new insight, so, too, does finding a suitable metaphor allow one access to novel concepts. When presented, a useful metaphor can assist one to not merely represent but also grasp the structure of a complex concept or the relationship between subordinate concepts. George Lakoff and Mark Johnson illustrate the point by a story of an Iranian student who initially misunderstood the phrase “solution of my problems” to be a metaphor not for finding a strategy for resolving problems analogous to solving a mathematic exercise but instead to be a metaphor for all one’s problems as

A large volume of liquid, bubbling and smoking, containing all of your problems, either dissolved or in the form of precipitates, with catalysts constantly dissolving some problems (for the time being) and precipitating out others. ... the chemical metaphor is both beautiful and insightful. It gives us a view of problems as things that never disappear utterly and cannot be solved once and for all. ... The best you can hope for is to find a catalyst that will make one problem dissolve without making another one precipitate out. And since you do not have complete control over what goes into the solution, you are constantly finding old and new metaphors precipitating out and present problems dissolving, partly because of your efforts and partly despite anything that you can do.

The chemical metaphor gives us a new view of human problems. It is appropriate to the experience of finding that problems that we thought were “solved” turn up again and again. ... We see this as a clear case of the power of metaphor to create a reality rather than simply to give us a way of conceptualizing a pre-existing reality.¹⁵⁹

¹⁵⁹ Lakoff and Johnson, *The Metaphors We Live By*, 143–44.

In Lakoff and Johnson's analysis, the novel metaphor inadvertently created by the student is itself insightful for linking the experience of managing the recurring problems of one's life to the experience of an ongoing process of managing the precipitates in a changing chemical solution. They observe that this alternative metaphor to the common western metaphor of solving one's problems permanently after the fashion of solving a math problem offers not merely a different description of the abstract reality of managing the problems of life but itself creates a new experience of one's problems and the process of addressing them. The discovery of a new metaphor is the means by which a new insight is acquired.

The innovative metaphor that Lakoff and Johnson found interesting serves as an illustration of the role that metaphor plays in arriving at a new insight. They observe that such a metaphor suggests that problems are not ineradicable, that life is a matter of continually managing them, and that as one cannot completely control all the chemicals added to the solution, problems will occasionally precipitate due to outside forces. Lakoff and Johnson, at least, regard this metaphor as offering an understanding of human life that is apposite to the common experience of problems as recurring and in need of continual management in a world in which external factors precipitate or dissolve problems in spite of one's efforts. This metaphor, they think includes the understanding that the continual challenge of one's problems is a matter to be accepted and that the temporary alleviation of problems is a victory, and their eventual resurgence does not mark a failure but a fact of life. This novel metaphor of problems as a chemical solution contrasts with the common metaphor of problems as mathematical problems to which a solution is found. In contrast, the common metaphor, links solving one's problems to the experience of the act of solving a math problem, which need only be performed once, after which the solution stands permanently. The common metaphor of personal problems as math problems, they think, creates the experience of problem-solving that includes the

expectation that solutions to problems are permanent, which implies that when a problem recurs one has failed to solve it. Since many personal problems in life are of a recurring nature, this metaphor of addressing one's problems captures other aspects of human experience than the metaphor of solving mathematical problems.¹⁶⁰

When viewed through the lens of the dichotomy described by McGilchrist between clichéd metaphors and living metaphors, the above example, namely the mathematical solution to one's problems vs the chemical solution of one's problems, stimulates a further insight. The metaphor of finding a solution to one's problems like finding the solution to an arithmetic problem is a cliché that is said almost as a rote utterance. In contrast, the metaphor of dealing with one's problems as a continual process of tending a changing chemical solution to mitigate the precipitation of problems is imaginatively engaging and generative of different perspectives on how to relate to problems, at least presented by Lakoff and Johnson.

In sum, Lakoff and Johnson provide an account complimentary to that of McGilchrist by which cognition is grounded in human embodied experience and provides the basis for complex intellectual activity. Like McGilchrist, they argue that metaphors are built from embodied experiences that are apposite for structuring concepts and assertions about things that are not themselves matters of direct bodily experience but nonetheless can be conceived and expressed in terms of the structures of experiences that are directly or indirectly resultant from embodiment. Further, as McGilchrist thought that metaphors could be creative when they are what he calls living metaphors and merely expressive of old understandings when expressed as clichés, Lakoff and Johnson illustrate the insights that can be gleaned from the perspective of viewing things from a novel metaphor.

¹⁶⁰ Lakoff and Johnson, *The Metaphors We Live By*, 143–44.

There is also some evidence that embodiment and metaphor are of value in the context of mathematical practice. Although mathematics is expressed and sometimes assisted by rule-based symbolic operations, the individual engaged in mathematical reasoning often relies more heavily upon their imagination than upon symbolism.

But conscious mathematics does not always proceed through abstract symbols. It may operate through a “number sense” or a “space sense” or a “kinesthetic sense.” For example, the problem “will this object fit into that box?” is answerable with high reliability on the basis of a mere glance. What lies behind these spatial senses is often not clear. Whether they represent stored experiences, analog-type solutions performed on the spot, inspired by partly random guesses, nonetheless the fact remains that this type of judgment can often be arrived at quickly and correctly.¹⁶¹

This text suggests that even mathematics, which is often considered one of the most abstract fields of human study, does, in fact, make frequently implied references to real or imagined embodied experiences. Examples include the references to a “number sense” or a “space sense” or a “kinesthetic sense,” as well as the reliability by inspection of answering the question, “will this object fit into that box?”

The link between embodiment and mathematical actions can be even more impressively established. As Davis and Hersh recall, a survey was taken of professional mathematicians concerning how they engaged in mathematical reasoning and what role, if any, formal symbolism played in their reasoning.

Practically all of them ... avoid not only the use of mental words, but also ... the mental use of algebraic or precise signs ... they use vague images. And ...the mental pictures of the mathematicians whose answers I have received are most frequently visual, but they may also be of another kind—for example kinetic.¹⁶²

¹⁶¹ Davis and Hersh, *The Mathematical Experience*, 302.

¹⁶² Jacques Hadamard, *The Psychology of Invention in the Mathematical Field* (Princeton University Press, 1945), 84–85. Quoted in Davis and Hersh, *The Mathematical Experience*, 1982, pp. 308.

In other words, professional mathematicians reported that, when engaged in mathematical thinking, they frequently used their imaginations and explored the mathematical objects visually or with imagined motion.

As another example, Albert Einstein in private correspondence recounted that

the words or the language, as they are written or spoken, do not seem to play any role in the mechanism of thought ... The physical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be “voluntarily” reproduced and combined... the above-mentioned elements are, in my case, of visual and some of muscular type. Conventional words or other signs have to be sought for laboriously only in a secondary stage...¹⁶³

In other words, in Einstein’s self-reported case, he approached mathematical reasoning not with linguistic symbols but through images grounded in his bodily existence. In particular, his mental images of mathematical objects depended upon the sensations of sight and tactile experience of his own muscles. These considerations, linking the understanding of abstract ideas with embodied experience in mathematics and physics, underline the role of embodied experience more generally.

In both language and in mathematics understanding is associated with a linkage between concepts of varying degrees of concreteness and of abstraction and embodied experiences that make those concepts and utterances about them meaningful. To understand the meaning of another’s utterance or representation, there is some need to create a similar direct or indirect link with an embodied experience, real or imagined, of one’s own. Polanyi’s observation that all knowledge is personal and the additional point that all knowledge is possessed from within a particular perspective deserves to be mentioned again, for embodied experience occurs in particular embodied beings. A mode of communication that focuses primarily on the

¹⁶³ Hadamard, *The Psychology of Invention in the Mathematical Field*, 142. Quoted in Davis and Hersh, *The Mathematical Experience*, 1982, pp. 308-309.

communication and sharing of personal perspectives and embodiment is that of narrative, to which this thesis now turns.

Narrative, too, has the power to facilitate and convey understanding as illustrated in the following passage from *Mansfield Park*:

Kept back as she [Fanny] was by everybody else, his [Edmund's] single support could not bring her forward; but his attentions were otherwise of the highest importance in assisting the improvement of her mind, and extending its pleasures. He knew her to be clever, to have a quick apprehension as well as good sense, and a fondness for reading, which, properly directed, must be an education in itself. Miss Lee taught her French, and heard her read the daily portion of history; but he recommended the books which charmed her leisure hours, he encouraged her taste, and corrected her judgment: he made reading useful by talking to her of what she read, and heightened its attraction by judicious praise. In return for such services she loved him better than anybody in the world except William: her heart was divided between the two.¹⁶⁴

The interactions of Edmund and Fanny allow him to educate her. This education consists in Edmund directing her attention towards narratives that inform her judgment and shape her tastes. Narrative allows vicarious experience and the tacit understandings it grants. Edmund directs Fanny's attention to imaginative experiences with the purpose of changing her tastes and judgments. That Edmund determines that the stories will have such salubrious effects implies that he has already read the stories himself, adopted these responses himself as part of a stance toward them, and wishes to bring Fanny to both an awareness of these texts and a similar stance. In other words, Edmund's education of Fanny is a matter of bringing her to share his stance toward these linguistic and imaginative objects.

This passage suggests that one significant way understanding may be shared is through a kind of pointing, literal or metaphorical, that not only calls something to attention but invites

¹⁶⁴ Jane Austen, *Mansfield Park*, ed. James Kinsley, Oxford World's Classics (Oxford; New York: Oxford University Press, 2003), 18.

one person to synchronise with another. These stories have the effect of forming Fanny's imagination, which informs how she subsequently makes sense of the actions of the subsequent characters. Appreciating these stories also brings her into deeper cognitive and emotional alignment with Edmund. He intends to shape her affective responses to things, and successfully does so, meaning that the two of them have a shared stance toward both the content of the narratives, and subsequently the novel's characters and their comportment (with the notable exception of Mary Crawford, whose character Edmund misrepresents until her recommendation that Maria Rushworth and Henry Crawford persist in adultery forces him to re-evaluate her character, allowing him to share the same stance toward Mary that Fanny does).

Although the *Stanford Encyclopedia of Philosophy* does have an article on fiction, it does not have an entry on narrative, which is broader than just the category of fiction, so while some philosophical examinations of fiction are valuable to examine in this thesis, the broader subject of narrative will still need to be explored.¹⁶⁵

The formative effect of narrative on understanding is discussed by Martha Nussbaum,

...the novel is a morally controversial form, expressing in its very shape and style, in its modes and interaction with its readers, a normative sense of life. It tells its readers to notice this and not this, to be active in these and not those ways. It leads them into certain postures of the mind and heart and not others.¹⁶⁶

Nussbaum's observation that the novel is morally controversial is due not to the nature of the novel as such, but to the debate about the beneficial or detrimental effects of engaging with literature for ethical formation. Although her analysis focuses on Charles Dickens' novel "Hard Times" and the competition there portrayed between utilitarian economic conceptions and

¹⁶⁵ For an overview of philosophical work on narrative taken as a reaction against rationalism in modernity, see Gregory L. Jones, "Rhetoric, Narrative, and the Rhetoric of Narratives: Exploring the Turns to Narrative in Recent Thought and Discourses," *Issues in Integrative Studies* 11, no. 1 (1993): 7–25.

¹⁶⁶ Martha C. Nussbaum, *Poetic Justice: The Literary Imagination and Public Life*, with Internet Archive (Beacon Press, 1995), <http://archive.org/details/poeticjusticelit0000nuss>.

representations of human life and the conceptions and representations of human life found in novels, Nussbaum links what she calls the morally controversial quality of novels to the misgivings expressed by philosophers for millennia about the formation of character through emotion and example directed toward particular and flawed human beings rather than through explicit and formal reasoning about human good.

Nonetheless, Nussbaum indicates that reading literature conveys to the reader a sense of how life ought to be lived and shapes how they ought to perceive and attend to the world. Further, narrative has a formative effect on the stances of intellect and affect. The narrative presents things to the attention of the reader, selecting them in favour of other things, and invites them to attend to them in certain ways. In effect, the narrative leads the reader to take a certain stance toward what is represented. The narrative engages the reader's imagination, and emotions, which is to say that the narrative stimulates the reader to virtually experience the sensations and feelings of the bodies represented in the story.

Unlike most historical works, literary works typically invite their readers to put themselves in the place of people of many different kinds and to take on their experiences. In their very mode of address to their imagined reader, they convey the sense that there are links of possibility, at least on a very general level, between the characters and the reader. the reader's emotions and imagination are highly active as a result...¹⁶⁷

In distinguishing narrative from historical works, Nussbaum draws attention to the difference between the experience of reading a description of actions that can, at its most basic, be a mere elaborated list of events and the experience of reading a work that invites the reader to understand the perspectives and experiences of particular characters. Moreover, Nussbaum indicates that the imaginative engagement with the possibilities expressed in fiction allows for

¹⁶⁷ Nussbaum, *Poetic Justice*, 5.

a kind of gaining of perspective on possibilities of action that informs one's practical grasp of how to act.

Nussbaum examines the novel by Charles Dickens *Hard Times* to explain how a novel can achieve formation of the intellect and character without explicit, abstract content.

This novel tells a story. In so doing, it gets readers involved with the characters, caring about their projects, their hopes and fears, participating in their attempts to unravel the mysteries and perplexities of their lives. The participation of the reader is made explicit at many points in the narration. And it is brought home to readers that the story is in certain ways their own story, showing possibilities for human life and choice that are in certain respects their own to seize, though their concrete circumstances may differ greatly.¹⁶⁸

This novel, like novels generally and narrative still more generally, invites the reader to share the characters' perspectives and experiences through imagination and emotion, gaining a sense of implicit aspects of their lives, which allows one to gain an understanding of the world from the perspective of another without having lived that life directly.¹⁶⁹

Aristotle's writings on the structure and nature of poetry are clarifying concerning the nature of fiction, as Martha Nussbaum notes. Fiction being an important species of narrative and containing the general structure of a story, it is useful to reference that structure here. In the first instance, narrative involves some sort of description of a complete action involving characters. There must be some sort of perceived unity to the action depicted. Merely being

¹⁶⁸ Nussbaum, *Poetic Justice*, 31.

¹⁶⁹ A caveat needs to be made for the cases in which a narrative is told without the existence of intelligent entities. In these instances, the narratives encourage a sharing of perspective with any characters through anthropomorphism. Even a simple story about wolves hunting buffalo will invite the reader to have some sort of imaginative or affective engagement with the wolves or the buffalo and creatures engaged in a hunt or fleeing danger, which humans understand by analogy with their own experience as animals. It is possible for even plants or inanimate objects to feature as the characters, as it were, of a story, but they must still be presented as undergoing some sort of action analogous to that of a human character, which would allow the growth and death of a tree or the rise and weathering of a mountain, say, to be described as something akin to the arc of a human's mortal life. It is not surprising that this can be achieved, as studies on human intentionality have shown that participants shown two abstract objects, a circle, and a triangle, say, in which one moves randomly and the larger follows it in terms of a deliberate chase. In the absence this ability to make sense events in terms of actions in which characters are involved, the text would not constitute a story, but a bare description.

about a single character is insufficient, and, Aristotle says, unnecessary for the work of art to contain an internal unity.¹⁷⁰

Narrative describes an action that occurs in time. As Aristotle observes of epics and drama, they have beginnings, middles, and ends, but also possess a unity of themselves. They are not a mere sequence of disconnected events, and they achieve a completion, which is to say that they have a *telos* toward which the events tend.

McGilchrist states that one of the key characteristics of narrative is that it is experienced as a continuous flow. This is distinct from an abstract account of an action through time as a series of discrete events in a sequence, which lack the apprehension of unity that characterises experience of narrative. Again, the importance of understanding as a singular and unifying cognitive act may be seen here. The movement through narrative is holistic.¹⁷¹

¹⁷⁰ “The unity of a plot does not consist, as some suppose, in its having one man as its subject. An infinity of things befall that one man, some of which it is impossible to reduce to unity; and in like manner there are many actions of one man which cannot be made to form one action... In writing an *Odyssey*, [Homer] did not make the poem cover all that ever befell his hero—it befell him, for instance, to get wounded on Parnassus and also to feign madness at the time of the call to arms, but the two incidents had no necessary or probable connexion with one another—instead of doing that, he took as the subject of the *Odyssey*, as also of the *Iliad*, an action with a unity of the kind we are describing. The truth is that, just as in the other imitative arts one imitation is always of one thing, so in poetry the story, as an imitation of action, must represent one action, a complete whole, with its several incidents so closely connected that the transposition or withdrawal of any one of them will disjoin and dislocate the whole. For that which makes no perceptible difference by its presence or absence is no real part of the whole.” The plot of a story, according to Aristotle, is principally an action involving a character that constitutes a whole. There is an element of selection involved on the part of the author to determine which events together make up the parts of a whole, and in this artistic act of seeing the whole, the faculty of understanding is employed. As in an act of insight a higher viewpoint is gained that unifies disparate phenomena, so in the authoring of a story is a unity grasped out of all the parts of what befalls the characters. This unity is, in turn, conveyed to the reader. Aristotle, “Poetics,” in *The Complete Works of Aristotle*, Revised edition, ed. Jonathan Barnes (Princeton University Press, 1984), 2:1451a15-35.

¹⁷¹ “Music – like narrative, like the experience of our lives as we live them – unfolds in time... Time is the context that gives meaning to everything in this world, and conversely everything that has meaning for us in this world, everything that has a place in our lives, exists in time. This is not true of abstractions and re-presentations of entities, but all that *is* is subject to time... What is called temporal sequencing is an ambiguous concept. Such sequencing, depending on what one means by that, may be right-hemisphere-dependent or, at least where the sequence has no ‘real world’ meaning, as it would in a narrative, left-hemisphere-dependent– the understanding of narrative is a right hemisphere skill; the left hemisphere cannot follow a narrative. But sequencing, in the sense of the ordering of artificially decontextualised, unrelated, momentary events, or momentary interruptions of temporal flow – the kind of thing that is as well or better performed by the left hemisphere – is not in fact a measure of the sense of time at all.” McGilchrist, *The Master and His Emissary*, 75–76, emphasis original.

The ability to engage with narrative depends upon imagination, which is to say that it requires the human ability to create phantasms that create a kind of virtual experience of what is being presented in the story. This depends upon embodied experience to create a kind of virtual experience.

Narrative is not restricted to training characters' decisions and habits of life through practical examples in the straightforward sense of a morality play in which human vices and virtues are portrayed directly with clear lessons meant to be taken. It is possible for narrative to feature characters whose understandings and decisions are in some way flawed, and the reader is left to make these assessments on their own. By way of example, Gustav Flaubert was reportedly criticised by contemporaries for omitting from *Madame Bovary* an explicit condemnation of the titular character by the narrator. In contrast, Flaubert thought that the reader ought to be able to make that assessment for themselves.¹⁷² The imaginative engagement with the story and the characters that narratives can facilitate need not depend on an explicit indication from the narrator as to what the reader's evaluation of characters ought to be.

Narrative allows the conveyance of experience through language not primarily in the form of an explicit communication of ideas but through creating the environment for sharing in the experiences of characters according to the selection of an author. To this extent, narrative depends upon a shared perspective being developed between the author and the reader. Much of what may be communicated is itself implicit, which is in part the explanation for the experience of an author conveying more through their work than they had fully apprehended in the process of writing. Still, this shared vision with the author toward the story expressed, does not imply that the reader's independent reasoning is not involved. The reader can form

¹⁷² George Sand et al., *The George Sand-Gustave Flaubert Letters*, with Roberts - University of Toronto (London, Duckworth, 1922), 353–55, <http://archive.org/details/georgesandgustav00sand>.

evaluations about the dispositions and actions of the characters within the story, and so unite the acquisition of understanding through a kind of virtual experience, as it were. The opportunity that narrative thus allows for sharing understanding relies both upon the shared experiences of embodiment between the author and the reader and upon the capacity to use language to evoke imaginative experiences of the cognitive, emotional, and physical experiences of other persons, communicating that which has been grasped without explicit utterance.¹⁷³

This section has argued for the fundamental importance of embodiment for cognition. Metaphor serves to link utterances and abstractions with embodied experience, which allows sophisticated concepts to be understood. Further, narrative provides a privileged means for gaining and communicating new understandings through its invitation to imaginatively engage with embodied scenes and the experiences of the characters depicted. This capacity to share understanding through shared experience between persons is examined in the next section.

3.3 Second-Person Aspects of Understanding

3.3.1 Apprenticeship and Sharing Understanding

Having explored the significance of first-person aspects of understanding, it is appropriate to explore what is meant by phrases such as “you understand”. This phrase is more than a simple assertion that there is something besides oneself that can say “I understand”, and entails at minimum a kind of addressing another. Moreover, the implications and conditions of such an addressing need to be explored.

¹⁷³ For an analysis of the importance of narrative for cognition and interpersonal communication in daily life, see Elinor Ochs and Lisa Capps, *Living Narrative: Creating Lives in Everyday Storytelling* (Cambridge, MA: Harvard University Press, 2022), <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://www.degruyter.com/isbn/9780674041592>.

To this end, I will first examine Michael Polanyi's discussion of what he calls "personal knowledge" and the means by which it is communicated. I will then comment briefly on Lonergan's discussion of pedagogy before examining the psychological literature on "joint attention" and what can be deemed the second-person perspective or second-person relatedness as offering privileged means of sharing in understanding with other persons.

Polanyi's theory of "personal knowledge" is most completely expounded in his eponymous book.¹⁷⁴ Polanyi describes knowledge as being personal in multiple senses, but for my purposes it is most relevant to discuss the way in which the understanding implied in having such knowledge can also be said to be personal. Polanyi denies that all that one knows can be fully articulated as a set of discrete propositions. There is always what he calls "tacit knowledge", which is known but not articulable (Polanyi stops short of denying that such knowledge can in principle be articulated. He makes the weaker claim that for any knowledge that is articulated there will remain unarticulated knowledge without which the articulated knowledge cannot be meaningful).

Polanyi finds that such knowledge is most clearly illustrated in skills and what he calls *connoisseurship*. As an example of skilful action, he analyses the skill of riding a bike, in which one maintains one's balance by continually adding a slight curvature to one's path that compensates for one beginning to fall. One need not be able to provide an explanation of the physics behind riding a bike to be able to exercise the skill. Nor, in the case of *connoisseurship*, is one always able to articulate how one discriminates between different perceived objects, yet one is capable of doing so reliably. Yet, it remains possible for such tacit knowledge to be communicated between persons as in the relationship between a master and an apprentice. The

¹⁷⁴ Polanyi, *Personal Knowledge*.

latter attends to the former and learns to perform themselves the appraisals and actions of their master. Such skilful action requires learning what to attend to in a given percept, in other words, to understand what is relevant to task for intelligent action at hand and, if need be, how to act in response to these attributes. When such personal relations become disrupted, it is possible for such knowledge to be lost, as Polanyi illustrates by the failure of modern artisans to recreate a violin of the calibre of a Stradivarius.¹⁷⁵ As a consequence, such tacit knowledge may indeed be transferred between persons, but it always belongs to persons. In other words, understanding belongs to beings with a first-person perspective.

3.3.2 Sharing Understanding and Catalysing Insight: Support from Data Science

Polanyi's description of master-apprentice relationships invites the question of whether or not there is a social dimension to understanding beyond that of transmission through shared attention or activity. There exists anecdotal support for the importance of the social context to insightfulness. Alex Pentland recounts that the most creative thinkers and artists he encounters in his work at the MIT Media Lab are those who are the most socially active in seeking and spreading their ideas.

The most consistently creative and insightful people are explorers. They spend an enormous amount of time seeking out new people and different ideas, without necessarily trying very hard to find the “best” people or “best” ideas. Instead, they seek out people with *different* ideas.¹⁷⁶

¹⁷⁵ “It follows that an art that has fallen into disuse for the period of a generation is altogether lost. There are hundreds of examples of this to which the process of mechanization is continuously adding new ones. These losses are usually irretrievable. It is pathetic to watch the endless efforts—equipped with microscopy and chemistry, with mathematics and electronics—to reproduce a single violin of the kind the half-literate Stradivarius turned out as a matter of routine more than 200 years ago.” Polanyi, *Personal Knowledge*, 53.

¹⁷⁶ Alex Pentland, *Social Physics: How Social Networks Can Make Us Smarter*, Reissue edition (Penguin Books, 2015), 26.

Pentland's anecdote suggests that social engagement with others facilitates understanding. Pentland implies that this understanding is not merely an aggregation of data from the persons one encounters by his own reference to an anecdote of Steve Jobs:

Creativity is just connecting things. When you ask creative people how they did something, they feel a little guilty, because they didn't really do it, they just saw something. It seemed obvious to them after a while. That's because they were able to connect experiences they've had and synthesize new things.¹⁷⁷

According to this passage, creativity, which Pentland links with insight, is a matter of establishing connections between things. To establish meaningful connections is distinct from merely collecting data and implies an ability to discriminate between what is relevant and irrelevant and to recognise some unity. The importance of sight as a metaphor for insight is here again affirmed. Pentland's approval of Steve Jobs' reference to insight as involving a synthesis of experience and Pentland's discussion of insightful thinkers as explorers seeking the ideas of others suggests that there is an extent to which having experiences of others or sharing an experience with others, including the experience of attending to an idea, contributes to insight. Pentland's analysis can be understood as an attempt to explain the observed positive effect on the frequency of insights that occurs when groups collaborate with one another, especially in contexts of face-to-face interaction. The primary thesis of Pentland's book is that, in a properly functioning network, idea flow occurs between individuals and facilitates individuals to acquire the best ideas for subsequent synthesis.

Idea flow is an important factor in facilitating insight, but there is a still deeper way in which being in relation to individuals facilitates insights. Namely, it allows for one to enter into

¹⁷⁷ George Beahm, ed., *I, Steve: Steve Jobs In His Own Words*, Later Printing edition (Agate B2, 2011) quoted in Alex Pentland, *Social Physics: How Social Networks Can Make Us Smarter*, Reissue edition (New York, NY: Penguin Books, 2015), 26.

the right kind of affective states and to develop appropriate habits of thought with respect to the object of shared potential understanding.

This distinctive quality of intellectual passions is largely due to the fact that they are attached to an articulate framework... Young men and women brought up in this culture accept it by pouring their minds into its fabric, and so live the emotions which it teaches them to feel. They transmit these emotions in their turn to succeeding generations, on whose responding fervour the edifice relies for its continued existence.¹⁷⁸

On this account, the passions which sustain intellectual life are part of culture and developed in individuals through interactions with the preceding generation. The enterprise of pursuing new knowledge and delighting in what is discovered and accepted as known depends on a shared awareness of and response to the objects of inquiry. In other words, it requires what psychologists call triadic awareness, which will be discussed in the next section.

On Polanyi's account, being in a trusting relation with others is necessary for the pupil to accept as true propositions that they strive to understand. As Lonergan observes, it is by trust in the assurance of the teacher that there is something to be understood that the pupil labours for an understanding that they do not yet have and may not initially anticipate. Further, the teacher aids the student in developing the sorts of passions that aid them in their efforts to acquire new knowledge both through a general desire to gain more understanding of a particular subject by particular means and, more subtly, to hone their ability to feel an affective response to approaches to the question likely to yield understanding. This analysis of the three-part awareness between pedagogue, pupil, and the object of understanding indicates that the shared awareness facilitates more than just a transmission of facts or propositions but moulds the habits and affections of the pupil's mind and that this is necessary both to appreciation of what has been discovered and to the continuation of making discoveries. Polanyi's view of the role of

¹⁷⁸ Polanyi, *Personal Knowledge*, 173–74.

personal relations with the teacher and their contribution to developing intellectual passions gains support from studies into joint attention in infants.¹⁷⁹

3.3.3 Empirical Support from Studies of Joint Attention in Infant Cognitive Development

Contemporary psychology has drawn attention to the enthusiasm and manner in which children typically interact with parents and caregivers. An everyday example is a child pointing to something in a supermarket, which is, in effect, an invitation to the parent to share attention to the object with the child. Children, similarly, will follow the gaze or the pointing fingers of their parents or caregivers to attend to object that they have picked out. Children demonstrate both the tendency to invite adults to share attention with them and to accept the adults' invitations to share in their attention.

When these tendencies to initiate or accept invitations to shared attention are significantly reduced, as in the case of severe autism spectrum disorder (ASD), cognitive development can be impaired. This suggests that shared attention is of importance to human cognition and is important for the acquisition of insights that comprise ordinary child development.¹⁸⁰

Indeed, psychological research suggests that such a privileged relation between persons for the sharing of attention does exist and that it bears special importance for the acquisition of insights. This relation of shared attention between persons is referred to in the literature as joint attention. One authoritative account of triadic joint attention is given by Peter Hobson as a

¹⁷⁹ Studies of the cognitive impairments shown by autistic children indicate that a capacity for second-person relatedness is important for typical cognitive development. Naomi Eilan et al., eds., *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology*, 1st edition (Oxford: Oxford University Press, 2005).

¹⁸⁰ Peter Hobson, *The Cradle of Thought: Exploring the Origins of Thinking*, Reprints edition (Pan, 2004), 124–27.

shared awareness of shared focus between persons, on a single, shared object of attention.¹⁸¹ Crucially, joint attention does not refer merely to cases in which two persons merely happen to be aware of the same object, as when two persons happen to look up at the moon simultaneously without any awareness that the other is doing so as well. In cases of joint attention, both persons are aware of a shared focus with the other person on a single object of attention.

There does exist a heterogeneity of viewpoints concerning how joint attention should be defined.¹⁸² Adjudicating between the different psychological and neurological accounts of the phenomenon is not necessary for the purposes of this thesis; it is sufficient to note that there is a reasonable consensus among researchers of joint attention that the human capacity to enter into shared perspectives with other persons is of fundamental importance to the development of human cognition and that this capacity for entering into mutual understanding remains an important component of cognition. To summarise, Hobson's words can serve as a working definition, namely that when persons have joint attention with one another, they will

... share awareness of the sharing of the focus, something that often entails sharing an attitude towards the thing or event in question.¹⁸³

In other words, when joint attention occurs, the participating persons have their attention focused on the same object, are aware of this shared awareness, and often adopt similar stances toward this object of shared focus.

Peter Hobson identifies a further aspect of joint attention that goes beyond an awareness of the other person simply as a co-attender:

¹⁸¹ Peter Hobson, "What Puts Jointness Into Joint Attention?" in *Joint Attention: Communication and Other Minds*, ed. Naomi Eilan et al., Issues in Philosophy and Psychology 3 (Oxford, United Kingdom: Oxford University Press, 2005), 185.

¹⁸² Axel Seemann, "What Is Joint Attention? Comparative, Developmental, and Definitional Issues," in *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, ed. Axel Seemann (MIT Press, 2012).

¹⁸³ Hobson, "What Puts Jointness Into Joint Attention?" 185.

The joining goes deeper than this, however—it is a linkage that exists in what can only be called *identification*. Identification is not reducible to imitation, any more than a subjective orientation can be reduced to intention or action. It involves appropriating something of *the psychological orientation* as well as the actions of the other person.¹⁸⁴

In other words, Hobson's contention is that when persons enter into joint attention, not only are they recognised as other subjects attending a common object, but they also enter into a kind of synchronisation with one another. They do not merely share information about the object or perceive the emotional state of the other. Rather, there can be an intense triadic harmonisation in which the psychological stance of the other person (including that person's emotions and possibly physical posture) is appropriated by oneself.

An early example in a person's life is the phenomenon of gaze-following by infants. Infants will look towards objects when they perceive adults to shift their attention to some object. This represents a simple case of the child attempting to sync its physical state with that of the adults it perceives. In older children, the infant responds to the emotions it perceives in the adult's face. A child that perceives the adult to be smiling will reciprocate. One that perceives anger in an adult's face will, in contrast, respond with an expression of fear.

Tomasello et al. even go so far as to suggest that the capacity to share the intentional stance of another person is a distinctive difference between humans and nonhuman apes that allows the expansion of human cognition far beyond the limited scope of other apes. It will be recalled that Nick Bostrom left the notion of qualitative superiority of intelligence vague and the nature of the superiority of human intellects to those of chimpanzees deliberately prospective. Whether or not one agrees with Tomasello et al.'s thesis that the capacity for second-person relatedness is the key distinction of human intelligence, they nonetheless marshal

¹⁸⁴ Hobson, "What Puts Jointness Into Joint Attention?," 201.

convincing evidence that such capacities do strongly contribute to the development of distinctive human cognitive abilities.

But there is another dimension to these interactions that goes beyond simple timing and contingency. Human infants and adults interact with one another dyadically in what are called protoconversations. These are social interactions in which the adult and infant look, touch, smile, and vocalize toward each other in turn-taking sequences. But as most observers of infants have noted, the glue that holds protoconversations together is not just contingency but the exchange of emotions (Hobson 2002; Trevarthen 1979). Evidence for this comes from Stern (1985), who found that during protoconversations adult and infant do not just mimic each other or respond randomly, but often express the same emotion using a different behavior (e.g., the adult expresses happiness facially and the child vocally). During protoconversations, infants gaze into the eyes of the partner face-to-face in what is called mutual gazing. It is a dyadic activity in the sense that the infant is not monitoring the adult's looking at her or any other object; it is direct engagement.¹⁸⁵

The existence of protoconversations supports the notion that human social interaction is a prelinguistic and embodied phenomenon that allows for sharing attention, emotions, and stances. Put more simply, the ability to understand and be understood by others and with others occurs early in development.

It being granted that autistic children and nonhuman apes alike do demonstrate an ability to make some sense of the intentions and goals of others, developmentally typical human children engage in unique kinds of dyadic and triadic awareness and interaction with adult humans. The case of children with severe ASD is particularly instructive because they engage in far fewer of these types of interactions and often suffer delays or impairments in cognitive development, plausibly as a result. Participation in what Tomasello et al. call protoconversation is a phenomenon in which prelinguistic children adopt the same emotional state as the adult with whom they are interacting, and thus a capacity to adopt the emotional state of an adult

¹⁸⁵ Michael Tomasello et al., "Understanding and Sharing Intentions: The Origins of Cultural Cognition," *The Behavioral and Brain Sciences* (New York, USA) 28, no. 5 (2005): 681, <https://doi.org/10.1017/S0140525X05000129>.

appears significant given that the reductions in exercising this capacity consequent on ASD seem to be detrimental.

The protoconversations Tomasello et al. describe are notable not only because typical human persons engage in them but because they evince the ability of young children to understand the emotional state of an adult rather than merely engaging in mimicry, since the child and adult can share these states even if they manifest them differently.

Tomasello et al. note that no other ape engages in protoconversations and conclude that these interactions evince the humans' unique drive to engage with other members of their species and understand their psychological state.

Although there may be differences in the way protoconversations take place in different cultures – especially in the nature and amount of face-to-face visual engagement – in one form or another they seem to be a universal feature of adult-infant interaction in the human species (Keller et al. 1988; Trevarthen 1993). Protoconversations require not only that the two interactants understand each other as animate agents, but also that they have a special motivation and capacity to share emotions with each other. This additional factor is clearly necessary, as the individuals of many nonhuman species appreciate others as animate agents, but are still unmotivated to engage with them in protoconversations (see sect. 4.1.2 on great apes). But sharing emotions in early infancy is just the beginning of a much longer developmental process. Important though they may be as a foundation, protoconversations do not involve joint commitments to any shared goals or action plans.¹⁸⁶

Protoconversations themselves do not contribute directly to cognitive development, but the strong desire of infants to align their emotional states to adults, in Tomasello's view, facilitates the acquisition of language and cultural knowledge that results in the dramatic expansion of cognitive abilities and interests that comprise adult action.¹⁸⁷

¹⁸⁶ Tomasello et al., "Understanding and Sharing Intentions," 682.

¹⁸⁷ "Thus, while ape infants interact with their mothers dyadically and are responsive to them behaviorally (Maestripieri & Call 1994) and they may even show some maternal gazing and social smiling (Mizuno & Takashita 2002; Tomonaga et al. 2004), there are no observations of anything like protoconversations between adults and infants. Personal observations of the authors suggest that although all primates display similar social emotions in terms of attachment between babies and mothers, human infants and mothers possess a much larger behavioral

It should also be noted that the dyadic and triadic interactions in which adults and children engage are fundamentally embodied. Most obviously, to gaze or point at an object requires moving parts of one's body so that they are oriented toward said object. More subtly, the facial expressions and gestures that are constitutive of the early interactions between adults and children are embodied and prelinguistic means of communication.

This kind of understanding leads to some powerful forms of cultural learning, especially imitative learning in which the observer must perform a means-ends analysis of the actor's behaviour and say in effect "When I have the same goal, I can use the same means (action plan)." This analysis is also necessary before one can ask why someone did something and whether that reason also applies in my circumstance ("rational imitation"). Without such an analysis, only simpler forms of social learning are possible (Tomasello et al. 1993 and see sect. 4.1.1). The main point is that 1-year-old infants use their newly emerging skills of intention understanding not only to predict what others will do, but also to learn from them how to do things conventionally in their culture.¹⁸⁸

Tomasello et al. observe that the ability to understand and share in the intentions and stance of adults allows children to engage in a sharing of cognitive content that allows them to gain not only informational knowledge or a kind of objective knowledge of other persons but the skills and relevant perspectives that other persons possess. Such psychological conclusions lend support to Polanyi's notion that master-apprentice learning is pervasive in human life and the means by which intellectual pursuits such as scientific practice, the art of research, and the aesthetic sense of intellectually satisfying theories are communicated between persons and generations. Tomasello et al. observe this fundamental means of sharing understanding in children and see it as the key that enables access to the entire range of human cognitive action.

repertoire for expressing a much wider range of emotions in their social interactions than do other apes (e.g., laughing, crying, cooing, smiling) – especially expressions of positive emotions serving to enrich the dyadic emotional engagement between mother and child." Tomasello et al., "Understanding and Sharing Intentions," 685.
¹⁸⁸ Tomasello et al., "Understanding and Sharing Intentions," 680.

The psychological research into the importance of joint attention compliments the account given by Michael Polanyi of the differentiating quality of human cognition from that of other animals. Polanyi ascribed the extension of human cognitive abilities to the possession of language, which allows the animal powers of tacit understanding to extend to abstract and formal representations and operations, as well as facilitating formulation and communication of concepts. Tomasello et al. suggest that acquisition of language belongs to the general class of cultural knowledge and skills transferred between adults and children through the mechanism of jointly attending, emotionally aligning, and mutually participating.

More complex cases exist in which the child does not appear merely to understand the direction of the adult's attention or their emotional state but appears to sync herself with the adult more dramatically. Hobson et al. performed a simple experiment in which they took a stuffed frog toy, wiped it across their foreheads, and told children to duplicate the action. Normal participants gathered the toy in their hand and wiped it across their foreheads in imitation of the experimenter. Such replication of self-referential actions bears a similarity to the mastery of I-thou language. When one has grasped how to use the first and second-person pronouns, one understands that when one is asked a question using the second-person for the form "How are you?" that one is meant to adopt the usage of the first-person I in making one's reply, and vocatively addressing the other person requires one in turn to use the word "you." This simple dynamic is suggestive, for the addressee of the question must understand that they must place themselves in the perspective of the other person to grasp that "I" and "Thou" function as pointers toward oneself and toward another respectively, and that one must, in replying, take the stance previously occupied by one's interlocutor, using the first-person to point to oneself and the second-person to point to the other person. The experience of this

relation of “I” and “thou” in the context of joint attention may be termed second-person relatedness.

The effects of autism spectrum disorder (ASD) on the above-mentioned phenomena are striking. Autistic children do not engage in gaze following, though they will look to an object when prompted by some other stimulus (say, a model train turned toward an object by a motor). This indicates that the failure to attend the objects of another’s attention is not due to an inability to focus when prompted, but some other deficiency. Nor do children with ASD perform the same on tests of ability to mimic self-referential actions. When told to replicate the researcher’s use of the stuffed frog to wipe his own face, the autistic child will not wipe his own forehead with the frog but will instead wipe the forehead of the researcher. Similarly, the child is delayed in her mastery of the use of I-thou language. When asked a question of the form “How are you?” many children with ASD will respond, “You is fine.” They will similarly refer to the researcher as “I”. In other words, they appear to use first and second-person modes of address in a third-person manner.

In addition, severely autistic children can show delayed language acquisition or even fail to acquire it at all. It is also a common characteristic of ASD that persons struggle to discern wholes or selectively attend to some details. They tend to be unable to see the wood for the trees, as it were, manifesting a comparative deficiency of one of the standard operations of the faculty of understanding. In other words, as well as inhibiting second-person relatedness, ASD can also impair the faculty of understanding, with the former as a possible cause of the latter. In other words, the failure to engage in typical self- and other-referential behaviours, in the context of joint attention, may be a contributory factor to a comparative inhibition of the faculty of understanding, leading to difficulties in, for example, distinguishing between what is of primary and secondary importance, or discerning wholes comprised of several individual parts.

Additionally, autistic children can often engage in highly repetitive behaviours. They arguably find it difficult to explore ways of interacting with the world in novel ways. In a sense, they are disinclined to venture from the world as they conceive and dwell within it, nor are they easily drawn from this limited experience and conception of the world to share in that of others.

The importance of joint attention may be illustrated negatively by an example of the consequences of failing to develop understanding as shown by the account by Clara Claiborne Park of raising a strongly autistic child. She discusses the apparent indifference of her daughter to interactions with her and other members of her family, failing to enter into interactions based on joint attention, never so much as pointing. As a consequence, her daughter did not develop shared experiences or skills necessary to enter into the culture of her family. She experienced severe delays in learning language, and she never achieved the full cognitive faculties of an adult.¹⁸⁹ It is of particular note that late in her daughter's childhood, Park was able to begin to bring her daughter into shared activities and interactions. Having missed multiple developmental milestones, she remained severely impaired over the course of her life, but she still proved able to acquire a crude grasp of language and an ability to be taught some basic skills and concepts such as to still place her intellectual capacities well beyond those of an ape (though she remained unable or unwilling to learn certain skills or practices such as toilet use).

Whatever additional cognitive impairments might result from a case of severe autism, the

¹⁸⁹ The daughter did nonetheless display instances of notable understanding. For instance, when she fell into ill health and spontaneously indicated to her mother that she wished to eat eggs by manipulating her mother's arms to stimulate her to prepare a meal with the desired ingredients. It should be noted that prior to this the daughter had never shown any apparent interest in or knowledge of the associated cooking process. During this experience, the daughter did not engage in any pointing behaviour or gaze following; she manipulated her mother as if she were an object in the world that would perform certain behavioural tasks (in this case food preparation) when given proper stimuli. This anecdote indicates that the daughter both was not relating to her mother as a second-person but as a sophisticated object in the world and that she was able to understand some complex things in the world. She remained, however, impervious to communication of understanding from her family. Clara Claiborne Park, *The Siege: A Family's Journey Into the World of an Autistic Child*, New edition (Boston: Little, Brown and Company, 1982).

longstanding incapacity to engage in joint attention with her parents left Park's daughter with apparently significant impediments to actualising her understanding, which were later partially mitigated when her parents were able to induce her to adopt a partial second-person perspective. This anecdotal evidence illustrates the capacity of second-person relatedness to stimulate insights and the necessity of such relatedness for acquiring the usual range of human understandings.

It is very possible that accelerating understanding is a social matter, and it is not obvious that technology will more effectively accelerate insight by novel means than by building on the already extant means of gaining insight through social interaction.

The above considerations bear a similarity to Lonergan's observation that a teacher does not merely impart information to pupils but primes them to recapitulate the teacher's own acts of understanding. What is implied is that understanding itself can be shared, but that it cannot be communicated in the same sense that one can communicate the facts, data, or phenomena that an act of understanding is about. Whatever else may be involved, the observations above imply that joint attention may play an important role in the acceleration of shared understanding.

3.4 Observations on Understanding in Theology

3.4.1 Light and Divine Illumination

As noted previously, one of the most striking metaphors for insight is that of gaining illumination, and this metaphor of understanding in terms of illumination or light hints at another question one might ask. Given that the experience of understanding is often compared to being illuminated, what then is the source of this illumination? Is this illumination internal or external to some working of the mind, or is the notion of an actual source a helpful or

misleading illusion? Given the epistemic value of insight, such questions would seem to be extremely important.

A particularly striking example of this in ancient literature is Plato's allegory of the cave in the *Republic*, in which Plato imagines that all human beings are born as prisoners in a cave, unable to see anything other than moving shadows cast upon a wall by the silhouettes of puppets that are behind them and backlit by a fire. For Plato the path to true enlightenment is to turn away from the shadows of the puppet show, first seeing the show itself and eventually escaping from the cave where it is possible finally to see real things, illuminated by the sun, culminating in the vision of the sun itself, which is also Plato's image of the divine.¹⁹⁰

This story is clearly about understanding, with understanding represented by the metaphor of sight, and the capacity for sight growing with the illumination that increases by stages from the ignorance of seeing only shadows to the wisdom of the full light of day. The image of sun implies something beneficial but also remote and uninvolved in the struggles of the potential philosopher. The prisoners need help to be drawn out from the cave, an action denoted by the Latin word *educere*, the root of the English word "education". Nevertheless, helping the prisoners is fraught with danger, given that, as Plato states, they would kill, if they could, someone coming from the realm of light to disturb them. What is also striking, however, is that the image of sun is also one of Plato's images of the divine, suggesting the perception of a link between insight and God.

One of the key lessons of Plato's cave, that supreme understanding is represented by the supreme source of light, which is also an image of the divine, can also be found explicitly in

¹⁹⁰ Plato, *Republic* 514a–520a, trans. G. M. A. Grube and C. D. C. Reeve, Ebook Central (Indianapolis: Hackett PubCo, 1992), 189, <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=735561>.

what Christians regard as the Scriptures, namely the Old and New Testaments. Indeed, the image of God as light is one that saturates the Scriptures. Examples from the Old Testament include “Yea, Thou art my lamp, O Lord, and my God lightens my darkness;”¹⁹¹ “...our God may brighten our eyes and grant us a little reviving in our bondage...;”¹⁹² and “the Lord opens the eyes of the blind.”¹⁹³ The sense of these statements is not only that God illuminates those who are in darkness, but exercises divine power to provide succour or healing to those who are in some way constrained or of themselves lack the power of vision, to enable sight and intellectual cognition related to sight.

Other references include “For with thee is the fountain of life; in thy light do we see light;”¹⁹⁴ and, “O send out thy light and thy truth; let them lead me; let them bring me to thy holy hill and to thy dwelling.”¹⁹⁵ In these references, the light is not only associated with truth, and by implication the cognitive acts pertaining to truth, but also to proximity to the divine. Further, these passages are taken from the Psalms, and read as prayers addressed to God with the hope of receiving a response. In other words, the source of light is treated in the manner of a *person* who may be responsive to a plea for assistance. Consider an excerpt from Psalm 119, “The unfolding of thy words gives light; it imparts understanding to the simple.”¹⁹⁶ God’s words are said to illuminate and give understanding to those who lack it. Taken together, these selected passages illustrate that God is strongly identified by Christian Scriptures as a source of intellectual light.

¹⁹¹ 2 Samuel 22:29 RSV

¹⁹² Ezra 9: 8 RSV

¹⁹³ Psalm 145:8 RSV

¹⁹⁴ Psalm 36:9 RSV

¹⁹⁵ Psalm 43:3 RSV

¹⁹⁶ Psalm 119:130 RSV

The New Testament is also replete with dramatic references to God in terms of light, such as at the very beginning of the Gospel of John,

In the beginning was the Word, and the Word was with God, and the Word was God. He was in the beginning with God; all things were made through him, and without him was not anything made that was made. In him was life (ζωή), and the life was the light of men. The light shines in the darkness, and the darkness has not overcome it.¹⁹⁷

In this text, the term life refers to divine life (ζωή), associated in the Gospel with the Father, Son, and Spirit, and this is described as the “light of men.” Hence God’s sharing of divine life is closely associated with light and hence also with some divine communication of understanding.¹⁹⁸ Indeed, on one occasion with a similar theme, God is even identified as light,

...God is light and in him is no darkness at all.¹⁹⁹

Christian thinkers have glossed these references to light explicitly in terms of divine understanding and divinely given inspiration. Aquinas considers this passage to show “that he produced things by his intellect and will”.²⁰⁰ Aquinas supports this interpretation by his identification of the Word as the act and content of divine understanding. For this reason, the Father creates through divine understanding, which is to say, through the Word. In other words, the Christian conception of reality is that intelligibility is logically prior to existence, and because it is through the divine intelligible ordering of the Word that all things were made, the same divine Word can also illuminate the human mind with the understanding of creation.²⁰¹

¹⁹⁷ John 1: 1-4 RSV

¹⁹⁸ This word ζωή, translated as life, used specifically to refer to divine life. By contrast, when Peter declares “I will lay down my life for you.” the word used for life is ψυχή, denoting the principle of animal life. John 13:17, RSV.

¹⁹⁹ 1 John 1:5 RSV

²⁰⁰ Thomas Aquinas, *Commentary on the Gospel of John*, trans. James A. Weisheipl, O.P. (Magi Books, Inc., 1998), 58.

²⁰¹ From this a rough answer may be given to the question posed by Eugene Wigner as to why mathematics appears able to describe structures and regularities in the natural world, both are united in the divine understanding which is the cause of the natural world. Eugene Wigner, “The Unreasonable Effectiveness of Mathematics in the Natural Sciences,” *Communications in Pure and Applied Mathematics* 13, no. 1 (1960), <https://www.dartmouth.edu/%7Ematc/MathDrama/reading/Wigner.html>.

A representative exegetical reflection on the metaphor of God and the meaning of the Word as light was given by Thomas Aquinas,

And so his statement, And that life was the light of men, can be understood in two ways. First, that the light of men is taken as an object that man alone can look upon, because the rational creature alone can see it, since he alone is capable of the vision of God who “teaches us more than the beasts of the earth, and enlightens us more than the birds of the air” (Jb 35:11); for although other animals may know certain things that are true, nevertheless, man alone knows the nature itself of truth. The light of men can also be taken as a light in which we participate. For we would never be able to look upon the Word and light itself except through a participation in it; and this participation is in man and is the superior part of our soul, i.e., the intellectual light, about which the Psalm (4:7) says, “The light of your countenance, O Lord, is marked upon us,” i.e., of your Son, who is your face, by whom you are manifested.²⁰²

Aquinas first identifies the Word as an object of understanding open to rational natures. Rational creatures are given to understand the divine, which is to say that they are to be able to share in God’s self-understanding by a participation in this understanding that is itself a divine gift. Furthermore, the way in which Aquinas describes the communication of understanding from God to enlightened humanity is also extremely suggestive. Aquinas describes this sharing in terms of a shining forth of the divine countenance in response to a prayer. In other words, this communication is likened to the turning of a face to one who has called out. An appropriate metaphor for this shared understanding is in the context of joint attention with another person.

Before further exploring the significance of the interpersonal component of the metaphor of God’s light as that of a face, it is necessary to explore more deeply the meaning of this Word (*λόγος*), and its role in the Godhead, that is said to be the light. Aquinas develops his account by analogy to the act of understanding in the human mind, so an examination of this divine act of understanding. Aquinas follows Aristotle in holding that an exterior word is a

²⁰² Aquinas, *Commentary on the Gospel of John*, 60.

vocal sound is a sign of a passion of the soul and that in the case of understanding this is a conception of the mind, which he calls the interior word, logically prior to any locution. It is according to the notion of the interior word that he wishes to interpret the evangelist's meaning of "the Word".²⁰³

To explain the meaning of the phrase "interior word", Aquinas asserts three things to be in the intellect, and the intellect as Aquinas describes it corresponds to what I in this thesis refer to as the faculty of understanding: the intellectual power, the species of that which is understood (or "intelligised", if one wishes to hew closely to the Latin), and the activity of understanding itself. The meaning of an external word, says Aquinas, does not correspond to any one of these three things but to the notion (*ratio*) that is understood, which the intellect creates by unifying and dividing. This meaning of "Word" is what Aquinas wishes to elucidate. To achieve this, he first notes that the notion is a likeness to what is understood, and that in the case of the self-comprehension of the intellect, there is a similitude to itself in the intellect. He then argues that the Word to which John refers must be the divine word as distinct from human and angelic words, since this Word is said to exist in the beginning, which is impossible for the created words of humans and spirits.²⁰⁴

According to Aquinas, this divine Word differs from creaturely interior words in three ways. First, creaturely understanding is not eternal, and the act of insight involves a change from non-understanding to understanding. In humans, this takes place through cogitation as the mind "casts about" for a suitable interior word for its object. Aquinas makes a possible exception for the first principles, which appear to be known *a priori*. What is of primary concern with respect to this human description of understanding is that one understands the first

²⁰³ Aquinas, *Commentary on the Gospel of John*, 31.

²⁰⁴ Aquinas, *Commentary on the Gospel of John*, 32–33; See also *ST I* q. 14 a. 4.

principles from which one reasons discursively, and the fruit of the discursive reasoning is itself an understanding. The operation of the intellect occurs after this cogitation, but these powers are metaphysically distinct. Second, this divine Word is perfect, embracing the divine mind itself and all created things, as opposed to the understandings of creatures, which consist of multiple conceptions failing to capture the totality of being and only expressible in locutions similarly various and limited. Third, this divine Word is identical with the act of Divine understanding itself.

There is yet another way in which this Word is “the light of man.” In Aquinas’ system, the blessed are granted to share in divine understanding. Thus, they are able to know God and through God, all things. In other words, shared understanding is a central feature of the Christian image of heaven.

For we would never be able to look upon the Word and light itself except through a participation in it; and this participation is in man and is the superior part of our soul, i.e., the intellectual light, about which the Psalm (4:7) says, “The light of your countenance, O Lord, is marked upon us,” i.e., of your Son, who is your face, by whom you are manifested.²⁰⁵

The understanding of the blessed is described as a kind of face-to-face encounter. As the contemplation of the divine is for Aquinas the end of human life, this means that understanding analogous to interpersonal awareness is the highest good depicted in Christian thought.²⁰⁶

²⁰⁵ Aquinas, 60.

²⁰⁶ cf. Aquinas, *ST I-II* q. 2, a. 8, q. 3, a. 8. Aquinas identifies God as the universal good of all things, and therefore the final good of human beings. He denies that any created good can offer more than imperfect satisfaction to human desire, which he regards as oriented toward the good and fully satisfied by the good. Humans, differentiated from animals by possessing a rational soul, require the satisfaction of their desire to understand, which remains unsatisfied as long as further questions can be understood (This failure to be intellectually satisfied while something yet is unknown is what Lonergan means when he describes the desire to understand as unrestricted and its object as being, understood as the sum total that is true and desirable for the mind to know). Apart from practical limitations to one’s time, Aquinas regards understanding gained through abstraction from sense experience or reasoning from first principles to be in principle insufficient for humans to grasp such things as the essence of God (except insofar as, he thinks, such can be proven to exist). Therefore, it is necessary on Aquinas’ account for humans to be granted participation in God’s own understanding by divine power rather than their own.

According to Christian belief, God's communication of divine understanding to human beings was accomplished in particular by the incarnation of the Word as the person of Jesus.²⁰⁷ By the narrative of the two testaments joined together, the symbolism of encountering God's countenance as described in the Psalms or in the cases of certain blessed persons is transformed into a literal encounter face-to-face with God made flesh.²⁰⁸

3.4.2 Narrative and Scripture

In evaluating the narrative treatment of insight and understanding in scripture, this thesis must provide interpretations of the meanings of these texts. Given the relevance of these texts to the theological tradition extending from Aquinas to Lonergan, this thesis treats the texts in a manner consistent with this tradition, as unified or unifiable works presenting material for theological reflection. This section does not attempt to evaluate the historical accuracy of these texts or attempt to provide a historical account of their composition as this is outside the scope of this thesis. Instead, this section follows the methodology provided by Eleonore Stump in *Wandering in Darkness*,²⁰⁹ in which she observes that both the medieval methodology and historical methodology aim to resolve inconsistencies and tensions in the texts, the former by making suppositions that render the text consistent (for example, explaining the Gospel accounts differing in the reported number of angels encountered by Mary Magdalene when she visited Jesus' tomb by positing that Mary Magdalene made multiple trips, each time

²⁰⁷ John 1:14 RSV "And the Word became flesh and dwelt among us..." For how Polanyi's notion of the tacit dimension relates to religious belief and the incarnation of Christ see Oliver Keenan, "Semantic Dogmatics in Outline: Cornelius Ernst's 'meaning of Meaning' as Michael Polanyi's Absent Doctrine of God" (2018), 259–331.

²⁰⁸ 1 John 1:1–3 RSV "That which was from the beginning, which we have heard, which we have seen with our eyes, which we have looked upon and touched with our hands, concerning the word of life—the life was made manifest, and we saw it, and testify to it, and proclaim to you the eternal life which was with the Father and was made manifest to us—that which we have seen and heard we proclaim also to you, so that you may have fellowship with us; and our fellowship is with the Father and with his Son Jesus Christ."

²⁰⁹ Eleonore Stump, *Wandering in Darkness: Narrative and the Problem of Suffering* (Oxford University Press, 2010), 22–38.

encountering a different number of angels), the latter by positing the existence of primitive, consistent sources from which the current texts were compiled and attempting to reconstruct these texts. Neither is capable of definitively proving that its interpretation is correct, and her purpose is not, at any event, to argue for a definitive interpretation of scripture but to defend the legitimacy of presenting an interpretation that resolves tensions in the text in a philosophically coherent narrative.

As a prelude to discussing the nature of narrative, a brief reference to the biblical use of non-expository language as a means of conveying understanding is appropriate. The distinction between narrative and poetry is not an exclusive one. After all Dante's *Commedia* is the story of a man's journey and spiritual regeneration and perfection told in poetry. The teleology of narrative requires temporal progress of its characters whereas poetry does not. Both narrative and poetry rely on a capacity for imagination to share experience and understanding, so an examination of Biblical poetry will prove illuminating with respect to the communicative capacities of Biblical narrative. Elaine James analyses Biblical poetry and finds it possesses several qualities that this thesis has found relevant to effect of narrative on understanding.

First, Biblical poetry evokes bodily experience.

On this view, we approach art with all the emotional, spiritual, physical complexity of our embodied human lives. This is true for worshippers who sing or pray the psalms and feel them in their lungs and throats and mouths, vibrating with the voices of their communities; it is true for readers who are drawn into the emotional drama and psychological depth of Job's debate with his friends; it is true for lovers who find themselves swept into the expansive, congenial world of the Song of Songs with its romantic saturation and lilting, springtime cadences; it is equally true for audiences who are compelled to disgust or shame or anger by some of the terrifying imagery of the prophets. These are bodily experiences. The poems ask us to feel with them.²¹⁰

²¹⁰ Elaine T. James, *An Invitation to Biblical Poetry*, Essentials of Biblical Studies Series (Oxford University Press, 2022), 10, <https://doi.org/10.1093/oso/9780190664923.001.0001>.

On Elaine James' view, Biblical poetry is, therefore, intimately connected with embodiment. The language evokes bodily experiences that allow the reader to imaginatively share in the perspective of the author and the speakers of the poems. The emotions evoked and experienced have an embodied element that is also conveyed. Implicitly, then, the body is the vehicle by which these shared experiences can be conveyed to the reader.

Biblical poems also possess the character of often being participative through some liturgical practice or ritual, which involves the very act of vocalising in an embodied context. Using language is an embodied act, but the attention of the speaker is usually not on the muscular actions needed to form phonemes, for instance. By attending to the context of ritual and communal participation, one is reminded that the recitation of biblical poetry was (and in modern liturgical practice remains) a communal act of embodied persons in which there is both mutual awareness of each other and a shared awareness of the vicarious experience evoked by the imagery of the poetry. The liturgical act of sharing an experience of Biblical poetry is possible due to the embodiment of the participating intelligences. The conveyance of understanding through the poetry and its shared reception by the congregation depends upon these sensorial and imaginative capacities.

The embodied experiences that the Biblical poems draw upon and which they evoke are not necessarily didactic even when they have some pedagogic purpose. They afford the reader the capacity to imaginatively share in an experience.

As art, these poems do intellectual work. They are deeply engaged in the questions and problems of human experience. I emphasize this in part because the Bible is often popularly characterized as didactic literature—its purpose is instructional, it aims to teach, and therefore the goal of reading is to extract the correct “message.” Biblical poetry *can* do these things, and some biblical poems are manifestly didactic (think, for example, of the many examples in the book of Proverbs). But to imagine that this is the only thing biblical poems do is needlessly limiting. As the example of Psalm 91 helps us see, the value of the poem is not only in our

ability to reduce it to a digestible point of instruction. Rather, the poems are creating experiences that invite our deep consideration and participation.²¹¹

Biblical poetry stimulates intellectual work not merely through a simplistic didactic format of giving explicit instruction but through stimulating imaginative engagement with the text through embodied language (among other poetical devices) that allows the author to evoke (and thereby share) experience with the reader. Through such imaginative experiences, the Biblical poems can evoke and shape the attitudes and perspectives of the reader, which can allow a sharing of implicit knowledge.

Additionally, poetry depicts personal relations, often in the form of dialogues between characters in I-thou relations. The Song of Songs provides a striking example of such address when the lover and beloved express their mutual longing for and appreciation of one another.

Admittedly, some poems are not narratives, but many poems are also narratives such as the Homer's *Iliad* or Dante's *Commedia*. The questions of which Biblical poems are also narratives and which ones are not narratives does not need be evaluated here. It is sufficient to observe that the characteristics of Biblical poetry identified by Elaine James above also belong to scriptural narrative. Biblical narratives, whether or not they possess prosody, present imagery and metaphors that rely on shared imaginative faculties and capacities for embodied experience between the author and the reader and not only allow a vicarious experience of the characters in the story. Narratives also possess the added detail of change through time and principally concern the journeys (metaphorical or literal) of characters, which are not essential to poems. Nonetheless, since a narrative may be a poem and even those narratives that are not poems still rely on images that rely on embodiment and since Biblical narratives are still recited in liturgical

²¹¹ James, 5, emphasis original.

contexts, the above comments on Biblical poetry are applicable to Biblical narrative. Biblical narratives depend upon embodiment for the representation and communication of their content.

Robert Alter's observations about Biblical narrative reinforce a similar point about the importance of embodiment for narrative.

A proper narrative event occurs when the narrative tempo slows down enough for us to discriminate a particular scene; to have the illusion of the scene's "presence" as it unfolds; to be able to imagine the interaction of personages or sometimes personages and groups, together with the freight of motivations, ulterior aims, character traits, political, social, or religious constraints, moral and theological meanings, borne by their speech, gestures, and acts.²¹²

Alter here distinguishes narrative events from summary or the simple provision of detail such as lengthy genealogies. A particular scene is, principally, a part of the whole that has an internal intelligibility but is also most intelligible when grasped as part of the greater whole. In these particular parts of narratives that are called scenes, it is possible to imagine a whole sequence of actions and interactions on the part of the characters. It is through such an act of imagination that the scene creates the illusion of presence Alter describes. Further, it is through descriptions of embodied actions including speech and gestures that the characters' perspectives, motivations, and contexts can be grasped as well as further moral and theological meanings. The essential role of embodiment in narrative is here emphasised.

Jesus' use of parables illustrates the well-known power of narratives to convey understanding. For instance, when offering a rebuke to the Pharisees for their refusal to accept them, he tells a story about unfaithful tenants who betray the owner of the vineyard they have been hired to keep. The Pharisees realise he was speaking of them, though he did not directly refer to them. Narratives can produce an explicit understanding even though their language

²¹² Robert Alter, *The Art of Biblical Narrative*, Ebook Central (New York: Basic Books, a member of the Perseus Books Group, 2011), 79, <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=679754>.

itself is implicit, and it is through the narratives of parables that Jesus primarily instructed the crowds.

All this Jesus said to the crowds in parables; indeed he said nothing to them without a parable.²¹³

The very fact that Jesus uses parables so extensively highlights that this genre is of tremendous importance to the Christian tradition.

The disciples are said to be granted the understanding that the people to whom Jesus primarily addresses parables lack. This passage makes use of the metaphor for understanding discussed above, namely that it is a kind of seeing, as evinced by the fact that Jesus explicitly states that those who are not his disciples hear without understanding and see without seeing. The second use of seeing is clearly in a different sense and is most reasonably interpreted as another sense of understanding.²¹⁴

In the passage, Jesus identifies those who have not been granted understanding as the audience for his parables. The purpose of these parables is identified as pedagogical. This illustrates the well-documented power of narratives to convey understanding. The situation is analogous to that of giving instruction to a child who cannot grasp abstract points of moral theory but may nonetheless learn how to act through hearing stories of right and wrong action.

Further, Mark comments that Christ spoke in parables to the people “as they were able to hear it,” and that he only spoke plainly to his disciples.²¹⁵ As recounted in the Gospel of Mark, Jesus tells his disciples:

And he said to them, “To you has been given the secret of the kingdom of God, but for those outside everything is in parables; ¹² so that they may

²¹³ Matthew 13:34 RSV

²¹⁴ Matthew 13: 10-11, 13 RSV “Then the disciples came and said to him, ‘Why do you speak to them in parables?’ And he answered them, ‘To you it has been given to know the secrets of the kingdom of heaven, but to them it has not been given... This is why I speak to them in parables, because seeing they do not see, and hearing they do not hear, nor do they understand.’”

²¹⁵ Mark 4:33 RSV

indeed see but not perceive and may indeed hear but not understand; lest they should turn again, and be forgiven.”²¹⁶

More interestingly, the passage illustrates that experience of sense data is not identical with understanding. Jesus indicates that his words may be heard by many but understood by few. Identical auditory experiences will elicit understanding in different persons. This is in harmony with Lonergan’s observation that many people had sat in the baths of Syracuse but only Archimedes had insight into the relationship between volume, displacement, and weight, which he attributes to a difference in their inner conditions.²¹⁷

To identify the inner conditions facilitating understanding according to the Gospel narratives, the cause of insights needs to first be articulated. Jesus identifies this as divine gift. The explicit identification of God with the cause of insight comes when Peter voices his belief that Jesus is the son of God.

And Jesus answered him, “Blessed are you, Simon Bar-Jona! For flesh and blood has not revealed this to you, but my Father who is in heaven.”²¹⁸

The passage indicates that this insight is one which could not have come from mortal power alone but was a divine gift.

Famously, a scribe approached Jesus to ask him what the greatest commandment is, and Jesus told him that there were two, first, to love God with all one’s will, soul, heart, and mind, and second, to love one’s neighbor as oneself. The scribe assents that these are greater than all other laws,

And the scribe said to him, “You are right, Teacher; you have truly said that he is one, and there is no other but he; and to love him with all the heart, and with all the understanding, and with all the strength, and to love one’s neighbor as oneself, is much more than all whole burnt offerings and

²¹⁶ Mark 4:11-12 RSV

²¹⁷ Lonergan, *Insight*, 29.

²¹⁸ Matthew 16:17 RSV

sacrifices.” And when Jesus saw that he answered wisely, he said to him, “You are not far from the kingdom of God.”²¹⁹

First, Jesus omits the use of parables in this exchange, for the scribe answers wisely. This accords with his words to his disciples that he has not used parables with them for they have been given to understand. This suggests that a purpose of parables is to catalyse understanding when one fails to attain it, but that when understanding has been acquired, one can proceed to other matters. Second, the affective component of understanding is highlighted as being of essential importance in Jesus’ teaching.

The notion above of loving God with all one’s understanding might at first appear obscure, but it illustrates a particular instance of a more general case of desire and affect in understanding. There are two senses in which one can be said to love something with one’s understanding: in pursuing an understanding one does not yet possess and, after acquiring it, having joy in the understanding. Lonergan’s description of understanding as motivated by the desire for intellectual satisfaction is pertinent here. The impetus to seek understanding, according to Lonergan, comes from the sense of wonder of which Aristotle spoke. This desire for understanding is, he says, unrestricted; it is directed toward all that is true.²²⁰ To love with one’s understanding in this sense is to seek to understand, so the scribe’s words plausibly indicate that the first commandment entails the one should desire to understand God above all else, which is to say that one identifies God as the most worthwhile object of inquiry. It is not necessary here to evaluate this claim, which is related to the ancient ranking of theology as the

²¹⁹ Mark 12: 32-34 RSV

²²⁰ In this Lonergan does not imply that all that is true can be affirmed by human beings due to the finiteness of their resources for pursuing truth, including but not limited to finite time and instruments for investigation. There may in principle even be things that cannot in principle be proven by human reason. Nonetheless, the desire to understand, he says, extends in principle to all truth. If one cannot by one’s labors attain all truth, one still longs for it nonetheless. Lonergan, *Insight*, 372–80.

highest science. Rather, it is sufficient to show that this passage is plausibly congruent with the claim that understanding is preceded by desire.

The first sense of loving with one's understanding is the most pertinent here, for it is a longstanding tenet in Christian philosophy that one cannot fully understand God while living, though one can understand propositions about the divine.²²¹

Nonetheless, there is scriptural affirmation that a knowledge of God is possible even for the unaided reason.

For what can be known about God is plain to them, because God has shown it to them. Ever since the creation of the world his invisible nature, namely, his eternal power and deity, has been clearly perceived in the things that have been made.²²²

Paul implies that at least some understanding of God is achievable by natural reason and that rejecting God requires a failure of rationality. Without here attempting a defense of natural theology, I observe that this passage indicates scriptural confidence in the power of humans to arrive at understanding of at least some theological matters.

Examined from a theological perspective, narratives are of great significance to the Christian tradition. Narratives in the form of parables are treated as a privileged form of communication and teaching by Jesus, and the poetry of the Old Testament serves the purpose of drawing the reader and the community into imaginative and synaesthetic engagement with the text. Simultaneously, fully understanding narrative at least in the context of parables is something that occurs according to one's ability to see it. Jesus' explicit instruction of the disciples is treated as supplementary due to a failure to understand, with the implication being that the parables, if one is rightly disposed, are sufficient to learn what Jesus later describes

²²¹ Aquinas, *ST I* q. 88, a. 3.

²²² Romans 1:19-20 RSV

explicitly. Beyond that, the scholarly observations made about biblical poetry may be carried over to Jesus' use of parables. The parables, too, afford opportunities for imaginative engagement with the embodied imagery of the story and invite the listener or reader to adopt a perspective and attitude toward the imagery of the story. Narrative in scripture is not treated merely as an aid to explicit instruction but as conducive to another mode of intellectual engagement. Biblical narratives and imagery present the notion that understanding of revealed truth is the result of a kind of illumination from God, and that narratives are proper object for this illumination of the intellect beyond just what can be acquired through a list of explicit propositions to credit. The further personal elements of such illumination and its effect on the human person will now be considered.

3.4.3 Encounter and Transformation

Understanding and pedagogy are illustrated in the famous reported experience of Jesus by his disciples on the road to Emmaus. According to this reported encounter, two disciples were traveling to Emmaus when Jesus appeared to them, yet they did not recognise him.²²³ They say that they had hoped that Jesus would prove to be the Messiah, and they are much distressed at his crucifixion, which apparently falsified these beliefs. Within the scriptural narrative, it is at this point established both that Jesus' messianic claims were veridical and that he has risen from the dead. The disciples, then, have a twofold failure of understanding: the failure to recognise Jesus as he stands before them and a failure to correctly make sense of the seemingly incompatible facts of his messianic claims and actual death.

And he said to them, "O foolish men, and slow of heart to believe all that the prophets have spoken! Was it not necessary that the Christ should

²²³ Luke 24:15-16 In the Latin: "*oculi autem illorum tenebantur ne eum agnoscerent*" meaning their eyes were darkened and they did not recognise him. This phrase again uses the motif of ignorance as darkness and, by implication, its opposite must be light. I suggest that the opposite of this darkness is the light of understanding.

suffer these things and enter into his glory?” And beginning with Moses and all the prophets, he interpreted to them in all the scriptures the things concerning himself.²²⁴

Jesus first rebukes the disciples for their failure to believe the sayings of the prophets and that it was to be expected that aspects of his life that seem incompatible, such as his claim to be the Messiah and his sufferings, are in fact coherent. Their failure to reconcile these aspects of Christ’s life can be described as a failure to understand, whereby they have details of Christ’s life that, according to the logic of the Gospels, are necessarily united but which they have yet to recognise as unified. Christ’s subsequent interpretation of scripture for them can be considered a kind of pedagogy aimed at helping them acquire a higher perspective that resolves the puzzle of how Jesus could make messianic claims apparently supported by the power of God whilst also enduring the rejection of his people and suffering execution. According to this new perspective, the foretold messianic mission consisted primarily in such suffering, so the apparent contradiction disappears.

If this gloss of Jesus’ discourse is correct, it may be further interpreted according to Lonergan’s notion of pedagogy as the removal of obstacles to understanding and the pointing out of those details which are likely to prompt insight. In this case, it is necessary for the disciples to attend to those details of the psalms such as the suffering servant songs that apparent defeaters of Christ’s messianic claims might instead be seen, according to the logic of the narrative, to be in fact supports for it.²²⁵

²²⁴ Luke 24: 25-27 RSV

²²⁵ Aquinas in his commentary on Luke 24 quotes Theophylactus’ gloss of Luke 24:25, in which examples of the scriptural passages pointing to Christ are cited. “For it is possible to believe some of these things and not all; as if a man should believe what the Prophets say of the cross of Christ, as in the Psalms, They pierced my hands and my feet; but should not believe what they say of the resurrection, as, Thou shall not suffer your Holy One to see corruption.” Emphasis original. These are quotations of Psalm 22:16 and Psalm 16:10. Thomas Aquinas, *Catena Aurea: Commentary on the Four Gospels*, vol. III Saint Luke (Southampton: The Saint Austin Press, 1997), 777. Still, the essential point in these quotations is that the disciples should have believed but have failed to do so. Situations in which apparent problems for a theoretical position can actually deepen it are abundant. Famously, the Copernican model of planetary motion initially produced less accurate predictions of planetary motion than the

The next event in the episode of interest with respect to insight comes when the disciples invite Jesus to join them for dinner. He blesses bread and breaks it, at which point the disciples suddenly recognise him before he vanishes from their sight.

They said to each other, “Did not our hearts burn within us while he talked to us on the road, while he opened to us the scriptures?”²²⁶

The disciples’ recognition of Christ illustrates the suddenness with which insight can come. The recognition of a face is sudden and akin to recognising a gestalt. More striking, however, is the implicit result of their recognition of Christ. In light of the explanations that they had just received, about the necessity of Christ’s death and resurrection, to see him was also to see that he had risen. This revelation cohered with his messianic claims and with the scriptural interpretations with which they had just been provided.

This passage also illustrates the affective element of insight that Polanyi discusses. The disciples say that their hearts burned, which is to say, were passionate, whilst they spoke to Jesus and had the scriptures opened to them. This highlights two aspects of the passion associated with understanding with respect to other persons. The first is that, even while they did not yet identify Christ, they yet entered a shared perspective with him in which to see the subject at hand, the meaning of scripture, as he saw it was a source of delight in itself, the delight of sharing understanding. Second, the disciples experienced passion as the scripture was opened to them, a metaphor signifying that they were grasping meanings that hitherto eluded them, the acquisition of which was delightful.

Ptolemaic model unless the planets were postulated to move in circular orbits with epicycles about the sun. This problem eventually prompted Kepler to posit his three laws of planetary motion, which more elegantly and accurately described the motion of the planets than Ptolemaic and strict Copernican theory. In this case, an apparent problem for the heliocentric model of the solar system led to an insight that strengthened the model and incorporated the correct understanding of a problematic observation as further evidence. Similarly, Christ in the road to Emmaus narrative attempts to elicit in the disciples a new understanding of scripture in which the apparent evidence against his messianic claims instead is seen to be supportive of them when the nature of the promised Messiah is grasped.

²²⁶ Luke 24:32 RSV

The coordination of these two sources of delight should not be overlooked. Within the narrative, the disciples are not merely given information or an interpretive key to a body of texts in isolation from human relations. The disciples express first the confusing facts that Jesus' messianic claims were supported by miracles yet apparently undercut by his crucifixion at the hands of their rulers. Jesus acts pedagogically by telling them that they think Christ's messianic claims, miracle working, and crucifixion are incompatible due to a failure to believe the prophets. Although this seemingly characterises the disciples' failure as a lack of faith, Christ's need to interpret scripture for them in light of his suffering suggests that there is a misunderstanding of prophetic texts that must be corrected. This is further supported by the disciples' initial expectations during his ministry that he would be an earthly king gathering the people of Israel for military conquest and their confusion when he states that he must suffer, die and rise again from the dead.²²⁷ Finally, the disciples in the story of the encounter on the road to Emmaus remark that their hearts burned as he opened the scriptures to them, which suggests that they had not simply understood the scriptures' meaning and failed to believe them, but that they had been closed in the sense of having a meaning they had not yet seen. In this, the opening of the scriptures to the disciples parallels the opening of their eyes to recognise Christ. In both instances, the whole was there to be seen and no new information was provided. Instead, a change occurred in the disciples that allowed them to recognise something previously hidden, which is compared to an opening of the eyes.

Jesus' rebuke may be seen as an effort by a teacher to make his pupils aware of a misunderstanding and direct their attention to the source of confusion. The report that the

²²⁷ Mark 11:9-10 RSV "And those who went before and those who followed cried out, 'Hosanna! Blessed is he who comes in the name of the Lord! Blessed is the kingdom of our father David that is coming! Hosanna in the highest!'" This passage, taken from Mark's narrative of Jesus' entry into Jerusalem illustrates the popular expectation for the restoration of the Davidic kingdom.

disciples' hearts burned within them as he spoke to them may be explained as an instance of the shared passion of pedagogue and pupils that gives the pupil the drive to come to understand. This shared attention on the matter which the pupil must understand is accompanied by affect on the part of the pedagogue and pupils in which each desire that the pupils might gain the understanding that allows shared awareness of the object of understanding. The second cause of passion mentioned, Jesus' explanation of the scriptures, may be taken to illustrate the excitement and satisfaction that comes from gaining understanding. The second cause of passion, understanding the scriptures, could in principle have come without the first cause of passion, desire to share Jesus' understanding, yet the former created conditions for the latter and allowed the teaching to progress quickly.

Perhaps the most famous story in the Gospels illustrating the association between insight and illumination is that of the conversion of Paul. Previously having the name of Saul, he zealously sought to extirpate the Christians of Jerusalem believing them to be a heretical Jewish sect.

As he neared Damascus on his journey, suddenly a light from heaven flashed around him. He fell to the ground and heard a voice say to him, "Saul, Saul, why do you persecute me?" "Who are you, Lord?" Saul asked. "I am Jesus, whom you are persecuting," he replied. "Now get up and go into the city, and you will be told what you must do."²²⁸

The sensory experiences of the bright light and the locution co-occurred and elicited in Saul a new understanding of Jesus' identity. This scene again uses illumination as a motif of the divine and of insight. The suddenness of the insight is portrayed both by the suddenness of the flash of light as well as the abrupt change in Saul's disposition. Previously, he had believed himself to be serving the Lord by opposing those who followed Jesus, but after his conversation, he

²²⁸ Acts 9: 3-6 RSV

instead has an understanding that the Lord is to be identified with Jesus, for he addresses Jesus as Lord and asks what he is to do. In other words, Paul's conversion has the character of an insight experience in which imagery of the divine as light and as a person are both present.

There is one further element of Paul's conversion worth remarking on. In Acts, Paul relates that he received the teaching of Christ not as one receives it from a human but directly from God. The implication is that, whether in the moment of his conversion or afterward, he was given knowledge and understanding directly by the person of God rather than from another human being. This is an extraordinary case of inspiration but is related to the notion of the gift of the Spirit pertaining to understanding, discussed below. For now, it is sufficient to show that the Christian conception of the divine is as not only an object or remote cause of understanding but is a personal catalyst.

Unlike in Plato's metaphor, the divine does not merely grant enlightenment passively to those who have been taken from darkness by human effort alone but is portrayed as actively giving illumination. While the divine light of Plato's myth is an agent only in the sense of acting on the person's eyes when they look to it or one of the objects it illuminates, in scripture the divine is portrayed as willing that those who suffer or are in ignorance be given illumination and intervenes to achieve this. In effect, the divine takes upon the roles of both sun and rescuer of Plato's allegory.

This brings out a notable trait of the Christian conception of the divine. God appears to humans not merely as an object of contemplation but as an interested agent. Further, God appears to humans as a person. A face-to-face encounter with God inaugurates the work of some prophets, such as Isaiah. From these experiences, they come to understand God's word and become able to communicate this to the people of Israel. In other words, they gain a twofold illumination, insight into God and into divine teaching. This suggests that the Christian

conception of insight places particular emphasis on the role of personal encounter in gaining understanding. Furthermore, there is a systematic account of this relationship and its implications in Aquinas' theological anthropology, which will now be examined.

3.4.4 Aquinas and the Gift of Understanding

Aquinas, speaking for the theological tradition, recognises God-given virtues and gifts, which are intrinsically different from the ones acquired by human practice and not commensurate with any natural or artificial means of enhancement. Aquinas recognises a greater range of God-given virtues than the Scotist tradition; in addition to Faith, Hope, and Charity there are God-given versions of Prudence, Justice, Courage, and Temperance. He develops a distinctive account of the Gifts (the Scotist tradition sees them as different names for what we already know as (infused, supernatural) Faith, Hope, and Charity, and (acquired) Prudence, Justice, Courage, and Temperance). Both the infused virtues and the Gifts are best understood as elements in an irreducible second-person relationship with God: the infused virtues mould our psyche to befit friendship with God; the gifts attune us to the Spirit's personal guidance. A second-person relationship cannot be manufactured by first person, individual enhancement. The gift of understanding usefully illustrates this point.

Aquinas argues that the final end of human beings is God and that a principal component of the satisfaction of this end is having the correct orientation of one's will, intellect, and passion to God.²²⁹ Of concern here, is the way in which the satisfaction of this end involves both personal relationship and understanding. That Aquinas regards understanding as of paramount importance in human life may be seen from his assertion that a disembodied soul would not be lacking in good because

²²⁹ Aquinas, *ST* I-II q. 1.

the end of the human soul is not to move the body, but to understand, in which its happiness consists...²³⁰

To understand is the ultimate purpose of the human soul and to attain understanding is happiness. Given that Aquinas considers the faculties of the soul and their ends to be the highest in human life, he is effectively making the claim that to understand is the greatest purpose in human life.²³¹ If the purpose of human life is to understand, the question arises if there is any particular object toward which the mind is to be directed. Aquinas answers the question positively and states that

For man and other rational creatures attain the ultimate end by knowing and loving God...²³²

If to know and love God is the highest end of a rational being, then the ultimate end of the understanding, too, must be to have the divine as its object. To know and understand the divine is not enough, however; human persons must also love the divine to fulfil their end. Thus far, Aquinas' account of the relationship between human understanding and the divine does not deviate significantly from those of Plato and Aristotle. Both held the divine to be the worthiest object of human desire and love and contemplating the divine to be the highest human good. Aquinas, however, also argues that there is a kind of understanding that can only be acquired by direct divine inspiration,

Therefore, a man needs a supernatural light in order to penetrate further, so that he might have cognition of certain things that he is unable to have cognition of by the natural light. And this supernatural light that is given to a man is called the gift of understanding.²³³

²³⁰ Thomas Aquinas, *On the Unity of the Intellect against the Averroists*, trans. Beatrice H. Zedler, Mediaeval Philosophical Texts in Translation; No. 19 (Milwaukee, Wisc.: Marquette University Press, 1968) para. 116.

²³¹ Ibid. Aquinas endorses Aristotle's argument for the supremacy of understanding among other human acts in *Nicomachean Ethics*, 1094^a1-3, 1177a11-1178^a8

²³² Aquinas, *ST* I-II q. 1 a. 8.

²³³ Aquinas, *ST* II-II q. 8 a. 1, see also I q. 88 a. 3.

This gift of understanding, described along with the natural understanding according to the metaphor of light, is described as capable of permitting one to know what is beyond one's natural powers of understanding.²³⁴

The gifts of the Spirit are

the gifts are habits that perfect a man so that he follows without delay the promptings of the Holy Spirit, in the same way that the moral virtues perfect the appetitive powers so that they obey reason. But just as the appetitive powers are apt to be moved by reason's commands, so all the human powers are apt to be moved by the promptings of God as a higher power. And so just as there are virtues for all the human powers that can be principles of human actions, viz., reason and the appetitive powers, so too there are gifts for all those powers.²³⁵

These gifts of the Spirit dispose human powers, including understanding, to be moved by the Spirit analogously to how the human intellect moves the appetites. Notably, the gifts are not equivalent to the natural virtues not only in their origin but in their operation. In contrast with the natural virtues, which, once acquired, may be actualized according to a human's natural capacities alone, the gifts dispose one to act in response to a further prompting. In other words, the gifts are not only not produced by independent action but necessitate continued relation with the divine.

This amenability of the human being to being moved by the divine is applied to the gift of understanding as well.

...the gifts of the Holy Spirit perfect the soul in the sense that it becomes easily moveable by the Holy Spirit (*secundum quod est bene mobile a spiritu sancto*). So, then, the gift of understanding is posited as an

²³⁴ Aquinas, *ST* II-II q. 8 a. 1 "Discursive reasoning always begins with understanding and ends with understanding, since we reason discursively by proceeding from things that are understood, and discursive reasoning is completed when we arrive at the point of understanding what we were previously ignorant of. Therefore, whatever we reason discursively to proceeds from some prior understanding. By contrast, a gift of grace does not proceed from the natural light but is instead added to it in the sense of perfecting it. And so this addition is not called reason, but is instead called understanding, since this added light is related to the things that become known to us supernaturally in the way that the natural light is related to the things that we have cognition of from the beginning." This understanding is grasped as something true but not achievable by any power of reasoning or any extension of the cogitative process.

²³⁵ Aquinas, *ST* I-II q. 68 a. 4.

intellectual light of grace insofar as man's intellect becomes easily moveable by the Holy Spirit. Now the thinking (*consideratio*) involved in this movement lies in a man's apprehending the truth concerning his end. Hence, unless the human intellect is moved by the Holy Spirit to the point of having a correct estimation of its end, it has not yet received the gift of understanding—no matter how much it understands about certain preambles [to the Faith] because of the illumination of the Holy Spirit.²³⁶

This receptivity to divine movement is described as a receptivity to grace, which, with respect to the gift of understanding, enlightens the mind with respect to truths concerning the supernatural final end of humanity. As discussed above, Aquinas interprets this final end to be an understanding of the divine. The gift of understanding is not equivalent to the beatific vision, however, which Aquinas holds to be achieved only by the saved souls. This understanding, rather, is partial, pertaining to particular propositions with respect to the divine and how to live rightly.²³⁷ Both sets of propositions concern the human end by being about either the end itself or about the means to attaining it. The nature of the grasp of these propositions is notable because they are simply grasped without use of ratiocination, that is, by reasoning from prior propositions grasped as true.

Further, there is a new interpretation of this account, according to which the kind of relationship with God that is spiritually fruitful is what is today called “second-person relatedness.” Second-person relatedness denotes the relations that occur between two persons when they are mutually aware of one another, and each is aware of the other's awareness. The phenomenon is common in daily life in interaction between persons and is evinced in

²³⁶ Aquinas, *ST II-II* q. 8. a 5.

²³⁷ Aquinas, *ST II-II* q. 8 a. 3 “...the gift of understanding has to do not only with those things that fall under the Faith principally and in the first place, but also with all the things that are ordered toward the Faith. Now certain good operations are ordered toward the Faith, since, as the Apostle says in Galatians 5:6, “Faith operates through love.” And so the gift of understanding extends even to certain actions—not, to be sure, as what it is principally concerned with, but insofar as, according to Augustine in *De Trinitate* 12, we are regulated in our actions “by the eternal conceptions (*rationibus aeternis*), which higher reason adheres to by looking to them and consulting them.” And it is higher reason that is perfected by the gift of understanding.”

behaviours as familiar as pointing and gaze-following and are especially evident in children. In such interactions, one person is aware of an object and wishes to bring it to the other's attention. The first makes a gesture that draws the second's attention, and the second, moved in some way by the first's gesture, turns their attention to the object of ostension. The two persons may be said to have a shared stance toward the object, e.g., that of awareness of, which becomes more sophisticated when the two persons share a further orientation toward the object, say of an affective response or intention to use for some purpose.²³⁸

The gifts of the Spirit, as noted above, are distinguished from the virtues in that they dispose one to be moved by the divine. The virtues may be described as having a dyadic relation of person to object, but the gifts have a triadic structure composed in which relations exist between the person, God, and the object. In the first case, the person is disposed to act with respect to the object, while in the second, the person is disposed to be moved by God to act with respect to the object. The relationship between God, person, and object bears a striking resemblance to the relationship between two persons and a third object in the summary of joint attention above. The divine has perfect understanding of all things, including speculative and practical truths necessary for human salvation, of which the person is ignorant, and the person is moved by the divine to understand and attend to those truths. For further exposition of this interpretation, see Andrew Pinsent, *The Second-Person Perspective in Aquinas's Ethics: Virtues and Gifts*.²³⁹

An additional point, to be further explored in the next chapter, is that Aquinas' philosophical account of the function of the human psyche, which assimilates and develops

²³⁸ Hobson, "What Puts Jointness Into Joint Attention?"

²³⁹ The gifts and virtues are alike insofar as they are dispositions to act and are about the same objects, but the gifts enable a person to be moved by God as by a second-person. See Andrew Pinsent, *The Second-Person Perspective in Aquinas's Ethics: Virtues and Gifts*, 1st ed (London: Routledge, 2013), 31–32.

from the Aristotelian tradition, recognises the importance of embodiment in human cognition, and sees the human soul as both a spiritual substance and the form of the body understood in an Aristotelian way. This philosophical account is affirmed by Christian doctrines of the resurrection of the body and the Beatific Vision. The doctrine the resurrection of the human body affirms that complete human fulfilment is not exclusively cognitive. Further, the resurrected body is radically different from the body as it currently is and as it could be made through enhancement. The doctrine of the Beatific Vision, as understood by Aquinas, requires the soul to possess an intellect and will that can be “enlarged” by grace to “embrace” God as the Known and Beloved, which is different from all forms of natural knowledge, including that of the angels. Consequently, the ultimate goal of friendship with God and sharing God’s own bliss is not attainable through any kind of enhancement, both because of what it is and because it is second-personal. These theological doctrines which affirm and enlarge the Aristotelian-Thomist philosophical perspective will be addressed in the next chapter’s thematic evaluations of the transhumanist perspective according to the grounding of insight in embodiment (4.2.2) and the role of second-person relatedness to insight (4.2.5), respectively.

3.5 Concluding Remarks on Insight

Whatever else may be said, a vast range of sources from everyday conversation to ancient texts to contemporary psychology suggest that there is some cognitive act that answers to the name “understanding” and that this act is foundational to human cognition. This existence and importance of this act is highlighted, in particular, by those changes in understanding that go by the name “insight”, along with concomitant terms like “Eureka!” or “Aha!” moments. Moreover, the notion of insight, along with its metaphor of enlightenment, has a long-standing perceived connection with notions of the divine. Examples include Plato’s cave and metaphors

of light in descriptions of God in the Judeo-Christian scriptures. In recent research, the connection between insight and the divine has also been framed in the terms of second-person relatedness. On this account, special divine insights are catalysed in the context of joint attention with the Holy Spirit, whom Jesus describes as the “Paraclete”, meaning the one called alongside.

As is well recognised, understanding remains difficult to analyse and its meaning and importance continue to be controversial. Nevertheless, this chapter has argued that there are certain plausible concomitant attributes and phenomena of understanding:

First, insights were experienced as occurring all at once in a fashion analogous to step changes and to yield both “higher” viewpoints and to be fruitful in the discovery of still further insights. Hence, insights are singular “ascending” operations.

Second, human insights are plausibly connected with and grounded in embodiment, involving the imagination, the senses, and things that are known implicitly rather than explicitly.

Third, metaphor, a word that means “to carry across”, seems to have some role in connecting embodied experience and language, a connection that assists in the communication of understanding.

Fourth, narratives, which can in some cases be regarded as metaphors extended over time leading to goals, are also a privileged means of communicating insights.

Fifth, second-person relatedness, arguably equivalent to joint attention, can be a catalyst for having and communicating insights.

With these attributes to serve as heuristic criteria, it is now possible to evaluate of the claims of the transhumanists in regard to the questions of insight.

CHAPTER 4 Evaluating Transhumanist Prospects for Enhancing Insight Directly

4.1 Introduction

Chapter two demonstrated that three prominent transhumanists – Marvin Minsky, Ray Kurzweil, and Nick Bostrom – all acknowledge the act of what they call “understanding” as being important for cognition, at least to some extent. In this chapter, I explore how their claims about the meaning of understanding compare with five attributes identified in chapter three, namely that acts of understanding called insights are singular and “ascending”, grounded in embodied experience, involve a privileged role played by metaphor, are communicated in a privileged way by narrative, and are often catalysed in the context of second-person relatedness.

4.2 Thematic Evaluations

4.2.1 A Singular “Ascending” Operation

The first theme to be addressed is that of insights as singular “ascending” operations. An insight is singular in the sense that it is a change in understanding that occurs all at once in an experience similar to a gestalt shift. Insights are also “ascending”, meaning that they have a progressive character. Insights lead to new perspectives (higher viewpoints, to use Lonergan’s turn of phrase) that make sense of more than was previously grasped. The cognitive shock of gaining this new perspective, and the reaction to this shock, is why insights are described as “Eureka!” experiences, a Greek word meaning “I have it!” What textual evidence might be found for characterising the transhumanist picture of insight vis-à-vis an operation of singular ascent?

One possibility is the way in which they address the question of how new ideas arise. On this topic, the work of Minsky, the teacher of Kurzweil, warrants especially careful appraisal.

An overview of Minsky's attitude to insight and its cognates can be gained from the following text in *Society of Mind*,

I see no reason to believe that intelligence can exist apart from a highly organized body of knowledge, models, and processes. The habit of our culture has always been to suppose that intelligence resides in some separated crystalline element, call it *consciousness, apprehension, insight, gestalt*, or what you will, but this is merely to confound naming the problem with solving it. The problem-solving abilities of a highly intelligent person lies partly in his superior heuristics for managing his knowledge structure and partly in the structure itself; these are probably somewhat inseparable. In any case, there is no reason to suppose that you can be intelligent except through the use of an adequate, *particular* knowledge or model structure.²⁴⁰

This key text draws attention to the following. First, Minsky implies that the true locus of intelligence is to be found in a “highly organised body of knowledge, models, and processes.” Second, he appears to denigrate the notion of what he calls a “separated crystalline element”, described as “*consciousness, apprehension, insight, gestalt ...*” Third, he associates high intelligence with “superior heuristics for managing [a] knowledge structure” and “the structure [of knowledge] itself.” Fourth, he reinforces this point by, in effect, stating that all intelligence requires “adequate, *particular* knowledge” or “model structure”.

From this passage, one therefore learns that Minsky denies a special faculty of insight, a notion that he dismisses as “some separated crystalline element” or otherwise magical faculty, the distinction of which is a “habit of our culture.” Indeed, elsewhere he makes this claim more explicitly as follows,

²⁴⁰ Minsky, *Semantic Information Processing*, 27, emphasis original.

We often use words like “insight” or “intuition” to talk about understandings that seem especially immediate. But it is bad psychology to assume that what seems “obvious” is therefore simple or self-evident. Many such things are done for us by huge, silent systems in our mind, built over long forgotten years of childhood.²⁴¹

Minsky acknowledges the apparent immediacy of understanding that is described as insight, but he claims that it is “bad psychology,” or otherwise methodologically unsound, to assume that those things about which insight occurs are themselves simple or, by implication, known simply. Instead, these understandings proceed from the operation of systems that were built up over one’s life, but the construction of which has been forgotten. In other words, Minsky concludes that the notion of a distinct intellectual faculty called insight is illusory without denying the subjective experience.

This conclusion raises the problem of what is going on with the subjective human experience of insight. Given that, on Minsky’s account, all intelligence is about “adequate, *particular* knowledge” or “model structure,” it is in these matters that an answer must be sought. Given that the subjective experience of insight at the very least involves cognitive change but does not seem to involve changes in “adequate, *particular* knowledge”, it is to be expected that Minsky locates phenomena associated with insight in terms of changes in model structures.

This expectation is fulfilled in his claims about what he calls “reformulation”. As noted in the second chapter (P. 26), Minsky does not acknowledge the actual existence of a distinct cognitive faculty of insight, but he does think that the process of describing known facts in new ways can be fruitful in generating what are commonly called insights:

...Reformulation is clearly very powerful—but how does one do it? How do people find new styles of descriptions that make their problems seem easier? Does this depend upon some mysterious kind of insight or upon some magically creative gift—or do we simply come upon them by accident? As I said when discussing creativity, these seem to me mere

²⁴¹ Minsky, “Society of Mind,” 128, emphasis original.

matters of degree, since people are always making reformulations of various sorts. Even when we contemplate those rarest and most revolutionary new ideas that come like revelations, suddenly to shed new light on entire fields of thought—like evolution, gravity, or relativity—we usually see by hindsight that these were variants of things that people knew before that time.²⁴²

Minsky implies in this passage that what is commonly called “insight” is neither mysterious, nor magically creative, nor reducible to mere accident. Instead, he sees great value in changing the styles of description of presumably known facts. As examples, he cites the cognitive revolutions of the theories of evolution, gravity, and relativity, all of which, he claims, were “variants of things that people knew before that time.” He calls these variations “reformulations” between a previous frame and a new frame, as described in chapter two.

In what sense are these reformulations singular, “ascending” operations? On one hand, a change of frame does, in fact, appear to be a step change between an old and a new formulation of a set of particulars. Moreover, a frame is, in itself, some cognitive structure that unifies many particulars and, based on the examples he chooses, reformulation appears to represent cognitive advancement. On Minsky’s account, reformulations may indeed be experienced as singular, ascending operations, consistent with philosophical and theological accounts of insights. Nevertheless, also on this account, as noted previously, there is no distinct faculty of insight and what is really going on in their experience can be reduced to mere changes of representation.

On this topic, Kurzweil and Bostrom have less to say, although Kurzweil was favourably disposed to Minsky’s interpretations of insights and representational change. Kurzweil endorsed a similar methodology in his own writing. Minsky and Kurzweil did differ in the particular processes and structures they hypothesized, but they shared the assumption that

²⁴² Minsky, “Society of Mind,” 133.

multiple levels of reductive description would be needed in psychology and in any description of insight.²⁴³

In a more general sense, Kurzweil did express what appears to be some sense that insights have a singular ascending character.

...increasing order usually means increasing complexity, but sometimes a profound insight will increase order while reducing complexity.²⁴⁴

In this case, profound insights are said to reduce complexity but increase order. Important insights have the effect of simplifying knowledge while also increasing it. He does not have a great deal more to say, however, about the particularities of insights as acts that lead to higher viewpoints or the relationship between them and changes in representation.

As regards Bostrom, there are at least some instances in which he takes a similar approach to Minsky on the topic of understanding,

Cognitive enhancement is the amplification or extension of core capacities of the mind through improvement or augmentation of internal or external *information processing systems*. Cognition refers to the processes an organism uses to organize information. These include acquiring (perception), selecting (attention), representing (understanding) and retaining (memory) information, and using this information to guide behaviour (reasoning and coordination of motor outputs). Interventions to improve cognitive function may be directed at any of these core faculties.²⁴⁵

In this text, on Bostrom's account cognition is "the process an organism uses to organise information", and an understanding is equivalent to an act of representation. In this case, a change of understanding involves a new and presumably improved representation, which is comparable to what Minsky calls reformulation.

²⁴³ Kurzweil, *The Age of Intelligent Machines*, 145–46.

²⁴⁴ Kurzweil, *The Singularity Is Near*, 653.

²⁴⁵ Nick Bostrom and Rebecca Roache, "Smart Policy: Cognitive Enhancement and the Public Interest," in *Enhancing Human Capacities*, 1st ed., ed. Julian Savulescu et al. (Wiley, 2011), 1, <https://doi.org/10.1002/9781444393552.ch9>, emphasis added.

The terminology of the following passage is also striking.

Much of what we learn in school is “mental software” for managing various cognitive domains: mathematics, categories of concepts, language, and problem solving in particular subjects. This kind of mental software reduces our mental load through clever encoding, organization, or processing. Instead of memorizing arbitrarily large multiplication tables we compress the pattern of arithmetic relationships into simpler rules of multiplication, which in turn (among very ambitious students) can be organized into efficient mental calculation methods like the Trachtenberg system (Trachtenberg 2000). Such specific methods have a smaller range of applicability but can dramatically improve performance within a particular domain. They represent a form of crystallized intelligence, distinct from the fluid intelligence of general cognitive abilities and problem solving capacity (Cattell 1987). The relative ease and utility of improving crystallized intelligence and specific abilities have made them popular targets of internal and external software development. Cognitive enhancement attempts the more difficult challenge of improving fluid intelligence.²⁴⁶

In this passage, Bostrom draws a sharp distinction between what he calls “crystallized” and “fluid” intelligence. Crystallized intelligence involves the application, often unconscious, of all kinds of clear and explicit rules, but it is notable that he describes so-called “fluid intelligence” as enabling problem solving. Although he does not refer to insight explicitly, the implication is that insights involve a somewhat vague kind of fluid intelligence that enables, perhaps, the establishment of what can become new crystallized intelligence. This account is not dissimilar to the notion of a singular, ascending operation.

Nevertheless, overall Bostrom remains somewhat vague on the notion of insight and his descriptions of it and related topics are at best implicit. Insofar as he addresses the topic at all, his descriptions do not diverge significantly from those of Minsky and Kurzweil.

In summary, the transhumanists all recognise at least the subjective experience of insights, of “eureka!” experiences, and consider the corresponding cognitive accomplishment

²⁴⁶ Nick Bostrom and Anders Sandberg, “Converging Cognitive Enhancements,” *Annals of the New York Academy of Sciences* 1093, no. 1 (2006): 208, <https://doi.org/10.1196/annals.1382.015>.

to be important. They do not, however, credit the existence of a distinct faculty of insight. Minsky and Kurzweil explicitly deny that such a faculty exists. Minsky and Kurzweil agree that intelligence is about representing facts and that the cognitive accomplishment corresponding to the use of the term “insight” is a reformulation of representations of facts. Bostrom does not explicitly deny that there is a special faculty of insight. He only speaks in terms of “crystallized intelligence”, “mental software” (which appears to correspond to representations and explicit techniques), and fluid intelligence (which seems to correspond to insight) as the means of producing new such representational structures. There is, at least, no claim that distinguishes him from the other transhumanists. The transhumanists, in any case, believe that insights are subjectively simple but objectively complex and are best characterised not as ascending to a higher viewpoint but methods of reformulating representations.

I note that claims about adequacy of this approach remain dubious, in particular the question of the ultimate grounding of the meaning of the representations involved in reformulations. No answer can be found simply in adding more representations, since there is a risk of an infinite regress. At the very least, regardless of the plausibility or otherwise of reformulations, it is necessary to examine the grounds of such representations.

4.2.2 Grounding in Embodied Experience

The second theme to discuss is the grounding of insight in embodied experience. The experience of the world occurs through the body and such experience is the raw material of cognition. Hence, understanding and insight are directly or indirectly embodied experience. In the case of transhumanism, there is a focus on the body as a site for modification and enhancement, so it bears examining the way the transhumanists speak of the role of the body for cognition and how it relates to their aspiration for cognitive enhancement.

With respect to the importance of embodiment, there are three points to discuss: the significance of having a body, the significance of having a body of a particular type, and the significance of having a particular body, my body rather than your body.

About the significance of having some kind of corporeality, it is not difficult to find passages from the works of the transhumanists that speak of the significance of some kind of corporeality. Bostrom makes a general observation that the body is needed for human cognition to function correctly.

Sensory or motor deprivation appears to produce intellectual and perceptual deficits within a few days time...²⁴⁷

If the body cannot function, then there are negative consequences for the function of the mind. Bostrom recognises, at minimum, that embodiment of some sort does have cognitive significance. This is an abstract and negative picture of what the body offers for cognition, but the transhumanists also have more concrete and more particular observations about the cognitive significance of embodiment.

Kurzweil provides a more positive example of the significance of embodiment in his book *The Age of Spiritual Machines*. Kurzweil often uses the rhetorical device of fictional dialogues between himself and imaginary characters to present what he thinks the experience of enhancement technologies will be like for the future. In one such passage he considers the utterances and concerns of an ordinary, unenhanced human, Molly.

Most of Molly's mental activities are directed toward her body and its survival, security, nutrition, image, not to mention related issues of affection, sexuality, and reproduction. But Molly is not unique in this regard. I invite my other readers to look at their own diaries. And if you don't have one, consider what you would write in it if you did. How many of your entries would make sense if you didn't have a body?²⁴⁸

²⁴⁷ Bostrom and Sandberg, *Whole Brain Emulation*, 30.

²⁴⁸ Ray Kurzweil, *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines* (Orion Business, 1999), 134.

Kurzweil's claim about the significance of the body for cognition is a strong one. The body is that toward which most mental activities are directed, presumably as a result of humans being social organisms who must tend their physical needs and maintain relations with other bodies. More importantly, Kurzweil suggests that lacking a body would make even something as prosaic as a description of a woman's ordinary daily life senseless.

The transhumanists recognise, at minimum, that the body motivates many cognitive acts and that much of everyday thinking is expressed in terms of the body. What do the transhumanists claim is the significance of the body for acts of understanding and of insight?

Minsky gives some sense of how embodiment relates to understanding and insight in his schema when discussing how particular objects are known. It will be recalled that in Minsky's lexicon something is understood when it is enframed, with a proper description applied, and what are called insights are successful reformulations of descriptions. Embodiment is crucial to such formulation and reformulation of descriptions of everyday objects, for instance chairs.

On one side, we need structural descriptions for recognizing chairs when we see them. On the other side, we need functional descriptions in order to know what we can *do* with the chairs... such knowledge about relations between structure, comfort, and posture could be used to understand when a box could serve as a chair... To be sure, such clever reasoning requires special mental skills with which to redescribe or "reformulate" the descriptions of both box and chair so that they "match" despite their differences. Until we learn to make old descriptions fit new circumstances, our old knowledge can be applied only to the circumstances in which it was learned.²⁴⁹

In Minsky's schema, the concept of a chair does not consist of only descriptions of its geometry but also of how it rests with four legs on the ground and how a human being sits upright upon it. What sort of descriptions could be written about the functionality of a chair to the body of a

²⁴⁹ Minsky, *The Society of Mind*, 123.

fish, for instance? Fish don't exist in an environment where there is a need to enter a posture of rest for one's limbs after holding oneself up from the ground or even the anatomy necessary to use a chair in such a way. Consider the even more extreme case of an amoeba, for which the chair is not an object in the environment but instead the vast and varied landscape of the environment itself. The propositions about chairs that Minsky discusses only make sense with reference to human bodies interacting with and using chairs. Moreover, the capacity to have the insight that two bodies in the world are in some way the same (in his example the recognition that a box can serve as a chair) is a result of special mental skills that allow the description of one object to be reformulated in terms of another. Insight is still a matter of reformulating descriptions, but one needs embodiment to be able to formulate these descriptions.²⁵⁰

Minsky makes an even stronger observation about the role played by the body when discussing the learning of spatial concepts:

We acquire our conceptions of space by using agencies that learn in accord with processes determined by inheritance. These agencies proceed to learn from experience—but the outcomes of their learning processes are virtually predestined by the spatial geometry of our body parts.²⁵¹

Whatever role biology and genetics play in the development of what Minsky calls the learning agencies the body exerts such influence on the concepts of space that emerge that gaining these concepts can be said to be predestined. In other words, Minsky thinks having a particular kind of body determines arriving at certain kinds of concepts about the world and how to represent it. One might say that the body partially but strongly determines the insights into space a human has and the kind of understandings at which they terminate.

²⁵⁰ Additionally, Minsky wrote appreciatively of the fiction of the transhumanist Vernor Vinge, especially his short story "True Names", which depicts human interfacing with computers through imagery and gestures in a kind of virtual reality. Minsky wrote a foreword to this story in which he expressed his hope that interfacing with computers would one day be an intuitive matter as easy as perceiving the world and responding with gesture.

²⁵¹ Minsky, *The Society of Mind*, 113.

To summarise Minsky's perspective, the body strongly shapes the kind of representations that constitutes an understanding of things. The body is for this reason important, but the strong influence the body plays on formulating these representations also has an implied limitation: one cannot but represent things with respect to the human body. Representations based on other kinds of bodily experience are inaccessible.

There is an additional subtlety, for Minsky specifically states that it is the spatial geometry of the body parts that determine the concepts of space that one learns. This spatial geometry might be a consequence of biology, but this level of description is still to an extent abstracted from biology. This raises the question of whether, on the transhumanist account, the cognitive significance of a particular kind of body is separable from the particular body itself.

With respect to the significance of corporeality, however, the transhumanists do not restrict the notion of corporeality solely to the realm of the physical corporeality. They also speak of the representation or abstraction of the body in virtual space. This would be done through creating virtual experiences for a simulated brain or other kind of thinking program.²⁵² The body is therefore "detachable" insofar as a suitable prosthetic or simulation supplies the needed sensation.

Kurzweil expresses high hopes for the possibility of virtual embodiment satisfying the needs of physical embodiment.

Ultimately, your experience would be highly realistic, just like being in the real world.²⁵³

Ultimately, Kurzweil predicts that the virtual experience would become indistinguishable from being in the real world, at least in terms of the ability of the virtual world to represent the

²⁵² Bostrom and Sandberg, *Whole Brain Emulation*; Kurzweil, *The Age of Spiritual Machines* (2001); Kurzweil, *The Singularity Is Near*.

²⁵³ Kurzweil, *The Age of Spiritual Machines* (1999), 144.

complexities of the real world. He further suggests that virtual reality will eventually be considered just as real and the sense of embodiment in it just as cognitively significant as that of the physical world.

Nick Bostrom's ambivalence about the importance of a particular kind of body is expressed in more detail elsewhere.

Some posthumans may find it advantageous to jettison their bodies altogether and live as information patterns on vast super-fast computer networks. Their minds may be not only more powerful than ours but may also employ different cognitive architectures or include new sensory modalities that enable greater participation in their virtual reality settings.²⁵⁴

The body may be something that needs to be jettisoned, like the booster of a rocket, in favour of existing as pure information and having a purely virtual form of embodiment. These virtual bodies are imagined as capacities for sensory experience beyond those of humans with the effect of "greater participation" in the virtual world. In other words, Bostrom proposes that one's own body may be discarded in favour of interactions with virtual representations and having one's embodiment be enhanced.

A key prerequisite of transhumanist aspirations is therefore revealed, namely that all aspects of the world or the body that are relevant to cognition can be fully represented. Bostrom articulates the degrees of success he describes for this project.

One can again make the distinction between an adequate environment simulation and a convincing simulation. An adequate environment produces enough input to activate the brain emulation and allow it to interact in such a way that its state and function can be evaluated. A convincing simulation is close enough to reality that the kinds of signals and interaction that occurs is hard (or impossible) for the emulated being to distinguish from reality.²⁵⁵

²⁵⁴ Bostrom, "Transhumanist FAQ."

²⁵⁵ Bostrom and Sandberg, *Whole Brain Emulation*, 76.

An adequate simulation is one that merely allows the mind to function while a convincing one approaches being indistinguishable from the real world. The purpose in the first case is to avoid the problem Bostrom highlighted about the detrimental effects of sensory and motor deprivation on the mind. What is useful about embodiment isn't really explored; there is just an awareness that too little embodied interaction with the world would be in some way suboptimal. What of the second goal, the creation of a simulation that is indistinguishable from the real world? Just how much realism would be required to achieve this is not clear, but there is reason to doubt that a simulation could achieve complete verisimilitude. The example of the three-body problem, for instance, indicates that many real-world systems cannot be perfectly modelled but only approximated. In some contexts, such as astronomical modelling, the error in the models can be kept low enough that the difference between predicted and observed orbital motion is negligible. In other contexts, such as modelling fluid dynamics, the error may be more significant when the chaotic motion is relevant to the phenomena one desires to model, such as neuronal activity. It is not yet clear to what extent the disjunct between descriptions and models of the world and the actual world would have an effect on humans exclusively interacting with such models, but the transhumanist aspirations require that models are not just tools to understand the world but rich enough to be proper objects of understanding themselves.

The aspiration to successfully creating a simulated world may be greater than is being fully acknowledged. The three-body problem, for instance, illustrates several points. The first is that there are real world systems that cannot be described with perfect accuracy but only with approximations. It may be that the claim that all of reality can be simulated perfectly is a false one. Further, it shows that predicting the outcome of chaotic systems is impossible since the model will contain slight variations on initial conditions that result in highly varying behaviour. Even if a simulation were, somehow, capable of modeling the physical world without

approximations, the model would necessarily deviate from the real world. Further, the difficulty can be extended to modelling the brain, which is far more complex than the gravitational interactions of three bodies and relies on, among other things, the chaotic motion of fluids. The transhumanist aspiration to abstract the physical operations of the brain into a model may be overly optimistic.

Given that the transhumanist notion of corporeality includes existence in a virtual space, a more subtle assumption follows, namely that bodies can in principle be interchangeable. There might be attachment to human morphologies due to the personal histories or ancestries of transhuman or posthuman entities, but there is no strict need to maintain even the broad structures of human experience and action except at a very abstract level (e.g., some forms of tactile and visual experience would possibly be preserved, but there would be no strict requirement to perceive a narrow bandwidth of light in the manner typical of a primate).

As observed earlier, Kurzweil acknowledges that many everyday human utterances and goals can only be understood by a being that has a human body, but his primary reflections on the relevance of the body to the cognitive processes are on the means of extending such experiences through direct and indirect means of corporeal enhancement.²⁵⁶ He speaks at great length about the opportunities afforded by coming advances in virtual reality and augmented reality. Throughout *The Age of Spiritual Machines*, he speculates that with sufficiently advanced virtual reality and augmented reality interfaces will create opportunities for immediately practical ends such as training surgeons and more open-ended prospects for freedom to gain otherwise inaccessible experience of the world including: the ability to experience the world with nonhuman morphology, the freedom to be present or absent from

²⁵⁶ Ray Kurzweil, *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines* (London: Orion Business, 1999), 134.

locations at will, or the ability to directly perceive phenomena outside the reach of ordinary human sensory capacity.²⁵⁷

One attribute I envision for version 3.0 [of the human body] is the ability to change our bodies. We'll be able to do that very easily in virtual-reality environments... but we will also acquire the means to do this in real reality. We will incorporate MNT-based [molecular nanotechnology-based] fabrication into ourselves, so we'll be able to rapidly alter our physical manifestation at will.

Kurzweil here imagines that in the future a radically redesigned human body will be characterised by tremendous plasticity being capable of being reshaped to allow new experiences and interaction with the world. This expresses a key theme of Kurzweil's work. The convergence between virtual and physical embodiment goes both ways. As virtual reality will eventually become sufficiently high fidelity to become experientially indistinguishable from physical reality, so will physical reality become easily modifiable in a way similar to that of virtual reality. In both cases, embodiment will become fluid and subject to one's own determination. What's relevant here is that the importance of having a particular kind of body is to a large extent denied. The kind of embodiment one has may shape cognitive experience, but any number of morphologies would be perfectly conducive to understanding and insight.

Admittedly, his speculations about the possible extension of human experience do suggest possibilities which, if realised, would indeed influence the way we come to new understandings. One might, for instance, create a virtual reality in which one can circumvent the limitations of a human body and the speed of human consciousness in order to experience nuclear fusion from the perspective of being within a tokamak.

Even relatively crude computer simulations can facilitate the grasping of concepts (this is discussed in greater depth in chapter five), so it is plausible that more sophisticated

²⁵⁷ Kurzweil, *The Age of Spiritual Machines* (1999), 144, 148, 198, 230.

kinaesthetic simulations will allow humans to engage bodily with an object of learning and gain intellectual mastery. That it may be possible to augment human embodied experience technologically to facilitate new understanding is a further point of consonance.²⁵⁸

Nick Bostrom takes a more equivocal view of the importance of embodiment. In his “Letter from Utopia” Nick Bostrom asks what the bodies of highly augmented beings in a utopian future might be like.

What is Body in Utopia? Body is a pair of legs, a pair of arms, a trunk and a head, all made out of flesh. Or not, as the case may be.²⁵⁹

Future intelligences might cleave to a human shape (or at least the shape of a terrestrial vertebrate), but equally they might not. Bostrom implies the particular kind of body one has is somewhat indifferent.

Although many passages in transhumanist literature indicate that embodiment is important for cognition and can convey benefits, an examination of the transhumanist descriptions of embodiment reveals a concept of embodiment that does not depend upon the actual possession of a body as such. Relatedly, the literature denies the necessity of having a particular kind of body even regarding the particularity of the body to be limiting. As a consequence of the non-necessity of having a body in general or having a particular kind of body, the literature has a notion of embodiment according to which having one’s own particular body is not necessary for the cognitive benefits of embodiment.

²⁵⁸ Consider, the Christian perspective of the body as good and its perfection as something worthwhile, though subordinated to the intellectual good of grasping truth and the supernatural good of sanctifying grace, resonated with scientific research that healed and expanded human perceptual experience. For instance, Roger Bacon OFM (d. c. 1292) contributed to the development of optics in the west, and the first depiction of a pair of spectacles worn by a person is a portrait of the Dominican Cardinal and renowned biblical scholar Hugh of Saint-Cher, painted by Tommaso da Modena in 1352. The technology of lens-making developed beyond restoring full power of vision to those with weak eyes to allowing, in principle, everyone to have visual access to both microscopic and very distant sights, which led to a progression of insights in biology and astronomy. To approve of technologies that might afford new insights through allowing new experience is consonant with Christian tradition.

²⁵⁹ Bostrom, “Letter from Utopia”, emphasis original.

In the transhumanist literature, there are three broad modes of interchangeability according to transhumanist conceptions of embodiment. In the first case, physical forms are seen as interchangeable with one another. The most straightforward way in which this is discussed in the literature is in terms of replacing one body, in whole or in part, with an artificial one. This replacement might simply take the form of creating an identical clone after fully representing and recreating the cognitive state of the original (this is usually imagined as a means of gaining a soft immortality) or by creating a simulacrum of the original body out of different components (which assumes the ability not only to represent fully the cognitive state of the original but the ability to abstract cognition and identity from the kind of matter of which the original body is composed. In both cases, the particular body is something from which a particular mind may be abstracted, and in the second case, the type of matter and fine structure can be abstracted away from the kind of embodiment the trans- or posthuman has.²⁶⁰

Transhumanist literature considers an even more radical interchangeability between physical bodies. Human bodies represent one kind in the space of conceivable morphologies that could be compatible with a sapient being's experience of the world, and it is imagined that a mind that has been abstracted from one body could be realised in a body with radically different morphologies and senses.

²⁶⁰ Eberl describes the contrast between the Aristotelian and Transhumanist anthropology's view of the related question of the relationship between one's identity and one's embodiment in the following way: "An Aristotelian account, on the other hand, understands a human person to exist as a unified substance that is a living, sentient, and rational animal: one's persistent existence as the numerically same person thereby includes the persistence of the same living animal body... The Lockean view allows for Transhumanists to envision a future in which a person may be able to have their cerebrum or entire head removed from their dying body and transplanted into/onto a younger and healthier body, transform their organic body into a more advanced cyborg form, or shed their body altogether by having their consciousness uploaded into a cybernetic network. By contrast, an Aristotelian anthropology would not appear to allow for one's personal survival in a total or near-total cyborg transformation or cybernetic uploading scenario due to one's essential organic embodiment." (In-line citations have been removed from this quotation. For the full bibliography, refer to Eberl's article.) Jason T. Eberl, "Who Wants to Live Forever? Transhumanist Immortality and Christian Eternity," *Christian Bioethics*, ahead of print, 2025, 83, <https://doi.org/10.1093/cb/cbaf004>.

The transhumanist literature has a particular focus on the possibility of interchanging the physical body with a virtual representation. Such an interchangeability is assumed in the literature on brain uploading, which the transhumanists discuss as a possible means of overcoming the limitations of cognition in a particular body. The assumption is that the mind can not only be abstracted from the matter of a particular body, but that a mind can be fully represented and instantiated in any matter capable of the requisite computations. Literal embodiment and virtual embodiment would then become indistinguishable.

Still, a purely abstract set of concepts could in principle exist and would be equally valid for the purpose of understanding on Minskian grounds. Lakoff and Johnson speak of the role of experience in providing the material, which is shaped to allow abstract concepts (linguistic, logical, or mathematical), but for Minsky the entirety of one's understanding can be contained in a system of representations. From this follows the principle that understanding may be completely abstracted from experience. Experience may allow a human to develop various concepts, but once the system has been constructed, reference to one's bodily experience is no longer necessary.

The peculiarity of Minsky's account of embodiment as significant due to the descriptions it engenders can be highlighted by comparison with the famous Mary in the Black and White Room thought experiment. The experiment imagines a scientist named Mary has been raised the entirety of her life in a room that is black and white and contains only black and white objects. She has never perceived colour, yet she has become an expert on optics and the physiology and neuroscience of colour perception. She has a thorough scientific description of colour, but she has never herself perceived colour. One day she leaves her black and white room

and sees the colour for the first time.²⁶¹ Intuitively, it seems that Mary has gained understanding from this experience that she lacked previously.

Mary, prior to experiencing red, would seem satisfy the conditions of having understanding of red on Minsky's terms. After all, she can describe the physical phenomena with exactitude and can also fully describe the operations of the body when experiencing red. She cannot increase in her ability to describe red, so it would seem that in Minsky's schema there is nothing new she can understand about red. Yet, intuitively perceiving red allows her to gain new understanding of red without any changes to her descriptions of what red is. This suggests that Minsky's account of embodiment ignores the importance of experience in favour of that of abstraction.

The consonances and dissonances between the philosophical and theological perspectives may be brought into sharper relief by making reference to Thomas Aquinas' writing on the human intellect's dependence on embodiment. There is an initial apparent consonance between Bostrom's observation that cognitive decline results from sensory deprivation and Aquinas observation that without the sensory power, the human intellect cannot

²⁶¹ Frank Jackson, "Epiphenomenal Qualia," *The Philosophical Quarterly* (1950-) 32, no. 127 (1982): 130, <https://doi.org/10.2307/2960077>.

function.²⁶² Nonetheless, when Aquinas asserts that intellectual cognition is derived from the senses,²⁶³ his meaning is significantly different than that of the transhumanists.

Aquinas writes of the intimate cooperation between the intellect and what he calls the interior senses.²⁶⁴ An example, as stated in chapter three, is the imagination, a corporeal power located in the brain.²⁶⁵

In the state of our present life, in which our intellect is joined to a passible body, it is impossible for it to have an actual intellectual understanding of anything unless it turns itself toward phantasms... But the senses, the imagination, and all the other powers that belong to the sentient part of the soul use a corporeal organ. Hence, it is clear that in order for the intellect to be actually engaged in intellectual understanding—not only for attaining knowledge *de novo*, but also for making use of already acquired knowledge—what is required are acts of the imagination and of the rest of the powers. For we see that when the act of the power of imagining is impeded by an injury to an organ, as in the case of those who are delirious

²⁶² *ST I-I* q. 84, a. 8. “As has been explained (a. 7), the proper object proportioned to our intellect is the nature of a sensible thing. Now a perfect judgment cannot be made about a thing if not everything that pertains to that thing is known, and especially if the terminus or end of the judgment is unknown. But in *De Caelo* 3 the Philosopher says, ‘Just as the end of productive knowledge (*factiva scientia*) is a piece of work (*opus*), so the end of natural knowledge (*naturalis scientia*) is that which is apparent principally to the senses.’ For a craftsman seeks to have cognition of knives only for the sake of his work, so that he might make this particular knife; and, similarly, a natural scientist seeks to have cognition of the nature of a rock or of a horse only in order to know the natures of those things that are apparent to the senses. Now it is obvious that the craftsman could not render a perfect judgment about a knife if he did not know how to make one (*si opus ignoraret*); and, similarly, no perfect judgment of natural knowledge can be made about natural things if sensible things are not known. But all the things we have intellectual understanding of in our present state are such that we have cognition of them in relation to natural sensible things. Hence, it is impossible for a perfect judgment of the intellect to exist in us when the sensory power is inoperative, since it is through the sensory power that we have cognition of sensible things.”

²⁶³ *ST I-I* q. 84, a. 6.

²⁶⁴ These four interior senses are the common sensory power (*sensus communis*), the imagination (*vis imaginativa*), the estimative power (*vis aestimativa*), and the sense-memory (*vis memorativa*). The common sensory power unifies the sensory inputs. It is how the animal has awareness of and reacts to the environment and what are today called affordances, so that it can react to them. The imagination is that which produces the image in the mind (this should be understood to embrace all sensory content) and which is both persistent in that the image can endure past the cessation of sensation and plastic in that it may be changed, even voluntarily. The estimative power apprehends intentions that aren’t provided by the exterior senses (for instance, the bear might not be actively injuring the wolf, but the wolf still perceives it as an enemy and threat, or a hermit crab perceives a shell as suitable for protection). The memorative power is a store of the such intentions rather as the imagination is a store of the past sensations of the common sensory power. In other animals, the perception of intentions is largely instinctive; in humans there is a much larger cognitive component. Aquinas, *ST I-I* q. 78, a. 4.

²⁶⁵ In his *Sentences Commentary*, Book 2, d. 20, q. 2, a. 2 ad 4, Aquinas explains that the intellect does not receive directly from the exterior senses, but from the interior. Hence, in young children the operation of the intellect is confused because “the imagination is confused on account of the excessive humidity of the brain.” In *Summa Theologiae* I q. 78, a. 4, he explains that different degrees of humidity of certain bodily organs makes them apt either to receive or to retain sense impressions.

(*in phreneticis*) or, similarly, when the act of the power of remembering is impeded, as in those who are groggy (*in lethargicis*), a man is prevented from having actual intellectual understanding even of those things that he previously had scientific knowledge of... Second, everyone can experience in his own case that when someone tries to have an intellectual understanding of something, he forms for himself phantasms as examples in which he inspects, as it were, what he is trying to understand. And so, too, when we want to make someone else understand something, we propose to him examples from which he can form phantasms in order to understand the matter at hand. The reason for this is that a cognitive power is proportioned to the thing it has cognition of... But we apprehend particulars through the sensory power and the imagination. And so for the intellect to have an actual intellectual understanding of its own proper object, it is necessary that it turn itself to phantasms, in order that it might inspect the universal nature as it exists in the particular (*ut speculetur naturam universalem in particulari existentem*)...²⁶⁶

Aquinas states that it is impossible for the intellect actively to understand without the operation of the imagination.²⁶⁷ The proof of this, for Aquinas, is that when the power of imagination is disrupted, say from sickness or exhaustion, actualizing one's understanding is impossible even when one had previously demonstrated systematic mastery of some area of knowledge. Further, Aquinas observes that whenever one is trying to understand something new, one creates mental images and tries to understand them. Similarly, sharing one's understanding is basically a matter of sharing imaginations. While God both gives the created intellect its power, and can imprint concepts (*species intelligibiles*) on it, even angels cannot put concepts directly into the human intellect. The human teacher, too, can assist, but from the outside, the intellect's intrinsic "light," by which the human being can apply universal principles to the particulars of memory and experience. The teacher can propose less universal propositions whose truth the pupil can recognise on the basis of what he or she already knows, and can offer sense-perceptible examples; the teacher can also strengthen the pupil's intellect by suggesting ways in which

²⁶⁶ *ST I-I* q. 84, a. 7.

²⁶⁷ Recall Lonergan's assertion that there is no natural act of understanding, not even of mathematical abstractions, that is not accompanied with some image which will often be a metaphor and the supporting evidence from Davis and Hersh that professional mathematicians experience the meaning of mathematical expressions kinaesthetically.

principles lead to conclusions, when the pupil is not able to formulate the syllogism independently.²⁶⁸ In other words, humans share understanding through shared attention.

Humans require sensory experience to have both practical and theoretical understanding, and incomplete or inadequate sensory experience results in incompleteness to that understanding. Moreover, for Aquinas the senses and the imagination allow the intellect to know the particular in the world, and through it know the universal. This is not a processing algorithm functioning with correct input but an actual contact between the mind and reality through the body. This is a more fundamental relation than that of abstraction or model-making.

The body is not just an accessory or prosthesis that is fitted to the human mind. The philosophical perspective, as represented by Aquinas, sees the human as a unified entity. The emphasis on the unity of humans as an intellectual organism may be seen by its contrast to the trend of anglophone philosophy to see human intelligence as a capacity for explicit and discursive reasoning siloed from animality.

Contemporary Anglophone philosophy needs an alternative to the view of human beings as brutes with thinking caps. The analogous challenge Aquinas faced was developing an alternative to the view that humans are souls wearing their bodies like garments. In both cases, what is at stake is the unity of the human being... Aquinas develops a unified account of the intellectual animal.²⁶⁹

This view of the grounding of human understanding by embodiment stands in contrast to both the perspective that human reason is reckoning divorced from embodied experience and the interlocutors of Aquinas who thought the intellectual soul was separate from the body. In both cases, the body was detachable and ultimately superfluous to the operation of the rational mind.

²⁶⁸ *ST I-I* q. 105, a. 3; q. 111, a. 1; q. 117, a. 1.

²⁶⁹ Candace Vogler, "The Intellectual Animal," *New Blackfriars* 100, no. 1090 (2019): 665, <https://doi.org/10.1111/nbfr.12503>.

By contrast, Aquinas sees the body as essential to the operation of the human intellect such that the detachment of the body and intellect in death is unnatural to the human being.

One kind of animal—the kind whose soul is intellectual, the one kind of animal whose soul *qua* intellectual can survive without its body, albeit in a radically impoverished and ultimately unsustainable way—integrates the higher cognitive powers with the operations of sentience. That animal is the human being.²⁷⁰

Even if the intellect can endure apart from the body, as it can in Aquinas' metaphysic,²⁷¹ it does not allow the normal functioning of human mind.²⁷²

In contrast to the transhumanist picture of the human as an aggregate of parts and powers, the human being is one entity.

What Thompson calls a “life form” and treats by way of careful consideration of the kinds of judgment and understanding that rely upon grasp of forms of life, Aquinas will treat as soul — a topic for metaphysics. The soul is the substantial form of a living thing. Following Aristotle, Aquinas gives an account of soul as the principle of unity for a living thing. As such, the soul makes an individual living thing the one thing that it is. If the living thing has parts, then the soul is responsible for both the diversity of its parts and for their order and functioning. If the living thing is an organism, then its body is made to be an individual body by the soul. As what unifies the individual and orders and organizes its activity, any living thing's substantial form is operative in the whole of that living thing throughout the living thing's life.²⁷³

In Aquinas' system, the soul is the principle of unity in an organism responsible for the diversity of parts, the order of these parts, and their functioning. The individual substance is unified, ordered, and organised in its activity by the soul, and the soul is operative in the whole of the body. This has the corollary that since human beings have an intellectual soul, the intellectual

²⁷⁰ Vogler, “The Intellectual Animal,” 673.

²⁷¹ *ST I* q. 75, aa. 2 and 6.

²⁷² In *ST I-I* q. 89, Aquinas does discuss the cognition of the separated soul. It is beyond my current scope to critique what Aquinas says, except to note that, by its natural powers, the separated soul is unlikely to be adept at the forms of cognition Aquinas recognises, and that the Beatific Vision does not rely on, and is not proportioned to, natural powers.

²⁷³ Vogler, “The Intellectual Animal,” 668.

soul is operative in the whole of the body. The whole of the body is involved in the operation of the intellectual soul, so the whole of the body is relevant to the operation of the intellect. To put it another way, the human is intellectual in an animal way and animal in an intellectual way.

Aquinas' theology of the resurrection in fact sharpens this point of the unity of the intellect and the body, for even gifted with God's own knowledge, the flourishing of the human being requires the restoration of the body so the intellect and the sentient powers to which it is united can operate according to their nature. The unity of the human being is strongly affirmed.

If we are talking about the sort of human beatitude that can be had in this life, then it is clear that a good disposition on the part of the body is necessarily required for it. For according to the Philosopher, this sort of beatitude consists in the operation of perfect virtue. But it is obvious that a man can be impeded in every sort of virtuous operation by bodily infirmities (*per invaletudinem corporis*). On the other hand, if we are talking about perfect beatitude, some have claimed that no particular disposition of the body is required for beatitude; in fact, they have claimed that what is required for beatitude is that the soul should be altogether separated from the body. Hence, in *De Civitate Dei* 22 Augustine quotes Porphyry as saying that "in order for the soul to be blessed, it must flee from all bodies." But this is wrong. For since it is natural for the soul to be united to its body, it cannot be the case that the soul's perfection excludes the body's natural perfection. And so one should reply that for beatitude that is perfect in every way, a perfect disposition on the part of the body is required both as an antecedent and as a consequence. It is required as an antecedent because, as Augustine says in *Super Genesim ad Litteram* 12, "If the body is such that it is difficult and burdensome to rule—e.g., flesh that is corrupting and burdening the soul—then the mind is turned away from the vision that belongs to the highest heaven." Hence, he concludes that since the body will be a spiritual body and no longer an animal body, a man will at that point be equal to the angels, and what was a burden to him will be to his glory. On the other hand, it is required as a consequence because there will be an overflow of the soul's beatitude into the body, so that the body itself will have its own perfection. Hence, in the letter *Ad Dioscorum* Augustine says, "God made the soul with such a potent nature that the vigor of incorruption overflows from its fullness of beatitude into the lower nature."²⁷⁴

²⁷⁴ *ST* I-II q. 4, a. 6.

On Aquinas' account, perfect bliss requires the body, but it must be a resurrected body that both possesses human life but is radically different from mortal life. In mortal life, happiness requires a natural perfection of the body because a disordered body is an impediment to the virtue in which natural happiness consists. Because human beings are embodied, the bliss of the beatific vision also requires the good of their bodies. Being without the body would be unnatural to human beings, so the perfection of the beatific vision cannot exclude the natural perfection of the body. Instead, the body will no longer be animal and corruptible (thereby liable to impede the virtues) but spiritual. The resurrected body will not be burdensome but a source of glory.

Further, the happiness of the soul partaking in the beatific vision will overflow into the body. The beatific vision, as Aquinas describes it, requires the resurrected body, which is different in kind from the mortal body and not achievable by any kind of enhancement. Among other things, even could the transhumanists achieve their ambition of producing a body capable of persisting until the heat death of the universe, it would not be incorruptible. It might be very good at keeping itself stable and dynamically repairing itself, but robustness to corruption and incorruptibility are not identical. Secondly, even the most robust hypothetical posthuman body would still eventually degrade and die, impeding the intellectual virtue needed for perfect bliss. The similarities between the resurrected body and any conceivable posthuman body are superficial. Further, on the Thomistic account, the body is good and will share in beatitude while on the transhumanist account the body is bad and must be jettisoned in favour of either a better body or no body at all.

On the transhumanist account, animality can be separated from humanity and human intelligence, but on the Thomistic account humans are essentially animals and are rational in an animal way. As Stephen Brock puts it in his book on Aquinas:

To say rational . . . is already, implicitly, to say animal . . . Rationality is a mode of intellect—of the quality of being, in a way, all things—and hence it implies a dimension of pure immateriality. But it is a mode that is intrinsically connected to the life of the senses, and therefore to the sense-organs; and to the vegetative functions and the physical ingredients that constitute the organs; and to matter itself.²⁷⁵

Intelligence implies both the pure immateriality of concepts and abstractions, but this mode of immateriality is also intimately connected to the senses, the organs that have them, to the vital functions of and physical components of the organs, and to matter. In effect, the view offered by Aquinas of humans as a unity that can be called an intelligent animal connects intelligence with the realm of pure abstraction and the realm of particularity and actuality from which abstractions are drawn.²⁷⁶ Embodiment grounds and is integral to understanding and insight.

In conclusion, the transhumanists do think that having embodiment, in some sense, confers cognitive benefit primarily because humanlike minds need perceptual and motoric interaction with an environment to function and shape their concepts based on these interactions. They explicitly affirm, however, that these benefits can be separated from one's actual body. Minsky and Kurzweil describe the body's importance for structuring concepts in terms of formulating descriptions, and it will be recalled that what is commonly called "understanding" is actually just the structuring of information in suitable representations. In Minsky and Kurzweil's thought, the body aids understanding in the sense that provides a locus for the representational structures that fully exhaust the cognitive content of "understanding".

²⁷⁵ Stephen L. Brock, *The Philosophy of Saint Thomas Aquinas: A Sketch*, 1st ed. (Cascade Books, 2015), 93; quoted in Vogler, "The Intellectual Animal," 675. See also John Finley, "Man More Animal than Anything: The Unity in Human Agency," *New Blackfriars* 100 (2019) pp. 677-697.

²⁷⁶ Given that all humans directly know of intelligence is their own experience of and acts of intelligence, it follows that all human conceptions of intelligence are either animal or by analogy to the animality of human intelligence. Daniel D. De Haan, "Approaching Other Animals with Caution: Exploring Insights from Aquinas's Psychology," *New Blackfriars* 100, no. 1090 (2019): 715–37, <https://doi.org/10.1111/nbfr.12509>.

In other words, there is nothing understood through embodiment other than or apart from representations of the body and the world.

The transhumanist stance on embodiment is evinced by the claim that, as Bostrom says, posthumans may choose to exist as information patterns in computing systems with virtual embodiment possibly richer than that of biological humans. In order for such a thing to be feasible, all the world and all the content of cognition must be fully representable, that is to say abstractable, without any remainder. This claim is counterintuitive, as illustrated by the Mary in the Black and White Room thought experiment, which seems to show one can have complete knowledge of descriptions and representations of what colour is and how it is perceived, yet still gain new understanding when perceiving colour for the first time. The claim that representation can be fully exhaustive of reality is in fact radical, as illustrated by the technical inability to perfectly model even simple chaotic systems like the three-body problem. Transhumanist claims about the deep interchangeability and malleability of the body ultimately are deeply bound with a radical metaphysic and deny that the body and bodily experience ultimately ground explicit utterance and representations ultimately allowing them to be understood as meaningful. In contrast, the philosophical perspective sees embodiment as unified to intelligence, so to insight and understanding, such that intelligence embraces both abstraction and implicit understanding of the particular precisely through the totality of one's embodiment.

4.2.3 Role of Metaphor

The third theme to discuss is the privileged role metaphor linking embodiment with language and facilitating insights. The transhumanist literature does show some sense of

metaphor being cognitively important, but, given the peculiar way in which transhumanists use the term embodiment, their discussion of metaphor warrants further examination.

Minsky characterises metaphors in the following way:

All our language is riddled and stitched with curious ways of portraying things as though they belonged to alien realms. We sometimes call these “metaphors,” our ways to transport thoughts between the various mental realms. Some metaphors seem utterly pedestrian, as when we speak of “taking steps” to cause or prevent some happening. Other metaphors seem more miraculous, when unexpected images lead to astonishing insights—as when a scientist solves a problem by conceiving of a fluid as made of tubes or a wave as an array of overlapping, expanding spheres.²⁷⁷

The first observation to make is that Minsky conceives of metaphor as “transporting” thoughts between mental realms. Within his idiom, this means to bring frames belonging to one realm of thought to bear upon another. For instance, the application of models developed in physics to economic activity would be an act of transport across mental realms by means of metaphor.²⁷⁸

The second observation to make is that Minsky treats the ability to find metaphors as enabling the generation of significant reformulations (i.e. insights). Such reformulations consist in the recognition of some commonality between different concepts that allows one to be illuminated by another.

Minsky states this link between metaphors and reformulation even more explicitly.

From where do we obtain our most productive, systematic, *cross-realm correspondences*? Some must be virtually born into our brains through the writing of our paranomes; other *metaphors* we discover by ourselves as individuals; but most of them are learned from other members of our cultural communities. Finally, from time to time, someone discovers a new *reformulation* that is both so fruitful and so easy to explain that it becomes part of the general culture.²⁷⁹

²⁷⁷ Minsky, *The Society of Mind*, 299.

²⁷⁸ See Mirowski for an account of the developments in economic theory making use of mathematical models developed for the physical sciences. Philip Mirowski, *Machine Dreams: Economics Becomes a Cyborg Science* (Cambridge; New York: Cambridge University Press, 2002).

²⁷⁹ Minsky, *The Society of Mind*, 299, emphasis added.

Minsky considers several possible ways that cross-realm correspondences are acquired. The most relevant portion of this passage here is not that Minsky's considers that most metaphors are shared within a community rather than innate or individually discovered (although this will be examined in section 4.2.5) but that metaphors are explicitly identified with the reformulations of description that Minsky identifies with insight.

Minsky thinks that metaphor can be labelled, but can it be described as something distinct?

What, then, is a metaphor? It might be easy to agree on functional definitions like "*A metaphor is that which allows us to replace one kind of thought with another.*" But when we ask for a structural definition of "metaphor," we find no unity, only an endless variety of processes and strategies. Some are simple, as when we make an analogy by stripping away so many details that two different objects seem the same. But other forms of metaphor are as complex as can be. In the end there is little to gain by cloaking them all under the same name "metaphor," because there isn't any boundary between metaphorical thought and ordinary thought.²⁸⁰

Minsky thinks that there is no structural unity to metaphor but that the sheer variety of kinds of metaphors makes providing a precise description of metaphor impossible. Like the terms "insight" and "understanding" Minsky regards "metaphor" to be a term of convenience that does not label any distinct faculty. Nonetheless, he does claim that what is called metaphorical thinking is so pervasive that it can be said that there is no distinction between it and ordinary thought. In other words, Minsky thinks that metaphor does have a privileged role with respect to insight, but he equally thinks that there isn't really a way to distinguish metaphors.

Although he uses different terminology and has a subtly different interpretation, Ray Kurzweil similarly remarks on the pervasive nature of metaphors.

²⁸⁰ Minsky, *The Society of Mind*, 297.

Finding a metaphor is the process of recognizing a pattern despite differences in detail and context—an activity we undertake trivially every moment of our lives.²⁸¹

Recognising the same pattern in different contexts and different details is similar to the Minskian notion that the same object can activate frames in different realms of thought, and this recognition is often fruitful for understanding.²⁸² Furthermore, to recognise a pattern in spite of differences in detail and context implies an ability to discard those details that are nonessential. More importantly, Kurzweil also describes metaphor as being a constant part of human mental activity, namely recognising patterns even when detail and context differs.

Finally, it should be noted that Bostrom does not discuss metaphor or its role in cognition in his corpus.

The transhumanist literature, then, has notable examples of engagement with the functional role of metaphor in acts of understanding and the generation of insight, but its account of metaphor nonetheless stands in contrast to the philosophical perspective on metaphor in both scope and grounding.²⁸³ According to the philosophical perspective, metaphor does not merely have function within a kind of algorithmic system. Metaphor carries across from language to meaning.²⁸⁴ Moreover, as discussed in chapter three, metaphor can be described not as some discrete process or algorithm but as a kind of kinaesthetic experience of similarity between things that not only results in new kinds of utterance but makes it possible to perceive the common grounds that makes it possible to refer to things in terms of other things. In other

²⁸¹ Kurzweil, *How to Create a Mind*, 121.

²⁸² Kurzweil reports having worked on many interdisciplinary projects in which the problems in one field could be solved by finding a suitable metaphor in another. This illustrates the ability to find metaphors can contribute to insights. Moreover, a functional equivalence may be seen between Kurzweil's account of using metaphors to bridge understanding between fields of study and Minsky's notion of using metaphors to transport between different realms of thought. Kurzweil, 123.

²⁸³ Metaphor here should be understood in the sense used by Ian McGilchrist as that which carries from linguistic representation to meaning.

²⁸⁴ McGilchrist, *The Master and His Emissary*, 115.

words, metaphors are distinct and recognisable in the same way that redness (to return again to Mary leaving her black and white room) is distinct and recognisable. It is the insistence that what is known must be exhausted by description that necessitates denying metaphors are distinct things.

Metaphor allows language to refer meaningfully to the world rather than to other symbols alone. Indeed, on McGilchrist's account as noted previously, metaphor allows a system of symbols to be about the lived world, pointing to actual or imagined experience rather than a mere definition in terms of other symbols. On this account, the definition of words in terms of other words, or more generally the definition of symbols in terms of other symbols, must eventually terminate in a pointing-toward embodied experience. A symbol system can be a representation by means of a pointing toward experience, which is accomplished by metaphor. Without the linking function of metaphor, a symbol system remains wholly abstract. Hence this account is obviously in conflict with that of the transhumanists, according to whom metaphor merely allows a connection between one set of symbols and another, with both sets constructed without reference or interpreted to have indeterminate or arbitrary reference.

The transhumanist account of metaphor may acknowledge that it has a function in allowing something to be expressed in terms of something else, but their account of embodiment does not afford metaphor the same grounding function that it does within the philosophical account. The philosophical account identifies a need for symbol systems to be grounded by a link with experience that must be recognised through acquaintance. The symbol user is apart from the system and using metaphor to give it meaning. In contrast, the transhumanists are committed to the claim that embodiment can itself be fully represented and therefore simulated. In other words, for the transhumanists metaphors are mappings between symbol sets, and they do not offer a solution to the problem of how to ground symbols in meaning.

Implicit to the transhumanist treatment of metaphor as exhaustible in terms of mappings between sets of symbols, is the assumption that all thought can be broken down into parts. The units of thought may be extremely numerous and arranged in a complex structure, but there exists some finite number of elements in the set. The philosophical approach offered here treats all representations as abstractions from the world that gain their meaning by pointing to something outside any finite set of components out of which they may be composed, and thought is taken to have as objects both representations and the objects of representations. In short, thought embraces both symbol systems and the world they are about.

The transhumanist perspective on metaphor, ultimately, stands in marked contrast to the philosophical perspective on metaphor. Where the philosophical perspective as represented by McGilchrist claims the power of metaphor to yield insights into unities between things is based on an embodied, synaesthetic experience of the similarity between things, the transhumanist account identifies the site of the discovery of unity to be at the level of identifying similar patterns in descriptions.

As cited earlier in this thesis (P. 93), McGilchrist writes that language activates connotations that makes things accessible to the whole of our being while at the same time words reach beyond themselves to the embodied experience of the world. There is nothing more fundamental than this experiential contact with the world, while the transhumanists suppose that, as discussed in the preceding section, the cognitive content of embodied experience is exhausted by representations. In other words, there is no gap between language (or representations more broadly) and the world that metaphor carries us across. Instead, metaphor is just about connotation and similarity between patterns of representation. To put it another way, McGilchrist identifies as the bedrock of meaning the kind of cognitive content that Mary points to when she leaves her black and white room as says, “So *that* is what red is,” as distinct

from all her explicit knowledge of the physics of colour and the neuroscience of perception. The transhumanist interpretation is that this foundation of embodied experience is as illusory as the distinct faculty of understanding. Hence, seemingly basic experiences are just composed of further representations outside of conscious experience. The transhumanists would deny, then, that metaphor consists primarily in experiencing a similarity between things that enables an articulation of these things. This need for everything to be fully representable is closely tied to their aspiration to enhancing cognition by reverse engineering intelligence in machines and uploading intelligence into computers. All that is computed must be represented as finite bit strings, which is to say that everything must be explicitly described. If there exists a fundamental gap between any explicit representations and the world that is represented, then their aspiration to fully describe and recreate a brain is put into question.

In conclusion, the transhumanists claim that metaphor has tremendous cognitive significance. They say that metaphor is pervasive in human thought, both in everyday cognition and acts of tremendous insight. The transhumanist and philosophical perspective, then, can be said to agree that metaphor is fundamental to understanding and deeply involved in metaphor. Nonetheless, they disagree strongly about what metaphor is. The transhumanist perspective interprets metaphor in terms of pattern matching between representations. While the philosophical perspective sees metaphor as grounded in an experience of similarity that is prior to an explicit articulation of similarity, the transhumanist perspective must deny that there is any understanding that is not exhausted by representation, hence an account of metaphor must also be in terms of representations that can, in principle, be fully modelled and enumerated.

Thus far, this chapter has examined the transhumanist perspectives on understanding, insight, embodiment, and metaphor. While the transhumanist perspective has sought to interpret them all in terms of explicit representations with no remainder of cognitive content, the

philosophical perspective instead maintains that there is a gap between experience and representation bridged by metaphor in which understanding operates. One direction for further inquiry is to explore a phenomenon that expands upon metaphor. Narrative is arguably a kind of metaphor extended through time, since narrative links a sequence of events into a kind of whole. An internal coherence is recognised in a story through the imagination in a way that allows one's whole being to engage with it. Experiences are shared through the medium of words that evoke the embodied sensations, actions, and emotions. What's more, if metaphor is used to express understanding and novel metaphors lead to new insights, then it would be expected that narratives, as metaphors extended through time, have an important, even privileged, role to play with respect to understanding and insight. That narrative has such a role was defended in chapter three. Having explored how the transhumanists interpret understanding, insight, embodiment, and metaphor in terms of explicit representations, it is worthwhile to explore how the transhumanists extend their account in terms of narrative.

4.2.4 Role of Narrative

What role, if any, do transhumanists assign to narrative in order to accelerate understanding? The question can be addressed directly, in terms of what they state about this role. The question can also be addressed indirectly, by considering the uses to which at least some of transhumanists put narrative in their own extensive works.

In the case of Minsky, he acknowledges that narratives are cognitively important for making sense of phenomena.

Perhaps our most familiar way to represent an incident is to recount it as a story or script that depicts a sequence of events in time—that is, in the form of a story or a narrative.²⁸⁵

Minsky describes representing incidents as stories or scripts to be possibly the most familiar kind of representation. This is at least an implicit acknowledgement that narrative is cognitively significant because narratives are a regular and recognisable part of cognition. Moreover, Minsky describes stories and scripts as depicted sequences of events in time. He doesn't provide further differentiating criteria, which means that his sense of narrative is arguably wider in scope than the commonsense meaning of the term. Even something as simple as a sequence of zeros and ones representing the outcome of a series of coinflips would seem to be sufficient for being at least a simple kind of narrative by this description. In his schema, narratives do typically have some additional structure, however.

Narrative thinking, according to Minsky, is only in a secondary sense about the written or spoken word. Narrative is primarily a means by which the expected is understood implicitly, so that those things which are important and suggestive of something novel can be given attention.

Whenever somebody tells you a story, you react less to what each separate sentence means than to how this differs from what you expected—and this also applies to our other perceptions.²⁸⁶

Functionally, for Minsky narrative thinking is what allows understanding of familiar situations and characters to occur in the background of one's cognition while highlighting those matters that require active attention to understand. Recurrent events and actions in the world are understood by and as narrative structures.

The structures are formed

²⁸⁵ Minsky, *The Emotion Machine*, 281.

²⁸⁶ Minsky, *The Emotion Machine*, 195.

...by grouping collections of smaller events into multiple levels of events, episodes, sections, and plots.²⁸⁷

Certain levels of detail or repetition are excised from the narrative at higher levels of grouping. Minsky uses the example of progressions through rhythms and themes in a movement of music, but the point may also be illustrated by narratives about human characters. The daily rhythms of waking and sleeping, eating and drinking, inhaling, and exhaling, are assumed to be part of the background that only needs to be emphasized in higher levels when they are relevant to some event at these levels, as for instance the meal at which a marriage proposal is made. There is an analogy here between higher groupings within narrative descriptions and the movement toward higher viewpoints as described by Bernard Lonergan.

As for the structure of narrative frames themselves, Minsky describes a known narrative as consisting of chained trans-frames. According to Minsky, a “transframe” is a frame that represents the conditions before and after an event by modelling the difference between the frames for the situation prior to and after said event. In effect, by Minsky’s account cognised narratives are equivalent to a representation of an ordered list of descriptions.

On the subject of scripts, which Minsky seems to classify as a kind of narrative in the above passage, Minsky describes their structure as follows:

Most computer programs are mainly composed of sequential actions like these—but at some points the flow is interrupted by some branching *If*s, at which the script can go in several directions, depending on specified current conditions. Nevertheless, once the process is done, one can simplify and summarize it by listing only the path that was actually taken—as in, “*I was trying to build an arch with my toy blocks, and I discovered that one has to build the supports before putting on the top*”—which omits any mention of all the diversions involved in learning this.²⁸⁸

²⁸⁷ Minsky, *The Emotion Machine*, 196.

²⁸⁸ Minsky, *The Emotion Machine*, 281, emphasis original.

In summary, Minsky thinks that the general structure of scripts as sequential events broken by forking conditionals is fundamentally the same as those of narrative, and he further thinks that our own cognitive uses of narrative are comparable to scripts.

What Minsky considers to be the general features of narrative, suffusing many cognitive domains, is expressed below:

...we tend to interpret such “narratives” in terms of how situations change on successively larger scales of time. In the case of most music, this structure is clear: we begin by grouping separate notes into “measures” of equal length, and we then group these into larger sections, until the whole composition is seen as a storylike structure. We do this in vision and language, too—although with less repetitiousness—by grouping collections of smaller events into multiple levels of events, incidents, episodes, sections, and plots.²⁸⁹

On Minsky’s terms, the ability to recognise expected sequences of events or patterns of information, often developing or changing according to some rules, is a function of narrative frames. Thus, Minsky considers understanding the progression of a piece of music to be a species of understanding through narrative. Sets of notes may activate a frame that recognises them as themes or motifs and their sequences and variations may activate frames recognising a piece of music’s overall progression. More generally, to grasp a sequence of changes with predictable progression is to grasp via narrative, according to Minsky. Narratives seem, on this account, to be means of creating descriptions that represent other descriptions, identifying the uninteresting from that which is interesting either because the latter is unexpected or marks the forks between the different possible progressions of the plot, script, or pattern that differentiate a given narrative from another.

There does appear to be some way in which Minsky’s notion of narrative involves embodiment, at least implicitly. Recall that Minsky claimed that many concepts were structured

²⁸⁹ Minsky, *The Emotion Machine*, 196.

with reference to the human body. Making sense of the descriptions in a narrative requires having many such descriptions.

Of course, each of the items named in a story or script may refer to other, more complex structures. For example, to understand what terms like *Joan* or *book* mean, the reader must already possess some structure or models that represent these.²⁹⁰

Narratives, on Minsky's account, make use of other descriptions, many of which will need to be built up from experience of the world. Consider, for example, a child learning to listen to stories. Minsky observes that the child will often ask for details that adults would consider extraneous.

Even the best storytellers find it hard to entertain children, who are prone to interrupt with questions that make perfect sense by themselves but drift away from the story's theme. "*Where does Mary live?*" "*Does she have a dog?*"²⁹¹

Minsky claims it is difficult to tell stories to children because they lack the experience to identify what is and is not of interest in a story, which prompts them to ask questions about background details that adults consider irrelevant to the story and readily supply themselves. Making sense of a narrative on Minsky's account requires considerable use of already developed mental representations.

Minsky's notion of narrative also builds upon his account of metaphor, which he said is a means to transport thoughts from one realm to another and is inextricable from ordinary thought.

Finally, our "*Storytellers*" interpret each piece as similar to events in other realms—such as depicting a journey through space and time, or a skirmish among personalities. One special appeal of music is how effectively it can depict what we might call "*abstract emotional scripts*"—stories that seem to be about entities about whom we know nothing at all except that we can recognize their individual characteristics—e.g., *this one is warm and*

²⁹⁰ Minsky, *The Emotion Machine*, 282.

²⁹¹ Minsky, *The Society of Mind*, 265, emphasis original.

*affectionate, whereas that one is cold and insensitive. Then we empathize with how they feel by interpreting those phrases and themes as representing mental conditions like conflict, adventure, surprise, and dismay—as in, Those horns are attacking the clarinets, but the strings are now trying to calm them down.*²⁹²

In this example, the cognitive agents that Minsky calls “storytellers”, the neural structures that create the kinds of descriptions that can be called narratives, transport descriptions of auditory structures and changes in structure of a musical performance to descriptions of progress through space and time or to descriptions of conflict, for example. Narratives, as Minsky describes them, thus depend upon his formulation of metaphors.

Minsky does, therefore, regard narrative as useful for communicating and, arguably, stimulating understanding. In particular, he acknowledges that narratives can help connect a great many descriptions, parsing out important from unimportant details. Moreover, he acknowledges that narratives can be connected with emotional response, which communicates not only understanding but a stance towards what is understood. Nevertheless, as regards the meaning of narrative, it is hard to see what he adds to the treatment of narratives as sequences of well-structured representations, reducible to scripts. Although Minsky acknowledges narrative as important to present-day human intelligence, he has a high confidence in the possibility of an exhaustive representational description of the world in which the role of narrative is largely a matter of convenience.

Ray Kurzweil develops ideas similar to those of Minsky regarding the role of narrative in human cognition:

A story can be regarded as a meaningful pattern of information.²⁹³

²⁹² Minsky, *The Emotion Machine*, 197, emphasis original.

²⁹³ Kurzweil, *The Singularity Is Near*, 5.

The distinguishing component of the above definition of a story is the adjective “meaningful”, which implies that the pattern carries some meaning. What, however, is the nature of this meaning?

Kurzweil’s definition of narrative seems to be very abstract, but he does elsewhere describe the role of narrative within ordinary language in a way that suggests he, too, has at least an implicit awareness of narrative’s pervasiveness and its importance for cognition.

Each idiom introduces a story or analogy that must somehow be integrated with the rest of the subject matter. Understanding language can thus require a knowledge of history, myths, literary allusions and references, and many other categories of shared human experience.²⁹⁴

Kurzweil focuses here on idiom, but what is interesting about this passage is the claim idioms introduce stories or analogies that must be integrated with the rest of the subject matter. Understanding ordinary language, then, requires knowing many kinds of stories including history, myth, and literary allusions all of which, crucially are identified as being examples of the wider categories of shared human experience. To restate the point, Kurzweil identifies shared human experience as a prerequisite for understanding ordinary language because narratives and analogies are often included in ordinary speech.

There is a crucial point about how Minsky and Kurzweil conceive of narrative in communication. Speech is often, for them, not a direct transmission of data but a parsimonious means of sending signals from a speaker to a listener that will activate what representations the listener already has. This will be returned to in the next section, for they come close to making the observation that communication between humans is less about a transmission of data than about achieving an alignment of perspectives and a sharing of understanding.

²⁹⁴ Kurzweil, *The Age of Intelligent Machines*, 307.

The transhumanists do not merely discuss how narrative relates to the rest of human cognition; they also put it into practice. Marvin Minsky employed narrative explicitly when cowriting a novel exploring his ideas on human cognition and engineering AI.²⁹⁵ Ray Kurzweil often presents his predictions about future trends in technological development in the form of a narrative, and he frequently employs the literary device of fictional dialogues between characters to convey the human experience of these changes.

The third key transhumanist of this study, Nick Bostrom, does not discuss the theory of the cognitive role of narrative explicitly. Nevertheless, it is striking that he employs narrative a great deal in his rhetorical texts. One such text is “The Fable of the Dragon Tyrant”, published in the *Journal of Medical ethics*, and it, provides an instructive example of how Bostrom employs narrative. What follows are some selected excerpts from this parable.

Once upon a time, the planet was tyrannized by a giant dragon.... It demanded from humankind a blood-curdling tribute: to satisfy its enormous appetite, ten thousand men and women had to be delivered every evening at the onset of dark to the foot of the mountain where the dragon-tyrant lived. Sometimes the dragon would devour these unfortunate souls upon arrival; sometimes again it would lock them up in the mountain where they would wither away for months or years before eventually being consumed...

Seeing that defeating the tyrant was impossible, humans had no choice but to obey its commands and pay the grisly tribute. The fatalities selected were always elders.²⁹⁶

Expectations had gradually adjusted and the dragon-tyrant had become a fact of life. In view of the evident futility of resistance, attempts to kill the dragon had ceased. Instead, efforts now focused on placating it.²⁹⁷

²⁹⁵ Harry Harrison, *The Turing Option*, with Marvin Minsky (Gateway, 2013); Minsky, “Society of Mind,” 395. The title of the novel as it is listed in the “Response to Four Reviews” is *When Tomorrow Comes*, but it was changed prior to publication to *The Turing Option*.

²⁹⁶ N. Bostrom, “The Fable of the Dragon Tyrant,” *Journal of Medical Ethics* 31, no. 5 (2005): 273, <https://doi.org/10.1136/jme.2004.009035>.

²⁹⁷ Bostrom, “The Fable of the Dragon Tyrant,” 273.

And as for the impossible notion that the dragon-tyrant could be killed, history books recounted hundreds of attempts to do just that, not one of which had been successful...

Meanwhile, the wheel of invention kept turning. Mere decades later, humans did fly and accomplished many other astonishing things.

A few iconoclastic dragonologists began arguing for a new attack on the dragon-tyrant. Killing the dragon would not be easy, they said, but if some material could be invented that was harder than the dragon's armor, and if this material could be fashioned into some kind of projectile, then maybe the feat would be possible. At first, the iconoclasts' ideas were rejected by their dragonologist peers on grounds that no known material was harder than dragon scales. But after working on the problem for many years, one of the iconoclasts succeeded in demonstrating that a dragon scale could be pierced by an object made of a certain composite material. Many dragonologists who had previously been skeptical now joined the iconoclasts.²⁹⁸

The narrative goes on to tell of the difficult process of convincing the people of the world to pursue the goal of slaying the dragon, their desperate hope, their eventual success, and finally their grief and realising just how many more people might have been saved had they studied how to kill the dragon sooner.

Bostrom's narrative employs imagery and evokes affective response to the experience of suffering apparently inevitable death at the jaws of a dragon to try to claim that there is an analogy with the ordinary experience of senescence and death. In the context of this parable, Bostrom presents something taken as a fact of ordinary human existence (aging and eventual death) as the effect of violence from an external entity. So doing, he intends to catalyse a shift in the reader's perspective towards death. In effect, this parable constitutes an attempt to express an extended metaphor for the history of humanity's experience of death from initial acquiescence to eventual, hard-won freedom from it. Further, the rhetorical force of the

²⁹⁸ Bostrom, "The Fable of the Dragon Tyrant," 274.

narrative comes from its ability to generate emotive and embodied reactions to the otherwise abstract and alien concept of technologically eliminating death.

The difference between the transhumanist and philosophical positions on the tacit is illuminated by these differing accounts of the understanding of narrative. Although both accounts recognise that narrative depends on both embodiment and metaphor to convey understanding, the transhumanist account shifts the focus from narrative as expressing shared experience and understanding to abstract descriptions. From the philosophical perspective, however, understanding a narrative is described in terms of imaginative and affective engagement with other experiences and perspectives. The transhumanists offer additional layers of abstraction and complexity to their account of narrative that is alien to the psychological account.

One further point may be made about the difference between the transhumanist and philosophical approaches to narrative. The transhumanist descriptions of narrative do not regard *telos* as integral to the structuring of a narrative. Structured and predictable change may occur from scene to scene to objects that remain sufficiently recognisable for a continuity to be ascribed to them, but the notion of a directedness toward some end is at best submerged in the transhumanist descriptions of narrative. In contrast, from the philosophical perspective the directedness toward some end is a necessary element of narrative, distinguishing it from bare description of detail on the one hand or the evocation of experience without the implication of change that is achieved in some poetry.²⁹⁹

²⁹⁹ It is the case that much literature is both poetry and narrative. The point is not to suggest that there is no overlap between the two, but rather to highlight that a narrative must be about things that change and the changes, which must be toward something. Poetry need not be about something changing and can be about experiences without reference to time. It is for this reason that Mystics often use poetry as the medium for expressing their experiences of the ineffable and the eternal.

In summary, transhumanists offer a somewhat ambiguous assessment of the role of narrative in accelerating insight. On one hand, they acknowledge that narratives can help connect a great many descriptions, parsing out important from unimportant details. They also acknowledge that a good deal of human understanding, including of language, involves at least implicit references to all kinds of narrative, including history and myths. Moreover, Bostrom, at least, makes powerful use of narratives for rhetorical reasons, to communicate understanding along with his preferred stance towards what is understood. On the other hand, in many texts, it seems that the transhumanists do not think that narratives are of fundamental importance, especially to potential transhuman and posthuman intelligences.

What is clear, however, is that narratives are important for present day human beings, especially children, in situations of shared understanding. This sharing raises the question of whether these situations of sharing are themselves cognitively significant for accelerating understanding. I therefore now turn to the issue of second-person relatedness.

4.2.5 Second-Person Relatedness and Understanding

The transhumanist aspirations to accelerating insight extend beyond just augmenting the performance of the individual but extend to increasing the power of shared cognition, hence shared understanding, between multiple humans. To what extent are these aspirations to enhancing interactive or cooperative intelligence based on a notion of intersubjectivity that is consonant with that of the philosophical account? Chapter two explored how shared understanding involves shared awareness, also called joint attention or second-person relatedness. How do the transhumanists characterise the relationships between second-person relatedness and understanding?

Before discussing how the transhumanists wish to augment capacities enabling shared understanding or the changes in understanding called insights, it is worth exploring what they think the role of second-person relations in ordinary human cognition is. The recent scholarship in psychology and philosophy exploring the importance of shared awareness with other persons for child development and the creation of new understandings, i.e. insights, uses the terms joint attention and second-person perspectives (the stance in which one shares awareness of shared focus with another person, characterised by the use of second-person pronouns) to characterise this intersubjective awareness. These terms “second-person perspective” and “joint attention” have relatively recently entered the philosophical lexicon, so it is not surprising that the transhumanists do not use them in their writings. Nonetheless, transhumanist literature does discuss phenomena included in these categories, especially in the context of explaining child pedagogy. These discussions will serve as the first point of departure for examining what the transhumanists have to say about second-person relatedness and understanding.

One passage from Kurzweil in which he discusses the role of the interactions between a teacher and a student in education is as follows:

We all have models of the world that we use to understand and respond to events and to solve problems, whether they be real-world situations or classroom exercises. If we make a mistake, it may be simply a matter of a few missing or inaccurate facts. But more often it reflects a structural defect in the organization of our knowledge. A good teacher attempts to understand the model that the student is using. Then if the student’s model, and not just his data base, is faulty, the teacher devises a strategy to modify the model to more accurately reflect the subject matter.³⁰⁰

In this passage, Kurzweil reaffirms that the world is understood and problems are solved through models. Mistakes are identified as resulting from either errors or deficits of fact or in a structural deficit in the organisation of knowledge, or on his terms, a mistake in the model. The

³⁰⁰ Kurzweil, *The Age of Intelligent Machines*, 431.

role of the teacher is to understand the student's model and determine a means of modifying the student's model to better represent its object. Within this framework, education consists of either conveying facts or creating optimal mental models, the latter being the means of sharing understanding. There is an element of intersubjectivity in Kurzweil's account insofar as the teacher needs to try to understand the mental models of the student, and have sufficient grasp of the student's habits of thought to be able to guide them to a more correct kind of model-making and understanding, but the shared objects of attention are still purely the abstract objects of representations, and the account of how a student might be guided to have a better mental model is not addressed.

Kurzweil's treatment of education as primarily about conveying facts and skill in representation finds echoes in other transhumanist authors. A passage from Minsky shows partial overlap with these concerns in his description of the qualities of a good mentor.

Minsky wants to provide teachers with ways to think about thinking, which he contrasts with what he considers the harmful belief that humans just rely on a mysterious power of intuition, a term which, as discussed above, he deprecated along with understanding in *Society of Mind*. Instead, he thinks that children should be taught to describe themselves as programmable computers capable of installing as mental software various practical strategies to help them generate solutions to problems.³⁰¹

³⁰¹ These learnable strategies include trying to solve problems by analogy, simplifying problems by breaking them down into parts, changing descriptions of the problem to reveal more relevant information, making the problem more specific if one has too many ideas, making the problem more general if one has too few ideas, reflecting on whether difficulty in solving a problem indicates whether a different approach should be tried or whether one should abandon the problem in favor of another, imagining how a more expert person would approach the problem, or resigning oneself to resting and letting nonconscious parts of one's mind try to find solutions. He additionally includes as a strategy simply applying the solution if one already knows it. Minsky includes as the final strategy asking another person for help. Marvin Minsky, *Inventive Minds: Marvin Minsky on Education*, with Cynthia Solomon and Xiao Xiao (Cambridge, Mass: MIT Press, 2019), 82–84.

Minsky acknowledges that most of the strategies he lists are already performed by children and adults but affirms his conjecture that teaching people to explicitly think of and put into practice these strategies will help them. Minsky does not, however, explain what benefit accrues from describing teaching these strategies according to the metaphor of programming a computer. A student can, after all, be taught to believe themselves mentally flexible and learn heuristics for problem-solving without believing themselves to be a computing machine.

Minsky's statement of what is needed in a good mentor or pedagogue summarises these points while also pointing beyond them.

...a good mentor must also be a Tutor, to teach the factual content of a field—and should also be a Coach, to train up the needed set of skills—and also should be a Role-Model who can inspire a good set of values and goals.³⁰²

Minsky identifies the three roles that a mentor should fill as being that of a tutor, a coach, and a role-model. The tutor communicates facts, the coach trains the student in skills, and the role model is meant to inspire good values and goals. Communicating facts and having skill in using them appears congruent with Kurzweil's representation of pedagogy as making sure that students have correct facts and structured representations of them. The cognitive skills Minsky discusses mostly concern strategies for using and reformulating the descriptions of these facts, so despite some slight differences in language, they offer a similar perspective. Minsky's identification of the function of role-models as inspiring good values and goals expands on this perspective. The use of the word "inspire" appears to imply that the student is not just a receptacle for the data of facts and problem-solving software but that there is a kind of affective response to the role-model themselves. What more does Minsky have to say on the matter, and to what extent does it map onto a transfer of understanding in a second-person relation?

³⁰² Minsky, *Inventive Minds*, 94.

This additional level of education occurs, often informally, through the relationships children form with these role-models, which Minsky calls “imprimers”.

An Imprimer is one of those persons to whom a child has become attached. For animals that raise their young, the function of infant attachment seems clear: remaining close to parents helps to nourish, teach, and protect their offspring. However in humans, attachment has another effect—on the children’s ultimate values and goals; when your Imprimer praises you, you feel a special thrill of pride that elevates the goal of your present activity to have some kind of higher priority. Similarly, a sense of shame depresses your present goal’s desirability.³⁰³

Minsky defines imprimers as persons to which children are attached and from whom they received not just physical care and practical training but the formation of their goals and values. This primarily occurs through the experience of feelings of pride and shame when praised or rebuked, respectively, by someone to whom an attachment has been formed. These sentiments prompt one to assign priority to those things the imprimer values and lower priority to those things the imprimer disvalues.

Yet, it may be asked whether Minsky’s description of the function of a role model actually captures what is meant when describing two people coming to share a stance toward something, in this case the stance that some goal or value is suitable. To be sure, in the ordinary course of child development, children will internalize what things parents, caretakers, and other important persons in their lives consider to be worthy values and those which they consider to be unworthy, but to say that they will feel pride when praised for valuing something or following some goal and shame when criticized for it will cause them to value and disvalue similar things is not to say that they will share the same stance towards these values and goals that the role-models do. This description is perfectly with a child adopting the rule that what their role-model likes is valuable and what their role-model dislikes is bad. In principle, this description allows

³⁰³ Minsky, *Inventive Minds*, 92.

for a child to value the same things as their role-model and to the same degree without actually sharing in the adult's attitude toward the values. This suggests that Minsky's description is incomplete, since children usually advance beyond just caring about their mentors' feedback to caring about the values themselves.

Although Minsky and Kurzweil do not use terms like joint attention or second-person relatedness in the context of education, they both show an awareness that interactions between teachers and learners do contribute to effective sharing of understanding. Minsky and Kurzweil, as stated above, describe what are commonly called acts of understanding in terms of structured representations or descriptions of facts and acts of insight in terms of reformulation of these representations. Hence, the transhumanist descriptions of educator interaction aiding the formulation of representations and of teaching techniques for reformulating representations in the context of problem solving can be interpreted as expressions in their terms of the capacity for interpersonal interactions between students and teachers to accelerate the acquisition of understanding and the catalysation of insights. Further, Minsky at least seems to show some awareness of the importance of relations between children and role-models for developing one's stance toward things, yet his description of the role personal relations in value acquisition largely consists of conditioning in terms of positive and negative reward signals rather than in the nature of the engagement between these persons. If Minsky and Kurzweil's language describing the mind as a computer and teachers as programmers is stripped away, no real content remains of their descriptions. Their perspective, at least with respect to actual human education, does not offer anything new, and what they do discuss of the personal relations between students and teachers does not seem to be wholly adequate to the phenomena.

In contrast, the account of pedagogy in chapter three describes a general approach to pedagogy for teaching cognitive skills in specific contexts, namely apprenticeship. The pupil

attends to the practice of the master, learns to imitate the master with the aid of the master's appraisal and correction, and finally assimilates the practice of the art. The apprentice may learn articulate rules and maxims as well, but these function as guides and are not exhaustive of the understanding acquired from imitating the master's example, as Polanyi discusses.

The large amount of time spent by students of chemistry, biology and medicine in their practical courses shows how greatly these sciences rely on the transmission of skills and connoisseurship from master to apprentice. It offers an impressive demonstration of the extent to which the art of knowing has remained unspecifiable at the very heart of science.³⁰⁴

In this passage Polanyi claims that the length of time students of various sciences such as chemistry, biology, and medicine spend in practical studies shows that much of science is unspecifiable, by which he means that it is not known explicitly. The skills and connoisseurship that are passed from master to apprentice require their interaction and mutual attention, which allows the student to align both practice and acquaintanceship with the teacher. In contrast with Minsky and Kurzweil's approach, there is no additional assumption that the understanding conveyed is encoded in representations.

Further, although Minsky says that he wishes to denigrate the idea that problem solving depends upon some undefinable notion of intuition (this is consistent with his description of "understanding" as a term of convenience), the strategies he describes are only superficially procedural. They provide general maxims or guides for action, but unlike the implementation of actual software, they require understanding on the part of the person employing them. Some discernment, itself relying on skills and familiarity with similar problems that must be gained rather than an explicit result of instructions to be followed, must be employed to determine that one needs to approach the problem in a different fashion. What's more, if any of the strategies

³⁰⁴ Polanyi, *Personal Knowledge*, 55.

employed are successful, it will be through making it possible for some sort of insight to occur. For instance, the advice that one seek an analogy between the current problem and another that one knows how to solve is simply an injunction to try to grasp a similarity between the problem and something already understood. Further skill in selecting useful analogies can come through either personal experience or through an imparting of implicit skill in constructing such analogies by an accomplished thinker. Any such pedagogy by sharing of implicit understanding remains undiscussed by Minsky.

Kurzweil also reflects on pedagogue-pupil interactions and the prospects for technologically enhancing education. His thoughts on the matter appear to be influenced by his experience of deploying a machine meant to assist students with visual and auditory disabilities.

It then reads the document out loud while highlighting the image of the print as it is being read. It essentially does what a reading teacher does—reading to a pupil while pointing out exactly what is being read.³⁰⁵

Although Kurzweil recognises that the educator performs the task of pointing to things while they teach and that this is beneficial, he reduces the function of the teacher to drawing attention to text. In other words, there is an implicit assumption that interpersonal interactions between pupils and teachers causes only the benefit of stimulating attention and can be replaced by an artificial stimulus. Kurzweil speculated that there was considerable opportunity for automating the role of the teacher.

Using the literary device of giving a description of the world from the perspective of someone in the future, Kurzweil weaves a narrative about future progress towards transhumanist transcendence that includes technological enhancement of education.

Intelligent courseware has emerged as a common means of learning. Recent controversial studies have shown that students can learn basic skills such as reading and math just as readily with interactive learning software

³⁰⁵ Kurzweil, *The Age of Spiritual Machines* (1999), 178.

as with human teachers, particularly when the ratio of students to human teachers is more than one to one... The traditional mode of a human teacher instructing a group of children is still prevalent, but schools are increasingly relying on software approaches, leaving human teachers to attend primarily to issues of motivation, psychological well-being, and socialization.³⁰⁶

Kurzweil's story about the future includes the prediction that interactive software will be able to take the place of human interaction for basic reading and math skills, which demonstrate more dramatically Kurzweil's assumption that at least the interpersonal interaction that occurs between student and teacher may be reduced to functions that may be realised in a program. As noted previously, the philosophical perspective is that the joint attention shared by a pupil and a teacher of an object, an instance of triadic awareness, allows a sharing of stance between the teacher and the pupil, allowing the pupil to see the object as the teacher does, catalysing insight and allowing a sharing of understanding. The act of reading to a child can contain a great deal of implicit communication as child and reader mutually respond to changes in intonation, cadence, and emotion. In effect, a great deal can be communicated about what the text expresses, especially if the content is a narrative, and how to respond to it. By contrast, Kurzweil thinks that the triadic awareness of a teacher instructing a pupil confers no benefits that cannot in principle be specified and automated.

Minsky and Kurzweil's writings on pedagogy both exhibit a tendency to abstract functional descriptions of interpersonal interactions that omit that which is tacitly shared and essential for the jointness in such joint attention and is often a condition for sharing understanding. Their broader speculations about how understanding might be shared among augmented persons reflect further such omissions.

³⁰⁶ Kurzweil, *The Age of Spiritual Machines* (1999), 192.

Kurzweil's own example of one possible avenue of increased communication facilitated by augmented reality and AI illustrates a situation where a seemingly more efficient communication system could actually stifle useful interactions. The example is that of a worker engaged in multiple simultaneous virtual committees with the aid of an AI assistant to help manage rote responses and identify the aspects of the conversation to which one most needs to turn. Kurzweil simply means to illustrate that technology will change the environments we are expected to survive in and shall provide the means to do so. Such a splitting of attention, however, would also have the effect of reducing the ability to engage successfully in any one group and reduce one's ability to maintain the kind of sustained attention that often facilitates insight. Further, if the communications from the other committee members are nontrivial, it means that one shall have trouble assimilating them. Although it is true that there is merit to Kurzweil's suggestion that AI assistants could be used to reduce unnecessary work (much as contemporary word processors can reduced editorial burdens through spellchecking functions and citation managers can lessen the required work for managing citations), it nonetheless remains the case that in facilitating insight, the reduction of distractions and dealing with rote tasks is advantageous, but the focus is not placed upon that which humans use to gain new insights.

In particular, the nature of shared understanding is not discussed. The most that Kurzweil says is rather vague, namely that it will be possible to directly share knowledge from mind to mind. Bostrom expresses much the same point. While we may imagine some undefined ability to share thoughts directly in some quasi-angelic communion, merely making such a description does nothing toward actually articulating the nature of such a communication. Note that it also presupposes an abstraction of meaning from the body of the one who understands the meaning. A further explanation of this may be worthwhile. All communication in some way

involves sensations produced and perceived by bodies. The understanding of a symbol resides in one person and is shared from one to other by leading the other to gain a new understanding either by expressing symbols that are then understood at once or attempting to lead the other through a progression of understandings culminating in a new understanding. Sometimes, one must simply point to something and affirm to a trusting interlocutor that there is something to be understood. Trust on the part of the one who receives instruction is one of the requirements of pedagogy, for instance, and it is sometimes required for such interactions to be fruitful.

Kurzweil imagines multiple speculative technologies that would alter human interactions and communications, in ways he thinks would ultimately deepen communication. As already discussed in the section on embodiment, he discusses possibilities for virtual reality and augmented reality communications systems, which he imagines will eventually allow experiences phenomenologically indistinguishable from real embodied interactions, with the exception that virtual interactions will allow possibilities not afforded by the constraints of actual physical interaction, such as the ability to meet in environments that would otherwise be inaccessible, to alter one's appearance as an additional mode of expression, or to jointly participate in actions that would be impossible in the real world.

Further, Kurzweil predicts that as algorithms grow in capability and complexity, some will not only become integrated into daily life and used in communications, but that they will be given conversational capabilities allowing users to interact with them in a second-person mode. Further, he speculates that such machines will increasingly be tasked with identifying relevant information for their users to attend to as the world grows more complex and the need to reduce human labour in sifting through and managing trivial data in favour of focusing on important data grows. Such machines would operate in a fashion analogous to a pedagogue or colleague identifying important information to which to attend in a given context. Kurzweil

thus evinces not only optimism about the prospects for future machine capabilities but a recognition of the privileged role of interpersonal interactions as a means of communicating.

Finally, Bostrom's musings on enhancement of collective intelligence describe ways to improve the quality of communication of ideas between persons. His thought experiment about increasing the human population by some orders of magnitude and his practical suggestion about using gene engineering or embryo selection to increase the total number of creative geniuses and high IQ individuals both illustrate Bostrom's belief that forming larger communities of highly intelligent persons will increase the total number of insights. On the one hand, this could occur trivially if the persons are having insights independently of one another, but Bostrom also considers the development of new systems of communication for greater efficiency and idea generation. On the other hand, his account does not actually explore what it is about collaboration that catalyses insights that one individual working alone could not achieve. He does appear to recognise that this collaboration does allow greater insight to occur, but to some extent this might be due to trivially increasing the number of insightful persons such that what allows multiple people to function more effectively than one is not really explored. At least, it is clear that there is more involved than mere parallelization. There is a suggestion that communication can be made more efficient and that this will benefit the group's ability to produce insights. Although it is true that rendering communication more efficient can benefit the ability to produce insights, the fruitfulness of the communication will still depend on other factors such as a shared desire to understand a question and capacity to enter a shared stance toward the objects of this inquiry. Increasing the number of intelligent persons in a group and increasing the efficiency of their communications will not necessarily increase the capacity of the group to produce new insights. They may equally just allow an increase in rote or procedural work and communication. The conditions favourable to interpersonal understanding

and shared pursuit of insights must be met if technology is to be used to enhance peripherally a collective drive to gain new shared understandings.

Bostrom assumes that increasing the population or improving methods of communication will positively affect the rate of generation of new ideas, but he does not explore the means by which this improvement is expected to occur. While in the abstract stating that allowing a greater number of persons to communicate more efficiently would facilitate more understanding, the actual catalysts of such understanding need to be grasped. Consider the example of *Social Physics*: better ideas are facilitated through rich social networks that allow communication of ideas, which allows shared conclusions to approve of and improve upon fruitful insights while rejecting the inadequacies of inferior ideas. Mere rapid communication, however, does not ensure that the kind of personal interactions will occur that facilitate the kinds of conversation that lead to the generation, sharing, and evaluation of insights.³⁰⁷

Bostrom's other suggestion concerning the development of collective intelligence is problematic for another reason. Bostrom explores the notion that communication technologies and computation could eventually reach a level of sophistication and integration that allows the assemblage of humans, computers, and communication infrastructure to become a unified intellectual agent. This assemblage would not merely be a coordinated set of smaller intellects and their tools but would itself constitute a thinking entity. On this view, the collective would become something like a global brain and the individual humans and their tools components analogous to individual neurons or neural circuits.

In responding to Bostrom's proposal, it is not necessary to examine the ontological questions concerning such a claim here. Rather, it is sufficient to observe that this idea does not

³⁰⁷ Pentland, *Social Physics*.

actually explore how humans share understanding or work together to achieve new insights. Positing that humans and their machines could together instantiate some super-computer or super-brain does not illuminate how humans actually cooperate to understand or how to enhance the human ability to gain new understandings. Further, even if this agent could be said to understand (and humans were still considered to have their own independent existence and understanding), it would not follow that humans share in this agent's understanding. The same issues that attend to other putative superintelligent entities would also apply to this collective agent. If the understanding is incommunicable, then humans could still benefit from the practical applications of the agent's understanding, but human understanding would not itself have been enhanced. If the understandings are communicable, then the question of how they are to be communicated to and shared among humans still remains unexplored.

One area where the lacuna in the transhumanist perspectives on insight and second-person relatedness is particularly evident is in their discussion of the kinds of knowing that belong to meditative experiences or, in the highest case, the non-representational knowledge of God termed the beatific vision. In *Deep Utopia*, Bostrom references Thomas Aquinas definition of the beatific vision and quotes this relationship as a “face to face” beholding of God by the saints and angels. He cites the definition of the beatific vision as knowing and loving God.³⁰⁸ When he glosses this definition, however, he omits any reference to the personal and fails to understand the crucial characteristics of the beatific vision.

If one does take this second approach, according to which one can score meaning-points simply by being good or by appreciating and loving the good, then it would be a lot easier for the utopians to have great meaning in their lives. They could reform themselves (with the aid of neurotechnology, if need be) to have virtuous characters that make it easy to approach each day with thankful relish for all the good they bring in

³⁰⁸ Nick Bostrom, *Deep Utopia: Life and Meaning in a Solved World* (Ideapress Publishing, 2024), 307.

train, and with warm appreciation of all the good in their fellow utopians, and with profound love and worship of divine goodness itself.³⁰⁹

Although Bostrom's gloss on the beatific vision is primarily his attempt to transpose its lessons into the context of engineering a utopian existence, the key lesson to draw from his attempt is that an attempt to render the beatific vision in transhumanist terms fails to capture the meaning of the beatific vision on its own terms, and fails to recognise the distinction between supernatural and natural knowing. In the theological conception, the beatific vision is a knowing and loving of God that goes beyond natural powers and is given by God directly in a kind of friendship that is described by the metaphor of seeing face to face.³¹⁰ In Bostrom's rendering, a utopian rendering of the beatific vision is just a creature (possibly a cognitively enhanced transhuman or posthuman) cognising and loving the divine with its own native powers. Even then, Bostrom's rendering reduces the loving of the divine to a simple utilitarian maximization of meaning points.

Bostrom's transposition of the concept of the beatific vision into transhumanist terms fails to even capture the meaning of God. Bostrom suggests one way of designing utopians would be to love the being with greatest goodness, but God is not to be conceived of as merely the best being alongside others.

³⁰⁹ Bostrom, *Deep Utopia*, 412–13.

³¹⁰ It is striking that Aquinas speaks of friendship between humans and God, given that Aristotle's account of friendship involves elements of equality and reciprocity. Yet, Aquinas does emphasize that humans are lifted up to share God's own life and bliss and empowered to know and love God with a kind of participation in God's own knowledge and love of himself. As Aquinas writes, "However, benevolence is not sufficient for the character of friendship; instead, a certain mutual loving is required, since a friend is a friend to his friend. Now this sort of mutual benevolence is founded upon something shared in common (fundatur super aliqua communicatione). Therefore, since man shares something in common with God insofar as God communicates His own beatitude to us, it must be the case that some sort of friendship is founded upon this sharing. 1 Corinthians 1:9 says of this sharing, '... the faithful God, by whom you have been called into the fellowship of His Son.' But the sort of love built on this sharing is charity. Hence, it is clear that charity is a certain sort of friendship of man with God." *ST II-II* q. 23, a. 1.

We are not to think of God as a powerful being within the universe but as creator of the entire universe and all that is in it.³¹¹

It must be stated explicitly that God is not part of the hierarchy of being but is outside of it as creator. The beatific vision is not simply a matter of knowing and loving the pinnacle of the hierarchy of being with one's creaturely intellect and emotions, even if they are enhanced by merely technological means.

The following quotations from Aquinas assert that knowledge of God in Himself is not accessible to creatures in the same sense as all other knowledge of entities within the universe.

I claim that the proposition "There is a God" is known *per se* as far as it itself is concerned, since the predicate is the same as the subject. For as will become clear below (q. 3, a. 4), God is His own *esse*. But because we do not know the real definition of God, this proposition is not known *per se* to us. Instead, it has to be demonstrated by means of things that are more known to us and less known by their nature, viz., God's effects.³¹²

Aquinas' meaning is that God directly understands his own existence, by something analogous to *intellectus* in the strict sense, a kind of direct insight. God's self-understanding is through direct insight. Humans, indeed all created beings, do not know God's existence through such insight into first principles, and natural knowledge of God must come through a demonstration based on God's effects.

And from this it is clear that God does not have a genus or differences; nor is there a definition of Him or a demonstration of Him, except through His effects. For a definition is composed of genus and difference, and the middle term of a demonstration is a definition.³¹³

This implies that God cannot be fitted into our creaturely categories, by which we relate creatures to each other and locate them within conceptual schemes. What is naturally known of God is not, therefore, God's own essence, since creaturely knowledge, human or angelic, cannot

³¹¹ Herbert McCabe, *The Teaching of the Catholic Church: A New Catechism of Christian Doctrine* (Darton, Longman & Todd, 2000), 3.

³¹² Aquinas, *ST I-I* q. 2, a. 1.

³¹³ Aquinas, *ST I-I* q. 3, a. 5.

encompass God in the way it can the rest of creation. Natural knowledge of God must be indirect, and any direct understanding of God would need to be of a different order than any understanding of the natural powers.

Thomas writes that whatever knowledge humans currently have of God is only by similitudes or images that are not of God's essence.

In *De Trinitate* 15 Augustine says, "When the Apostle says, 'We see now through a mirror and dimly' (1 Corinthians 13:12), by the terms 'mirror' and 'dimly' he can be taken to be referring to any likenesses that are accommodated to understanding God." However, seeing God through His essence is not a dim or mirror-like vision, but is instead contrasted with such a vision. Therefore, it is not the case that the divine essence is seen by means of likenesses.³¹⁴

The direct, beatific understanding of God is contrasted with understanding by means of concepts. For Aquinas, the beatific understanding of God is not that of which humans by means of concepts.

In Aquinas' parlance, God is most intrinsically knowable as Truth itself, but no concept in the intellect can be portioned to God. At the same time, no concept is needed to know God.

Therefore, one should say that in order to see God's essence what is required on the part of the visual power is a certain likeness, viz., the light of glory, which empowers the intellect to see God. This is spoken of in Psalm 35:10: "In Your light we shall see light." However, God's essence cannot be seen by means of any created likeness that represents God's essence as it is in itself.³¹⁵

It might be said that God is *too knowable* on Aquinas account, but that the created intellect is too weak to receive God unless strengthened.

Now when a created intellect sees God through His essence, God's very essence is the intelligible form of the intellect. Hence, it is necessary for some supernatural disposition to be added to the intellect in order for it to be elevated to such great sublimity. Therefore, since, as has been shown (a. 4), a created intellect's natural power is not sufficient for seeing God's

³¹⁴ Aquinas, *ST I-I* q. 12, a. 2.

³¹⁵ Aquinas, *ST I-I* q. 12, a. 2.

essence, a power of intellective understanding has to be added to it by God's grace. This addition of intellective power is what we call the illumination of the intellect, just as the intelligible thing itself is called light or a source of light. This illumination is the light of which Apocalypse 21:23 says, "The glory of God will illuminate it [read: the society of the blessed who see God]." And by this light the blessed are made godlike, i.e., similar to God—this according to 1 John 3:2: "When He shall appear we shall be like to Him, and we shall see Him as He is."³¹⁶

The natural power of the intellect, according to Aquinas, cannot understand God, so God must by grace give the human intellect illumination so that the human can share in God's own self-understanding. The beatific vision is a kind of supernatural shared insight that unlike all creaturely insight and understanding does not involve concepts or images and is best described by metaphors of a sharing between persons.

For Aquinas, while the sense cannot receive understanding of the divine nature because they are limited to the particular,³¹⁷ the intellect is not limited in the same way because it can know universals that are abstracted from the particular.

Our intellect has cognition of the infinite in the same sense in which it is infinite in power. For its power is infinite in the sense that it is not terminated by corporeal matter. And it has cognition of the universal, which is abstracted from the material individual, and, as a result, it is not terminated in any individual, but, as far it itself is concerned, it extends to infinitely many individuals.³¹⁸

³¹⁶ Aquinas, *ST I-I* q. 12, a. 5. See also *I-I* q. 12, a. 1 "Each thing is knowable to the extent that it has actuality, and so God, who is pure actuality without any admixture of potentiality, is in Himself maximally knowable (*maxime cognoscibilis*). But what is maximally knowable in its own right may not be knowable to some intellect because its intelligibility is too great for that intellect—just as the sun, which is maximally visible, cannot be seen by a bat because its light is too great." According to Aquinas, the more real a being is, the more true it is, hence the more knowable it is. God is too real for a created intellect to know through its own powers, analogous to how a sufficiently bright light might induce blindness in beings with sensory powers not suited to it. God's transcendence means that all creaturely intellects will be insufficient to understanding Him through their own powers. This would remain true for an archangel and for reckoning being, human or posthuman.

³¹⁷ Aquinas, *ST I-I* q. 12, a. 3.

³¹⁸ Aquinas, *ST I-I* q. 86, a. 2.

God is still beyond the inherent power of the intellect, but the intellect can still be supernaturally strengthened to receive God himself.³¹⁹

It follows on Aquinas' account that the beatific vision is beyond the natural powers of any created intellect, human or angelic. As a corollary, the Thomistic conception of the beatific vision is one that would be beyond any powers of cognition enhanced by merely technological means.

It is impossible for any created intellect to see God's essence by its own natural powers. For cognition occurs insofar as what is known exists in the knower. But what is known exists in the knower according to the mode of the knower. Hence, every knower's cognition accords with the mode of its own nature. Therefore, if the thing known is such that its mode of being exceeds the mode of a given knower's nature, then the cognition of that thing must lie beyond the nature of that knower... to know subsistent esse itself is connatural only to the divine intellect, and that this lies beyond the natural power of every created intellect. For no creature is its own esse; instead, every creature has participated esse. Therefore, no created intellect can see God through His essence except insofar as God, through His grace, conjoins Himself to the created intellect as something that can be understood by it (*se intellectui creato coniungit ut intelligibile ab ipso*).³²⁰

Intelligent beings know according to their natures, and no created being understands in a way that would allow it through its own powers to understand self-subsistent being. Knowledge of God's being must be shared by grace, which puts it beyond the limit of any possible cognitive enhancement.

Moreover, while loving God is a component of bliss, it does not constitute this bliss. The goal is rather sharing in God's own self-understanding, which may be called friendship with God, and issues in a sharing in God's own bliss.

³¹⁹ It is worth noting that there are two very different views of the soul here. The transhumanist view is the human "soul" is a complex AI device, albeit an biologically evolved one, so the whole is the sum of its parts in their interaction. On the Thomist view, there is an immortal soul that is the form in the Aristotelian sense of the whole, and this form as the intellect with a built-in capacity for infinity. This means the soul can be expanded to receive the vision of God, but an AI couldn't.

³²⁰ Aquinas, *ST I-I* q. 12, a. 4.

As was explained above (q. 2, a. 6), two things are required for beatitude: (a) the essence of beatitude and (b) its *per se* accident, as it were, viz., the delight adjoined to it.

Thus, I claim that as regards what beatitude is in its essence (*essentialiter*), it is impossible for it to consist in an act of the will. For it is clear from what was said above (aa. 1-2 and q. 2, a. 7) that beatitude is the attainment of the ultimate end. But attaining an end does not consist in an act of the will. *For the will is directed toward the end—an absent end when the will desires it, and a present end when it delights while reposing in it.* But it is clear that an act of desiring an end (*desiderium finis*) is not itself an act of attaining the end (*ipsum desiderium finis non est consecutio finis*); rather, it is a movement toward the end. Moreover, *an act of delighting comes to the will because the end is present, whereas it is not the case, conversely, that something is present because the will delights in it.* Therefore, there must be something other than the act of the will (*oportet aliquid aliud esse quam actum voluntatis*) such that through it the end itself becomes present to the one who wills...

...at the beginning we will to attain some intelligible end. But we attain it by the fact that it becomes present to us through an act of the intellect. *And then our delighted will comes to rest in the end now attained.*

So, then, the essence of beatitude consists in an act of the intellect, whereas what pertains to the will is an act of delighting that is consequent to this beatitude. Accordingly, in *Confessiones* 10 Augustine says that beatitude is “rejoicing in the truth” (*gaudium de veritate*), because *the act of rejoicing is the consummation of beatitude.*³²¹

Aquinas says that true happiness is not the orientation of the will to God or the delight in receiving God, though both are naturally present in the act of loving God. It is actually having the end of sharing in God’s own knowledge and life that is what Aquinas calls happiness. Happiness, for Aquinas, is not one’s subjective satisfaction but one’s state of being. Further, the happiness does not consist in the orientation of one’s will to its true end, an orientation that is loving God, but happiness consists in sharing in God’s self-knowledge. Contra the lesson Bostrom draws from the beatific vision, happiness cannot be reduced to having the best possible object of love or to knowing much about it through one’s own powers.

³²¹ Aquinas, *ST I-II* q. 3, a. 4. emphasis added.

The beatific vision, it may be seen, is of a different order from natural understanding entirely. Any transhuman or posthuman with its expanded powers would still be capable only of a creaturely understanding of the divine, however more sophisticated than that of humans it might be. Bostrom's rendering of the beatific vision into the terms of meaning maximization by loving that which is maximally good shows that he is unable to describe the beatific vision not only because of his physicalism, but because of the absence of the personal dimension of human insight from his thought, a personal dimension that can be elevated by grace into a personal communion with God.

This attempt to render the beatific vision into terms that can be achieved by transhumanist means shows a failure to understand the original theological meaning of the beatific vision. Further, the failure of Bostrom's attempt to render the theological concept of the beatific vision in transhumanist terms highlights the incommensurability of enhancement to understanding with divinely infused understanding.³²²

Kurzweil does not mention the beatific vision, but he does discuss spiritual experiences both for humans, hypothesized thinking AI's, and prospective posthumans. He suggests that all will share neural machinery (or software equivalents) that allow for subjective experiences of the divine or the transcendent, and that such experiences will be valuable to superhuman intelligences as they are to current day humans, but he provides no further analysis.³²³

Minsky does not discuss either the beatific vision, but he does consider meditative and religious experiences of transcendent or higher knowledge. He suggests that these are a result

³²² Eberl also observes the incommensurability of transhumanist aspirations and the beatific vision. "The eternal—i.e. atemporal—happiness of loving union with God cannot be compared on any scale with the temporal joys of this life..." Eberl, "Who Wants to Live Forever?," 89.

³²³ Kurzweil, *The Age of Spiritual Machines* (1999), 252.

of the misfiring of neural structures that give a sense of reward when something is known despite the experience being empty of content.

There is also a different and more sinister way to make the world seem orderly, in which the mind has merely found a way to simplify itself. This is what we must suspect whenever some idea seems to explain too much. Perhaps no problem was actually solved at all; instead the mind has merely found some secondary pathway in the brain, through which one can mechanically dislodge each doubt and difference from its rightful place! This may be what happens in some of those experiences that leave a person with a sense of revelation—in a state in which no doubts remain, or with a vision of astounding clarity—yet unable to recount any details. Some accidental mental stress has temporarily suppressed the capacity to question, doubt, or probe. One remembers that no question went unanswered but forgets that none were asked! One can acquire certainty only by amputating inquiry.

When victims of these incidents are compelled to recapture them, their lives and personalities are sometimes permanently changed. Then others, seeing the radiance in their eyes and hearing of the glory to be found, are drawn to follow them.³²⁴

Minsky elides over several concepts in this passage. Initially, he speaks of ideas that seem to have the simplifying effect of insights, but which are false and perverse because they involve a simple refusal to see the complexity of the phenomenon. These would not be true insights constituting higher viewpoints that unify and clarify disparate phenomena but errors akin to confusing a toy model of the world with the world itself and circumventing any cognitive or emotional reflection on the toy model's adequacy. Minsky then shifts to discussing experiences of revelation, which he says are contentless yet have a subjective experience of being without doubt and having clarity. The evidence, for Minsky, that these experiences are sinister malfunctions of the brain and those who have them victims, are that these experiences do not consist of questions asked or answers given. In other words, there cannot be knowing that is not explicit.

³²⁴ Minsky, *The Society of Mind*, 65.

In contrast, the Aristotelian-Thomist Perspective can acknowledge knowing or understanding another person as an encounter that does not require explicit offering and giving of information. The theological perspective supports and deepens this by describing the knowing that comes from God as an infused knowledge describable by the metaphor of a relationship with God.

Thomas' notion of the infused intellectual virtue of prudence, and the Gifts of the Holy Spirit, is helpful here. These can be given in the absence of strong natural powers of reasoning but nonetheless carry an implicit understanding. In the case of infused prudence, for instance one is given understanding of how to live in keeping the life of grace,³²⁵ these judgements often being of particular situations and contexts. One has this understanding although no explicit knowledge has been given. Thomas uses the metaphor of God being a master and the soul with the Gifts being an apprentice.³²⁶ As we saw from Polanyi, the sharing of understanding between a master and an apprentice is a kind of implicit, personal communication. As Pinsent observes, this relationship can be described by the metaphor of joint-attention. One's understanding is akin to sharing perspectives with God. The most famous case of divine enlightenment in Christian scripture, that of St. Paul's conversion,³²⁷ takes the form of a personal encounter.

The Aristotelian-Thomist perspective, then, can make sense of the claim that meditation and mystical experience³²⁸ can impart non-explicit understanding in contrast to Minsky's interpretation of such experiences as neural misfiring.

³²⁵ *ST* I-II, q. 65, a. 4 ad 1.

³²⁶ *ST* I-II, q. 68, a. 2.

³²⁷ Acts 9: 1-21.

³²⁸ It does not follow that all the experiences of one who meditates, or all claimed mystical experiences, are from God or provide authentic understanding about the true God. St John insists that we should "test the spirits to see whether they are of God," and offers criteria (I John 4:1-4).

In summary, transhumanists cannot make sense of the claims of shared understanding with God because of their inability to consider understanding that is not exhausted by explicit representation. It renders them unable to describe understanding that comes from personal encounter.

In conclusion, the transhumanist perspective of the role of second-person relatedness on understanding is, from the philosophical and theological perspectives, extremely limited. Minsky and Kurzweil describe sharing understanding in education in terms of transmitting facts and skill in reformulating representations. Although they sometimes acknowledge an apparent significance to personal awareness between students and teachers, they describe such awareness abstractly in terms of functions of data transfer that omit much of the actual psychological content shared implicitly by the persons. In describing the use of technology to allow new modes of communication, possible obstacles to obtaining shared perspectives are overlooked in favour of new modes of communicating data. Speculations about how collaboration can be enhanced do not actually address how shared understanding can be improved. The matter is either treated trivially by suggesting that merely increasing the number of collaborators will increase the rate of insights occurring simply by increasing the number of people having insights at the same base rate, abstractly by suggesting that some communication technologies might reduce inefficiencies that slow down the communication of ideas, or very speculatively by imagining the possible emergence of a distinct and superior intelligent entity which includes humans as components. In all these cases, the facilitating conditions in which second-person relations enable sharing understanding or catalysing insights are largely not the focus of transhumanist aspirations. It seems unlikely, then, that transhumanism will prove helpful in understanding either second-person relatedness or how technology can be deployed to help such states facilitate understanding and insight.

4.3 Concluding Observations on Transhumanist Prospects

The philosophical and transhumanist pictures of understanding do have similarities in describing the functional benefits of understanding and some means by which it can be catalysed. There is agreement that changes in understanding, namely insights, are experienced as singular and result in new or higher understandings and that actions evince these changes to understandings. There is agreement that bodily experience contributes to the internal representations associated with understanding and is useful for the creation of new understanding. The transhumanists and the philosophers agree that metaphor catalyses insights and facilitates the building of new understandings. Both agree that narrative is a particularly powerful vehicle for sharing understanding. Likewise, in both the philosophical and transhumanist pictures of understanding, joint attention, at least in the form of communication, is recognised as facilitating insight.

Although the two perspectives do have many such points of agreement about the inputs and outputs of the black box of understanding, efforts to describe what understanding is intrinsically show considerable disagreement. The transhumanists treat the apparently singular quality of acts of insights to be a merely subjective appearance due to the assumption that some computational, and ultimately mechanistic, explanation of the phenomenon must exist. Consequently, the status of insight as basic to cognition is denied and description of insight in terms of some algorithm is supplied. Similarly, the transhumanists ultimately deny that embodiment is more fundamental than and necessary for the role of abstraction. Bodily experience may shape abstractions, but sufficiently detailed representations can capture the whole of an understanding and are equivalent to that understanding. Having denied the necessity of a body to understand representations, transhumanists similarly treat metaphor not as a

connection between embodied experience and meaning but instead as a complex algorithm or set of algorithms for finding mappings between different representational structures. The role of narrative as involving an implicit sharing of characters' understandings and a sense of ends is displaced by an overbroad description of narratives as patterns with, at most, end states to a process. Finally, the role of shared awareness and implicit communication in human interactions is obscured by an attempt to describe functions humans perform in conversations that may be either altered or replicated by machines.

The contrast between the transhumanist and the philosophical account are brought into sharper relief by the Thomistic account of the intellect. The human intellect is, for Thomas, purely spiritual, yet it operates with and through the body and its senses in a unified way. The human as a rational being, is naturally an embodied and hence social being. Further, because the human intellect is immaterial, it is capable of sharing in the Beatific Vision, which is not accessible by any intellectual enhancement of a natural or scientific kind.

Ultimately, transhumanist perspectives on understanding attempt to reduce understanding to a computational account of mind, by which the generation of insights is ultimately a matter of algorithm and understanding is a system or systems of data storage and retrieval. The transhumanist account is an attempt to assimilate the implicit into what is explicit, in the sense of a computer program. To use McGilchrist's terminology, the transhumanist perspective on understanding is an attempt to capture a "right-hemispheric" cognitive act in a "left-hemispheric" idiom. The resultant description of understanding is not commensurate with the phenomenon of understanding, and the prospects for achieving transhumanist aspirations toward cognitive enhancements are harmed through this misunderstanding of the nature of understanding and the conditions that catalyse it. It remains to examine in chapter five how, if at all, technology might nevertheless be used to enhance understanding.

CHAPTER 5 Towards a Synthesis: Human-Machine Cooperation

5.1 Introduction

The previous chapter argued that the transhumanist perspectives on understanding and insight are deficient, denying the reality of them as distinct phenomena and neglecting the importance of embodiment, which leads them to mischaracterise important concomitants of insight, specifically metaphor, narrative, and joint-attention.

If the prospects for a transhumanist approach to seeking cognitive enhancement through becoming computational entities are underwhelming, are there nonetheless opportunities for enhancing cognition, specifically accelerating insight, that can be described in a non-transhumanist framework? In this chapter, therefore, I turn to a broader question, namely whether such technologies may nevertheless be of indirect help in accelerating insight. In particular, I consider ways in which such technologies increase the scope of human senses and intra-human communications; cognitive tasks in which only the results of insight matters; and situations in which human insights might be triggered indirectly by the knowledge of these results.

5.2 Enhancing the Conditions of Insight

To examine how enhancements might benefit insight, it is first necessary to distinguish between internal enhancements and external enhancements. Internal enhancements are technologies that result in a direct alteration to the human body. These include genetic modification, implantation of cybernetics, and treatments (surgical or pharmacological) that are meant to bring about a permanent change in the body. Internal enhancements are those which

are most paradigmatic for transhumanist authors, for such technologies promise to bring about a change in the powers of the human organism.³²⁹

External enhancements, by contrast, are tools and techniques that extend in some way the powers of the human person without modifying the body directly.³³⁰ While there has been considerable change in the kinds of tools that humans use to extend their powers, the practice of using tools is ancient and extends to prehistory. The transhumanists make the claim that the use of such tools is ultimately contiguous with their project of human enhancement. The radical differences between the lifestyles and capacities of present-day humans and those of our hunter-gatherer ancestors is such that transhumanists suggest that we could already in a sense be considered transhuman with respect to them.³³¹ In other words, they are most concerned with technology radically increasing human powers. Since the human body does entail certain limitations on possible actions and design constraints on external tools (a human body cannot survive accelerations beyond a certain threshold, for instance, limiting options for vehicle

³²⁹ Whether or not cybernetic implants should be considered metaphysically part of a human or should be considered a metaphysically distinct tool is a question that does not need to be investigated here. Whether a human that has received cybernetic implants is to be considered an aggregate of a human and a metaphysically distinct set of tools that just happen to be embedded in flesh or is to be considered a metaphysical unity isn't fundamentally important to the transhumanists. If the technology is experienced as increasing one's intrinsic powers and contributes to a trajectory of technological change resulting in the change of human entities into posthuman entities, then it can be considered to be altering one's natural powers for the purposes of the transhumanist project.

³³⁰ For an examination of how internal enhancements and external enhancements can contribute to or impede human flourishing according to a Catholic perspective, see Andrew Pinsent and Sean Biggins, "Catholic Perspectives on Human Biotechnological Enhancement," *Studies in Christian Ethics*, February 8, 2019, <https://ezproxy-prd.bodleian.ox.ac.uk:2246/doi/full/10.1177/0953946819826769>.

³³¹ 'Our current extended phenotypes (and the lives that we lead) are markedly different from those of our hunter-gatherer ancestors. We read and write, we wear clothes, we live in cities, we earn money and buy food from the supermarket, we call people on the telephone, watch television, read newspapers, drive cars, file taxes, vote in national elections, women give birth in hospitals, life-expectancy is three times longer than in the Pleistocene, we know that the Earth is round and that stars are large gas clouds lit from inside by nuclear fusion, and that the universe is approximately 13.7 billion years old and enormously big. In the eyes of a hunter-gatherer, we might already appear "posthuman".' Although humans have not actually changed as a species, Bostrom's point is that the conditions of life, interests, and activities of contemporary humans are different and arguably superior to those of our hunter-gatherer ancestors to the extent that contemporary human lives might seem to them to be not what they would expect of human beings but what they would instead expect of superhuman beings. Nick Bostrom, "In Defense of Posthuman Dignity," *Bioethics* 19, no. 3 (2005): 213, <https://doi.org/10.1111/j.1467-8519.2005.00437.x>.

design), they are interested in modifications that overcome human biological limitations, and, ultimately, the limitations of biologically realised intelligence entirely.

The notion that there can exist gradations in the effectiveness of the faculty of insight finds precedence in Aquinas. In examining how different persons can exhibit different degrees of understanding, he writes

...someone can understand the same thing better than someone else, in virtue of having more power of understanding—in the same way that someone who has a more perfect power and in whom the visual power is more perfect sees a given thing better by a corporeal act of seeing than someone else does. There are two ways in which this occurs in the case of the intellect: One way is on the part of the intellect itself, when it is more perfect. For it is clear that to the extent that the body is better disposed, the soul is better. This is manifestly obvious in the case of beings that are diverse in species. The reason is that actuality and form are received into matter in accord with the matter's capacity. Hence, since even among men there are those who have a body that is better disposed, they receive a soul that has a greater power of intellectual understanding. Hence, *De Anima* 2 says, "We notice that those with soft flesh are more mentally gifted."³³²

In this picture, it is true that different persons can possess a greater or lesser ability attaining a grasp of things. The difference in capacity at understanding is explained through differences between the actualization of the soul's powers, which is due to a disparity between the fitness of each person's matter to actualize the form. This implies that, if the deficiencies of the body were corrected, the power of understanding could itself be better actualized.

Further, a difference in the facility in understanding may be ascribed to secondary powers with which the faculty of understanding cooperates.

The second way is on the part of the lower powers that the intellect needs for its own operation. For those in whom the power of imagining and the cogitative power and the power of remembering are better disposed are better disposed for intellectual understanding.³³³

³³² Aquinas, *ST I-I* q. 85 a. 7.

³³³ Aquinas, *ST I-I* q. 85 a. 7.

On Aquinas' view, understanding will be actualised more perfectly when the other cognitive faculties, themselves dependent on the matter of the body being properly in-formed, are themselves well-actualised. As before, this account leaves open the possibility that can there be improvements in the physical structures by which the cognitive faculties operate, improvements that will also benefit understanding.

Appreciating the implications of Aquinas' view for the aspiration for enhancement of insights requires reference to his account of the proper functioning of understanding in humans and its reliance on the functioning of the senses and the imagination. As he states, without the phantasm, there is no understanding (or, to put it into the terms of Bernard Lonergan, all understanding requires some sort of image of that which is understood), and without sensation, the phantasm cannot be of an object in the world. Further, corruption of the senses and other faculties cooperating with understanding can result in errors in understanding. As a corollary, these faculties can be restored through returning the flesh to function and health.

With Aquinas taken as representative of the philosophical perspective presented in chapter three, at least on this topic, the lesson may be drawn that there exist targets for efforts to produce some enhancement of an ability to understand. Namely, some improvements to the body can yield corresponding improvements in enhancements either to understanding itself or to the faculties with which it operates. It should be observed, however, that while the notion that a more perfect body will be capable of more perfect acts of understanding is entailed in the tradition Aquinas represents, he does not consider the possibility of enhancing a human being's cognitive faculties by altering the body.

On Aquinas' perspective, deficiencies in the instantiation of the human form can result in deficiencies in its faculties, and a more suitable instantiation of its form with respect to the organs of thought will result in superior faculties, but only up to what is proper to humankind.

To the extent that natural faculties potentially exist that are superior to those of humankind, they would be proper to a different kind of organism.

Although there exist some approaches to internal enhancement of the body that can produce benefits for human cognition, such as taking drugs which can slightly increase some measures of cognitive performance, they can be equivocal and do not yet represent radical innovations.³³⁴ Proposals such as germline genetic engineering to increase intelligence or the modification of whole brain emulations are highly speculative.

At present, prospects for the enhancement of understanding directly, say by modification of brain structure, are matters of speculation. Bostrom may speak of the possibility of creating a mind qualitatively superior to those of humans whilst acknowledging that any discussion of such entities must be vague, but such speculations offer no real illumination to the question of how insights may be accelerated.

How may insight be enhanced through technologies affecting the “raw material” of insight, the objects of the faculties that are cooperative with insight? This is not an exploration of how the faculty of insight may be directly enhanced in some fashion analogous to how muscle mass may be stimulated by artificial steroids, for the neural concomitants of understanding are not well understood, but rather of how faculties that are understood to a greater or lesser extent, may have their capacities enhanced through technological means.

The power of tools to act as enhancements to human faculties of perception and manipulation are well-documented. There are, therefore, possibilities of using technology to enhance human faculties that facilitate and accelerate understanding. Michael Polanyi discusses the use of tools as extensions of the body in which one becomes tacitly aware of the tool and

³³⁴ Bostrom and Sandberg, “Converging Cognitive Enhancements,” 11–15.

proximally aware of that with which the tool is interacting.³³⁵ By this, he means that the tool becomes a means by which we experience and ultimately come to know the world. The blind man does not feel the walking stick in his hand but instead is aware of whatever the end of his stick touches.³³⁶ This is a simple example illustrating the power of tools to extend the power of perception.

Ample evidence does exist, moreover, that insights can be accelerated through technologies that provide indirect benefits, such as by bringing otherwise inaccessible phenomena into the spheres of human perception and manipulation. The former may be illustrated by such tools as the early telescopes, which allowed perception of both new astronomical bodies and previously unknown features of those bodies to be perceived and so understood. Sonar creates a visual representation of the propagation of sound waves that can be sensed and apprehended. Similarly, the tools of chemistry facilitate experimentation on substances that could not be manipulated by hand. In both cases, inherent human capacities are not extended but previously imperceptible phenomena are made accessible to human perceptual abilities, so human understanding may engage with them.

The trigger system of the ATLAS detector in the Large Hadron Collider at CERN illustrates the possibilities for scientific research that were beyond human reach before the maturation of computing technology. In each second there are forty million beam crossing

³³⁵ Polanyi explains that when using tools to sense and manipulate the world, we experience them as extensions of our bodies. “Our subsidiary awareness of tools and probes can be regarded now as the act of making them form a part of our own body. The way we use a hammer or a blind man uses his stick, shows in fact that in both cases we shift outwards the points at which we make contact with the things that we observe as objects outside ourselves. While we rely on a tool or a probe, these are not handled as external objects. We may test the tool for its effectiveness or the probe for its suitability, e.g. in discovering the hidden details of a cavity, but the tool and the probe can never lie in the field of these operations; they remain necessarily on our side of it, forming part of ourselves, the operating persons. We pour ourselves out into them and assimilate them as part of our own existence. We accept them existentially by dwelling in them.” Humans already use tools to bring otherwise inaccessible objects into contact with their abilities to perceive and manipulate. Polanyi, *Personal Knowledge*, 59.

³³⁶ Polanyi, *Personal Knowledge*, 55–61.

events in the central detector. Apart from the superhuman speed required to detect the events, it is also necessary for the detections to be screened for statistically significant events to ensure the data storage is manageable. If all raw data were stored without being filtered by triggers and subsequently compressed, ATLAS alone would produce one petabyte of data per second. The level one trigger located at the hardware of the detector selects one hundred thousand events per second (a reduction by a factor of 400) for analysis by the high-level trigger, which uses approximately forty thousand CPUs to select the one thousand most significant events for storage. After compression algorithms are applied, this is equivalent to sixty-four terabytes of data per second. The trigger system alone illustrates the practical impossibility of present high energy particle physics in the absence of tools capable of physical operation and symbolic processing at superhuman speeds. In other words, the equipment and experiments at CERN allow otherwise utterly inaccessible phenomena to be brought into human awareness and hence into the reach of insight.

It is also possible for simulations and virtual experience to aid humans in understanding. Davis and Hersh recount the striking use of software that created an interactive representation of a hypercube. A hypercube, the four-dimensional equivalent of an ordinary three-dimensional cube, cannot ordinarily be experienced, given that everyday life is three-dimensional. This software enabled the rotation and manipulation of the hypercube in three-dimensional space. One of the authors describe this experience as follows:

I tried turning the hypercube around, moving it away, bringing it up close, turning it around another way. Suddenly I could *feel* it! The hypercube had leaped into palpable reality, as I learned how to manipulate it, feeling in my fingertips the power to change what I saw and change it back again. The active control at the computer console created a union of kinesthetics

and visual thinking which brought the hypercube up to the level of intuitive understanding.³³⁷

In other words, it is, in fact, possible to gain intuitions about four-dimensional objects, thanks to the virtual manipulation of such objects in a three-dimensional space. Moreover, such an intuition is clearly regarded, by the author, as an advance over merely abstract descriptions.³³⁸

At present, computer simulations trade depth of physical experience for computational power (turning knobs on a computer interface is less kinesthetic than physically manipulating a physical model), but continuing advances in virtual reality may allow for more sophisticated and realistic models that further engage our physical senses. Note that this will be more useful for some kinds of models than others. Mathematics necessarily deals with abstractions, so manipulating knobs or sliders on an interface to effect rotations of an object may already be sufficient to gain a grasp of the object in question, since one only needs to grasp the way that physically manipulating position leads to a change in the represented relative positions of vertices and faces in a three-dimensional slice of a four-dimensional object. For simulating something in greater particularity, it becomes necessary for more and more detailed approximations of reality.

These examples underline, again, the role of embodied experiences for promoting understanding. They also show that contemporary technology can be used to help to extend understanding beyond the limits of what can ordinarily be imagined without requiring modification of the faculty itself, or indeed of any human faculties.

³³⁷ Davis and Hersh, *The Mathematical Experience*, 404.

³³⁸ To one extent, this kind of simulation is just an outgrowth of the already extant practice of model-making. Kepler, for instance, used paper models to perform advanced topological analysis. I owe this observation about Kepler's use of paper models to a conversation with Professor Nick Jardine of Cambridge. In another example, Antoni Gaudí used inverted rope models to determine and grasp the required structure for the basilica of *La Sagrada Família* according to the principle that ideal arches are similar to inverted catenary curves, which are the shapes formed by ropes or chains that have reached a state of equilibrium.

In conclusion, Aquinas, at least, does not seem to exclude the possibility of enhancing the faculty of understanding, albeit by conditioning the body rather than modifying the body in the manner of the transhumanist speculations. As regards these speculations, there is, at present, very little evidence of the actual direct enhancement of understanding, or any plausible path to attain this goal by technological means. Nevertheless, as I have also noted above, there are many plausible ways in which understanding may be enhanced indirectly. Examples include extending the senses by means of instruments that transform the imperceptible into the perceptible, extending the temporal scope of action by machines that work faster than the human mind, and software that can give an intuitive grasp of higher dimensions.

5.3 Outsourcing Those Cognitive Tasks in which Only the Results of Insight Matters

One of the key transhumanist goals is the augmentation of human cognition. They take the development of increasingly powerful AI systems as a short-term goal that serves as a stepping stone towards realising human-equivalent and above faculties of understanding and insight in a machine that can in turn be transferred to humans through cybernetic modification or as an upgrade to an uploaded consciousness. The transhumanist project would wish to interpret developments at the time of writing as progress toward this goal. Nonetheless, even when systems are designed that can bypass the need for human understanding in specific domains, they may be better thought of as an outsourcing of cognitive tasks in which only results matter, which does not compete or displace but ultimately serves to indirectly benefit insight.

There are instances of producing results that would, if generated by a human, be the result of understanding. The Galaxy Zoo, a large citizen scientist project for classifying images of galaxies according to their shapes, provides a case study of the advantages that can accrue

from saving humans not only from rote formal operations but also from rote, albeit informal, acts such as recognising and classifying images of objects. The project was originally conceived because the quantity of images of galaxies captured by robotic telescopes exceeded the capacity of astronomers to process, and human volunteers were more accurate than algorithms at correctly classifying galaxies. Although humans were still performing acts of insight, the researchers outsourced the acts of insight across a huge number of participants.

In 2019, the project announced that Bayesian convolutional neural networks had grown sufficiently reliable, in spite of the uncertainties of the accuracies of the labels used in the training data, that, the majority of image classification could now be performed automatically. From that time, human volunteers would only continue to classify galaxies to provide further data for the training of the algorithms. While this classification is not a formal task for volunteers, requiring recognition and assessment on their part, correctly labelling the data was the purpose of the volunteers' actions rather than achieving whatever understanding the scientists analysing the data hope to gain. The classification is results oriented, and it is the results that matter rather than the understanding of a human producing the results.³³⁹

Another example of an area of research where only the results of insight matter and it is an indirect enhancement for human insight to be bypassed, at least partially, is protein folding prediction. Protein modelling is computationally intensive whether it is approached with brute force methods or when human intuitions into possible protein shapes are evaluated algorithmically. Indeed, there exist citizen scientist projects for leveraging human intuition and creativity for suggestions as to possible folding to be computationally evaluated. As in the case

³³⁹ Mike Walmsley et al., *Galaxy Zoo: Probabilistic Morphology through Bayesian CNNs and Active Learning*, Oxford University Press, 2019.

of Galaxy Zoo, determining the shapes of the phenomena under study historically relied to an extent on human insight, but it the results of this insight were what mattered.³⁴⁰

Impressive milestones in computational protein modelling were achieved by Google DeepMind's AlphaFold 2 algorithm, which was able to predict protein structure with near parity to experimental results during the Critical Assessment of Structure Prediction (CASP) challenge of 2020.³⁴¹ Two years later, AlphaFold had predicted the structures of 200 million proteins, which included most known proteins. Thirty-five percent of the predictions were determined to be highly accurate, comparing favourably to experimental results, and forty-five percent of the remaining predictions were found to be usable for many applications.³⁴²

Automatic text generation provides another example of an application in which, for some applications, only the results of insight matter. For many types of routine communication, it is sufficient to structure and present information in a human readable format. An algorithm could, for instance, identify noteworthy trends in sales in a dataset and create text that highlights this and is easily grasped. Additionally, human guided use of AI text generation can be used to more rapidly compose texts in which human understanding is only required to identify what is important to communicate but not to compose the actual text for the communication.

There is a related opportunity for using automatic text generation that is specifically designed for tasks such as coding. Large language models are currently capable of producing workable code with some human oversight. This production is particularly useful for

³⁴⁰ David Baker et al., "Predicting Protein Structures with a Multiplayer Online Game," *Nature* (London) 466, no. 7307 (2010): 756–60, <https://doi.org/10.1038/nature09304>.

³⁴¹ Ewen Callaway, "'It Will Change Everything': DeepMind's AI Makes Gigantic Leap in Solving Protein Structures," *Nature* (BERLIN) 588, no. 7837 (2020): 203–4, <https://doi.org/10.1038/d41586-020-03348-4>; John Jumper et al., "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature* (BERLIN) 596, no. 7873 (2021): 583–89, <https://doi.org/10.1038/s41586-021-03819-2>.

³⁴² Ewen Callaway, "'The Entire Protein Universe': AI Predicts Shape of Nearly Every Known Protein," *Nature* 608, no. 7921 (2022): 15–16, <https://doi.org/10.1038/d41586-022-02083-2>.

performing much of the rote work of writing the text of the code after the programmer has made decisions about the structure and functionality of the code, as well as suggesting optimisations for human-written code. Generative text algorithms offer a tool that allow humans to employ their understanding and insight to establish the higher-level order of a practical task, while bypassing the need to employ human insight to complete secondary and derivative work.

A similar symbiosis between human insight and mechanical outsourcing of insight can be seen in applications of machine translation. Machine translation engines can successfully translate much of the text but they struggle when the meaning of a phrase is highly contextual, idiomatic, or specific to a culture. These are the phrases that require the most effort and often ingenuity on the part of human translators to render correctly in a different language. On their own, a human translator can spend several additional hours working on the straightforward translation, and a machine translate a text in seconds but make serious errors on the passages of a more ambiguous nature. If human translators use the translation engines as tools, however, then they can devote their insight to translating the most context-dependent text and ensuring the overall sense of the translation is correct and relegate the uninteresting translation work to machines.

Other areas in which work that would require insight in humans can be outsourced to machines include performing medical diagnoses. The correct identification of the disease is more important than any particular doctor making the identification, and the application of doctor insights into patient care is better facilitated when correct diagnoses can be achieved more quickly and reliably.

Still other areas include the use of generative AI to assist in engineering projects by creating optimised geometries of parts and load-bearing structures. As with the algorithms that assist in the writing computer code, the human makes decisions about the high-level structure,

and the AI performs lower-level work to more quickly and easily bring the practical fruits of the human insight into the world.

AI can be used for more general image classification. There are many applications in which there is only a need to identify what an image is of, as, for instance, when verifying identity when passing customs in an airport. In cases where a correct output is all that is required, the need for a human to make the identification can be bypassed.

AI image generation of visual media such as images and video has continued to grow in sophistication. The systems are still limited to making predictions as to pixel patterns or sequences of pixel patterns, but they can be used for making cheap images or videos.³⁴³

The technologies discussed in this section all, in some fashion or another, serve to extend the reach of human understanding indirectly. There are tasks in which only the results of insights matter rather than the new understanding itself. For these tasks, it is adequate if the results can be reliably produced that are comparable to what would be produced by a human, or in some cases to an extent that is superior. These technologies do not need to be described in a transhumanist idiom, and an attempt to speak of the machines as seeming to possess something akin to human understanding, with their performances considered analogous to a Turing Test perhaps, obscures the relationships that humans have to them. Namely, bypassing work for which only the results of insights matter not only saves human effort but can serve as external enhancements that indirectly accelerate insight. Machines can facilitate the fruits of being realised more quickly and tested such as through the use of generative AI and by doing the lower-level work that enables insights into interesting questions such as through the astronomical census of Galaxy Zoo.

³⁴³ For more creative tasks, the images generated can be used to help stimulate the ideas of human artists.

5.4 Using Computers to Discover Missing Insights

The opportunities for enhancing human understanding using computers extends beyond merely allowing humans to gain the results of mere brute work in which the actual understanding of the computing agent is a marginal consideration. Advances in the field of artificial intelligence have shown that there are algorithms that can produce results that are similar to what would be produced by a human as the fruits of an act of insight. Such machines can be used to find what may be termed missing insights. That is, they may be used to return results that reveal what humans could have grasped through an act of insight, but which humans have not yet had the time or chance to achieve.

The previous section explored the use of machines to bypass the need for insight in those tasks where only the result of insight matters, and this section explores how machines can instead be used to aid human insight by discovering missing insights. The previous section discussed how machines could be used to bypass the need for human insight for some tasks, specifically those in which it was only the results of an act of insight that mattered.

The research of Google's DeepMind project into neural net engines to play Go and Chess is particularly illustrative of these points. The first of these systems to be developed, AlphaGo, surprised Go grandmasters with its innovative play.

During the match AlphaGo played many original, creative and beautiful moves, most famously move 37 in game 2, which in many cases overturned centuries of received wisdom. It wasn't just regurgitating or copying human knowledge. Subsequently Fan Hui and our many friends in the Go world have told us this has revolutionised how the game is played, and

many books have now been written about AlphaGo's unique playing style.³⁴⁴

The effect of AlphaGo on the theorising and practical play of Go is of note both because of the subjective evaluation of its moves as being like that which would be produced by human creativity, which would be a style of play resulting from a novel insight, and for the effect this new style has changed expert play. The effect AlphaGo had on Go theory differed from the effect of classic analysis engines on chess, as Demis Hassabis, the CEO of DeepMind, observed. Classical chess engines could perform searches of move trees with great rapidity and depth, but their strength lay primarily in their capacity to avoid mistakes rather than to generate new strategies that would result from a human having insight into the structure of the game.

Impressive as the success of the AlphaGo engine was, DeepMind developed a successor program called AlphaZero which was generalisable to other games than Go, including Chess. Matthew Sadler, the English chess grandmaster, after reviewing two hundred ten games between AlphaZero and Stockfish 8, the then current version of one of the most popular chess engines, he observed, firstly, that AlphaZero's play style was unique and individual, and

Secondly, when I tested key positions on the chess engines I use for analysis, I saw that AlphaZero's assessment was different to that of the other engines. In fact, AlphaZero's assessments were closer to my intuitive (human) feeling about the position.³⁴⁵

Sadler's description of the chess engine's play style illustrates that its board assessments were in harmony with his expert understanding and that AlphaZero's innovative gameplay seemed to him to be more like that of a creative human player than the gameplay of more traditional chess engines.

³⁴⁴ Matthew Sadler et al., *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI* (New In Chess, 2019), 62.

³⁴⁵ Sadler et al., *Game Changer*, 36.

AlphaZero's play is notable for being automatically developed through self-play with human input only in the design phase. Previous iterations of chess engines such as Stockfish, although capable of checking many millions of possible board positions to determine optimum moves, would still rely on some human designed evaluation functions to determine the strength of positions to avoid wasting processing time on depth searches of unpromising lines of play. These evaluation functions made use of human written rules such as assigning points to each type of chess piece and partially determining the strength of a position based on which player is ahead in material. AlphaZero was provided only with a numerical representation of the chess board and rules for moving pieces, and from initially random play to evolutionary optimisation of the weightings of the connections between the nodes of its convolutional neural network it developed a facility for play that not only was competitive with other chess engines but produced real innovations in chess strategy.

Chess engines before AlphaZero could use their speed of calculation to test enormous finite numbers of board positions and avoid making tactical blunders. They could be used to evaluate the strength of strategies by performing depth searches on possible board permutations, but they were useful for testing human insights rather than for supplementing them.

However, what we are finding with AlphaZero is that some of the improvements or the advancements are actually strategic in nature, and that is something that we can potentially incorporate in our own play.³⁴⁶

AlphaZero demonstrated the ability to produce new strategies that in humans would be innovations resulting from insight. Previous chess engines merely allowed players to use brute force calculation to evaluate their intuitions about novel strategies, but AlphaZero made it possible to generate move sequences that in humans would result from strategic insights.

³⁴⁶ Sadler et al ., 64.

Unlike in the case of the use of convolutional neural networks for galaxy classification, the results of an act of understanding are not the only things that are desirable. In the case of games such as Chess and Go, the goal is not merely to update an entry in a database but to improve the skills and abilities of human chess players. The use of deep learning algorithms in games such as chess and Go is of especial interest because the goals of mastery of these games is twofold. One goal is to win in a competition with another player, and there is a sense in which achieving victory is a concern with the bottom-line: chess and Go are zero-sum games and in competition one wins, draws, or loses. There is also a sense in which the games are valued aesthetically in which games are prized for the ingenuity and beauty of their play. In the second sense, the goal of the player or the analyst is not merely to calculate the means to victory but to achieve or share an understanding of the progression of the board states that elegantly navigates the branching possibilities of future board positions. To the extent that chess and Go can be approached not primarily as problems to be solved but as objects that it is valuable to understand, the goal in understanding them is terminal rather than proximate, and the significance of algorithms that can produce results not yet discovered by insight lies in the changes they effect in humans' holistic approaches to these games. More generally, although machines cannot yet be said to understand, advances in AI demonstrate that in some domains they can make discoveries that prompt the ascent to a higher viewpoint in a human person.³⁴⁷

³⁴⁷ The philosopher James Ross argues that computation is not equivalent to human cognition even in cases when a computer can return the same outputs as a human being given a set of inputs. Consider the case of an arithmetical operation such as addition or a logical operation such as modus ponens. In each case, performing the operation depends upon fulfilling a particular function that is both formal and abstract. "Size and length of computation, for example, are external to the form of thinking, accidental to what is done. I am squaring just in case my thinking is of the form mentioned. If it is of any impossible form, or is indeterminate among impossible forms, it is not of the form, 'N times N = N squared.' It is not then squaring, however much its products may look like it, and however long a sequences of its outputs do." (138-139) Ross' point is that a given abstract excludes other formal operation that are in some way similar to but distinguishable from it. The classic example to which Ross refers is Kripke's example of quaddition, which is defined as a function that performs the same as addition unless one of the operands is 57 or greater, in which case the function always returns 5. Ross asserts that when one performs the operation of addition this necessarily excludes the possibility that one is performing the operation of quaddition,

The reception of AlphaZero's play by chess experts is instructive as an example of how humans can respond to the results produced by an algorithm as they would to the communication of an insight from another human. Sadler recounts a eureka experience that he had while viewing a game between AlphaZero and Stockfish 8 in which the board developed into a position similar to many games that he had played against Stockfish, each of which Sadler lost to counterattacks along the h-file facilitated by Stockfish's rook on h8. When in a similar position to Sadler, AlphaZero instead developed a kingside pawn position that kept the h-file closed, prevented the h8 rook from moving again in the game, and allowed AlphaZero to achieve a checkmate.³⁴⁸ Sadler's experience was akin to that of seeing another person approach another problem in a novel way and having a shift in his understanding to align with the view of the situation evinced by the other's actions.

and vice versa. "The fact that I cannot process every case of modus ponens, because most of them have premises too long for me to remember, sentences too long to say, or words I do not understand, is adventitious, like my not being able to do modus ponens in Portuguese. Those are features of the functors, not of the function. The function that has to be realized in every case is the one wholly realized in the single case. That point is to be taken literally: that the function is wholly present, not by approximation, exemplification, or simulation, but by realization in the single case." (139) Recalling Lonergan's assessment that it is insight that forms a pivot between the particular and the universal, it may be said that it is by the act of understanding that a universally applicable function such as squaring is realised in a particular act of calculation. In each case of performing a particular abstract operation, the abstract operation is realised in the particular case, and the claim to be an instance of performing a function according to a universal form is not undercut by the practical inability of humans to perform certain kinds of operation. The practical impossibility of squaring an integer represented by a sequence of a trillion random digits does not negate the human ability to conceive of such an act as an instance of squaring, and the ability of a machine to return outputs for operations too large for a human to perform does not of itself entail that the machine has realised an instance of this function. Relatedly, a function that returns an approximation of another function, while it may be practically useful and referred to as if it realises this function, does not in fact do so. "No physical process is so definite as to determine among impossible abstract functions that one rather than another is realized, and thus to settle for every relevant case what the 'outcome' is to be. That indeterminacy remains no matter how long the physical process is 'repeated,' even infinitely. In a word, with a machine it is indeterminate among impossible functions what it is doing, no matter what it does." (140-141) Ross argues that the indeterminacy of different functions that could be realised by a physical process coupled with the impossibility that the physical process actually realise one and yet fail to realise "There is no doubt, then, as to what the machine is doing. It adds, calculates, recalls, etc., by simulation. What it does gets the name of what we do, because it reliably gets the results we do (perhaps even more reliably than we do) when we add by a distinct process. The machine adds the way puppets walk. The names are analogous. The machine attains enough reliability, stability, and economy of output to achieve realism without reality." (142) James Ross, "Immaterial Aspects of Thought," *The Journal of Philosophy* 89, no. 3 (1992): 138-42, <https://doi.org/10.2307/2026790>.

³⁴⁸ Sadler et al., *Game Changer*, 42.

The AlphaZero and AlphaGo engines have demonstrated that machines can be used to find what may be termed missing insights; they can produce results that humans have not yet discovered for themselves through insights. Humans can examine the results and perform the necessary acts of understanding to see why the results are valid.

The AlphaZero engine has also been recently repurposed to perform automatic theorem proving, and when combined with a large language model capable of converting ordinary language into mathematical formalisms, it was able to perform at the level of a silver medallist in the 2024 International Mathematics Olympiad, creating non-obvious proofs at the level of a skilled pre-university student. Although the technology is still immature at the time of writing, this suggests that neural net approaches to AI may be capable of constructing novel proofs for interesting mathematical questions in the future.³⁴⁹

There is a related sense in which machines can aid humans in seeking mathematical insights through the use of algorithms that test their intuitions. Google DeepMind created algorithms to help skilled mathematicians investigate the behaviour of knots and permutations. Several passages from the conclusion of the research paper are worth commentary.

This case demonstrates how non-trivial insights about the behaviour of large mathematical objects can be obtained from trained models, such that new structure can be discovered.

Rather than use machine learning to directly generate conjectures, we focus on helping guide the highly tuned intuition of expert mathematicians, yielding results that are both interesting and deep.

Our case studies demonstrate how a foundational connection in a well-studied and mathematically interesting area can go unnoticed, and how the

³⁴⁹ “AI Achieves Silver-Medal Standard Solving International Mathematical Olympiad Problems,” Google DeepMind, May 14, 2024, <https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/>; There have also been impressive results with a geometry proof constructing algorithm called AlphaGeometry 2. A combination of more traditional symbolic approaches and AlphaGeometry was able to collectively perform at gold medallist level in proof construction. Shiven Sinha et al., “Wu’s Method Can Boost Symbolic AI to Rival Silver Medalists and AlphaGeometry to Outperform Gold Medalists at IMO Geometry,” arXiv:2404.06405, preprint, arXiv, April 11, 2024, <https://doi.org/10.48550/arXiv.2404.06405>.

framework allows mathematicians to better understand the behaviour of objects that are too large for them to otherwise observe patterns in.³⁵⁰

The algorithm aided the researchers in discovering non-trivial insights by allowing mathematicians to gain a sense of the behaviour of the mathematical objects. The algorithm did not generate conjectures independently of a human but instead indirectly enhanced the insight of the mathematicians by acting as an external tool that allowed them to test their intuitions about mathematical objects that would otherwise be larger than they could observe. Importantly, the results the mathematicians achieved were foundational and interesting in fields that were already well-studied, but such results could not be obtained until machines augmented the conditions of insight by bringing these phenomena into human attention and imagination.

It should also be noted that there is a special class of problems, addressed by what are called non-surveyable proofs, in which machines can aid human understanding in a still further extended sense. Non-surveyable proofs are proofs which are impossible for a human mathematician to check due to the length of time required to perform the computation. In other words, they are proofs that cannot be known by humans in their entirety and that can only be known to be correct contingent on the successful operation of a machine.

A classic example of a non-surveyable proof is the Four-Colour Theorem. The Four-Colour theorem states that four colours are sufficient to colour any map such that no two bordering regions have the same colour. It was proven by exhausting cases that would have taken more than a single human lifetime to evaluate by hand. The specifics of the proof need not be addressed here. It is sufficient to observe that the proof that was conducted was such that

³⁵⁰ Alex Davies et al., “Advancing Mathematics by Guiding Human Intuition with AI,” *Nature* 600, no. 7887 (2021): 70–74, <https://doi.org/10.1038/s41586-021-04086-x>.

finding a verifying the result depended upon the use of machinery. This was the first instance of computing machinery allowing humans to determine the validity of a non-surveyable proof.

The epistemic status of such mathematical knowledge is controversial. Although the results of the Four-Colour Theorem are accepted as established mathematical knowledge, Tymoczko argues that computer-dependent proofs constitute a departure from previous mathematical knowledge by depending upon empiricism, namely the experiment of running the algorithm on a computer. The results returned depend upon the algorithm being free of errors that would produce unexpected behaviour (that is operations resulting from a difference between the internal logic of the algorithm and the abstract logic of the intended proof) and that the physical computer is operating as intended.³⁵¹

More substantively, the Four-Colour Theorem illustrates a common feature of non-surveyable proofs that is considered undesirable, namely, that they are inelegant and require an exhaustive proof by cases. A proof by cases establishes the validity of a theorem by examining all possibilities that follow from the axioms and demonstrating that in all cases the proposition to be proved is entailed. Proofs by cases are notable for lacking what mathematicians call elegance and intuitive appeal. Rather than gaining an insight distinct from and superior to the rote work of deriving the proof, the rote work comprises the whole of the proof. One might understand the individual steps in the proof but there is no higher perspective to be gained about the proof.

In effect, the class of non-surveyable proofs is a group in which it is possible to establish knowledge that given theorems are truly contingent on the successful operation and programming of the computers, but they are a class of theorems that it is not possible for human

³⁵¹ Thomas Tymoczko, "The Four-Color Problem and Its Philosophical Significance," *The Journal of Philosophy* (New York) 76, no. 2 (1979): 57–83, <https://doi.org/10.2307/2025976>.

beings to understand in the sense of grasping of the proof as a whole. Instead, one may only understand the general procedure by which the proof is established and to also understand that the proof has been established.

The use of computers for proving theorems in this fashion does have the effect of enhancing human understanding in the sense that theorems that may only be known by non-surveyable proofs would not otherwise be known to be true. Still, the use of computing machinery to create non-surveyable proofs does shift the object of understanding away from the mathematical proof itself to the algorithm used to perform the proof. That is, while a non-surveyable proof is not itself capable of being fully present to a human act of understanding, the design of the algorithm that constructs the proof may be understood (and may itself be the product of mathematical insights), and from an understanding of the algorithm one may understand the validity of the proof.

Mathematicians have also been able to gain indirect assurance of the veracity of non-surveyable proofs by using theorem-checking algorithms that are themselves surveyable to establish that the proof is valid and that an unexpected bug or failure of hardware did not result in a false report of the proof's validity.

The significance of the advent of non-surveyable computational proofs is that it became possible for humans to have knowledge of the truth of theorems for which the proofs could not themselves be understood due to the limits of the quantity of work that humans could achieve even in the timeframe afforded by a lifetime. In effect, the use of computing machinery can compensate for the indirect limitation of human understanding by transcending other human limitations. In such instances of extending human capacities, the enhancement to understanding is indirect. A human still cannot gain a holistic and simplifying view of such proofs in virtue of them being proofs by cases, but the inability to apply understanding serially to each case to

arrive at an understanding that the conclusion is true is compensated for by allowing an understanding of a formal method and the means by which it is implemented to produce an understanding that the results must necessarily be correct.

Although proving a theorem through non-surveyable means with a computer does mean that no human will be able to understand the whole proof, the nature of at least the Four-Colour Theorem is that it does not proceed from a mathematical insight nor require insight for a student to rise above a rote performance of the steps of the proof. As said before, in a proof by cases, however short or exhaustive, knowledge of the mathematical truth can only come from a completion of the rote work.

As there is no higher viewpoint to be gained from understanding the whole set of steps in the proof and it is the result that is of interest, the goal of using computer-assisted non-surveyable theorem proving is similar to that of the nineteenth century use of computers to construct logarithm tables. In each case, there was not a fundamental insight to be gained, merely mathematical truths that were desirable to be known. In both cases there is something that is valuable to be known, but there is not a higher viewpoint to be gained from the process of attaining the goal. The confirmation of the Four-Colour Theorem illustrated that lengthy computation is sufficient to establish intellectually interesting theorems and not only produce numerical values of interest.

In this vein, it is necessary to conceive of the enhancement of the faculty of understanding and the facilitation of insight from a perspective in which the distinction between an act of understanding and an act of tool use is maintained. The tool may serve the purpose of catalysing insight, but its role as subordinate needs to be properly articulated.

In the case of automatic theorem proving, the results can only be said to be understood when a human views and grasps the results. An algorithm run on a computer is ultimately a

pattern of physical processes corresponding to the steps of a discrete state machine and, like the perceptible output, only meaningful in the context of some human action, such as testing the validity of a mathematical proposition.

The results of an algorithmic proof can only be said to contribute to a greater understanding of mathematics if they are grasped by an intelligent being. The mere derivation of statement encoded in bit strings does not of itself lead to an advancement of mathematical knowledge. It is necessary for humans to know the meanings the bit strings denote, to grasp the proposition they encode and its general truth.³⁵²

The use of neural net algorithms to discover missing insights is similar to the circumstance discussed by Polanyi concerning symbol systems outrunning human understanding. There is a subtle difference, insofar as the inference engines discussed by Polanyi have been explicitly programmed and are grounded by human understanding in that sense. At the time of writing, many neural net AI systems, such as AlphaZero and AlphaGo or large language models like ChatGPT, are not fully understood in that how the systems generate outputs is opaque. Nonetheless, human engineers still understand that these systems are structures of linear algebra that can be optimised in the outputs they map to inputs. It requires human understanding to identify the problems for which AI systems are useful and to judge the mappings as useful. In the Polanyian sense, the understanding of a human is still needed for the machines to be properly said to function. To borrow some parlance from Aquinas, the use of

³⁵² “This is the difference between machine and mind. A man’s mind can carry out feats of intelligence *by aid of* a machine and also *without* such aid, while a machine can function only as the extension of a person’s body under the control of his mind... Since the control over the machine by the user’s mind is—like all interpretations of a system of strict rules—necessarily unspecifiable, the machine can be said to function intelligently only by aid of unspecifiable personal coefficients supplied by the user’s mind.” Emphasis original. Polanyi, *Personal Knowledge*, 262.

AI systems can be described as an aid to cogitation insofar as they aid the discovery of new cognitive words and images.

When the results of insight are all that matter, then for some problems a computational process can produce these results without human involvement, thus saving time and acting as an indirect enhancement understanding. When the results of insight are themselves valuable, then they prompt new insights as systems whose results outrun understanding. In these ways, modern neural nets can serve the role of external enhancements that indirectly serve to accelerate insight.

5.5 Judicious Use of Technologies to Stimulate Second-Person Relatedness

Although the paradigmatic case of joint attention is shared gaze either with the other in a dyadic relation or toward an object of mutual awareness in a triadic relation, it is possible for purely textual communications to allow a qualified mode of shared awareness. The act of pointing to something and picking it out as the object of one's attention to invite others to attend to it as well, is performed metaphorically in the selection of subject matter discussed by the text. The author imagined the object and disclosed their explicit thoughts concerning it, with word choice, implied tone, and other facets of rhetorical style and conversational implicature substituting for tone and body language in communicating emotional stance and unstated aspects of one's orientation toward the object. If it is possible to have a sense of shared awareness with an author by reading their text, then it should in principle be possible to gain a shared awareness with another person of some sort through any act of communication when there is mutual interaction, for then there is also mutual awareness to some extent.

Crucially, such forms of communication as letter writing requires attention and engagement of the imagination to gain a sense of the other person's mental state, to enter

alignment with that person, to approximate the experience of being with another person, seeing their posture, hearing their inflection, and physically or imaginatively turning their attention to some object. Communication through the medium of text, nonetheless, does not necessarily lead to the same degree of emotional intimacy, at least, that is achieved from direct in-person interaction. According to the Ethnographer of Technology Sherry Turkle, relationships that occur primarily through online interactions tend to feature less emotional bonding than those formed from in-person interactions.³⁵³ Although reduced emotional bonding implies reduced sharing of stance with another, it does not follow that relationships depending upon such communication are completely lacking in the benefits that accrue to second-person awareness, but it does indicate that there are limits to the efficacy of at least common modes of interaction via modern technology.

Forms of communication that do not require this degree of imaginative engagement with other by supplying some more direct form of awareness can run the risk of impairing second-person awareness. There exists the risk, for instance, that video chat services can be designed in such a way that low resolution video and audio make it difficult to determine facial expression and intonation while simultaneously prompting one to attend to the other person as if these inarticulate facets of communication were not impaired. Ironically, this can result in less nuanced communication than if both parties had simply produced text with the awareness that any nonexplicit communication needed to be expressed through the words.

The market demand during the pandemic for video conferencing tools that afforded better audiovisual resolution and reduced latencies illustrates the human preference for methods of communication that facilitate second-person awareness of conversation partners.

³⁵³ Sherry Turkle, *Reclaiming Conversation: The Power of Talk in a Digital Age* (Penguin Press, 2015), 23.

Nonetheless, notable limitations on facilitating second-person relatedness between the conversational partners do exist. For instance, the design constraints of commercially available video conferencing software are unable to simulate direct eye contact because the camera used to record the user is not in the same location as the onscreen representation of the other person's eyes, which means that even virtual gaze sharing is not possible.

The importance of implicit communication that is not easily captured in video conferencing suggests the possibility that advances in virtual reality could allow a better approximation of embodied interaction. Technology will need to allow faithful representation of posture, facial expression, and gaze following for a successful simulation of embodied experience of another. Using the technology most effectively, however, it is necessary to be mindful of its limitations. Every avatar is a representation of quantitative measurements taken of the other person's physical shape, body posture, and facial contours among other attributes. To the extent that such representations are necessarily abstractions there is the possibility that some aspects of embodied communication would be omitted. This is not to deny that virtual reality communication could facilitate second-person relatedness and the alignment of awareness that catalyses insights. Rather, the role of virtual reality should be seen as supplementary to embodied communication.

Sherry Turkle gives a more recent example of a use of virtual reality that had the potential to allow some understanding of others' experiences but was deployed in a fashion she considers counterproductive. Specifically, a virtual reality apparatus was set up to allow participants to virtually experience being in an impoverished African village for the stated purpose of eliciting empathy for the villagers. She reported that there was a positive response among participants. She also observed, however, that the response to the initiative generally lacked a consideration about the limits of the virtual reality experience in contrast to the

experience that would result from making the journey and partially sharing the experience of living in the region of the world depicted by the simulation. She observed the danger that a technologically facilitated substitute for a given experience originally considered “better than nothing” can have its deficits overlooked and be appropriated as equivalent to the original experience or, in some cases, superior to the original.³⁵⁴

Turkle has observed that there was a further risk that promoters of the technology could present the substitute experience as superior to the original one. She cited the example of a colleague who claimed that offering psychotherapy to patients through video calls was superior to practicing in person because they were able to accomplish more with their patients in less time. She noted that studies show that stress levels are often higher during video calls and that her colleague’s evaluation of the superiority of remote rather than physically proximate practice may have been due primarily to the increase in focus stimulated by the stress. In other words, her colleague would perhaps have performed still better if both equally focused and physically present to the patients. Turkle also noted a resurgence of interest in creating language parsers to offer automated counselling. These cases indicate persistent and recurring interest in the use of technologies to replicate or increase the benefits that result from second-person interactions with other persons. These cases also illustrate the possibility, however, that technologies that do aid one in acquiring the benefits typically associated with second-person interactions with other persons do not necessarily allow one to acquire all the benefits of second-person interactions. The ability to gain some benefits of second-person awareness for insight in the

³⁵⁴ Sherry Turkle, “Reclaiming Conversation: Our New Silent Spring in a Digital Age,” University of Cambridge, 2016.

absence of actual physical proximity does represent an enhancement of human capacity as long as they are not used in preferment to embodied interaction.³⁵⁵

The technology will be most effective when the limitations of the experiences it can produce are acknowledged. One instructive example is the reception of Joseph Weizenbaum's early chatbot, ELIZA, and the attempt to employ it for psychotherapy. Language parsing algorithms do provide many uses such as serving as intuitive UI's and enabling language translation. Persons who encountered one of the first language parsers, ELIZA, which was designed to engage in teletype conversations under the guise of being a psychoanalyst, asked its creator how it might be employed in psychoanalytic therapy. He subsequently argued that the inability of ELIZA to actually understand the person interacting with it meant that the conversations that users could have with ELIZA would not offer the benefits of speaking with a real therapist. A psychotherapist would describe their work as coming to an understanding of the patient's psychological ills and guiding the conversation in a way that would lead to healing or at least improved function for the patient. Weizenbaum's point was that those who thought that ELIZA could meaningfully be used in therapy were impressed by a superficial appearance of comprehension. The simple language parser could not understand the patient's psychological needs or communicate them to the patient.

5.6 Concluding Remarks on Human-Machine Cooperation

As noted in chapter three, possibilities for accelerating insight directly through internal enhancements do not at present seem to be well grounded. Nevertheless, this judgement does not necessarily exclude opportunities for accelerating insight indirectly, especially by means of

³⁵⁵ Turkle, "Reclaiming Conversation: Our New Silent Spring in a Digital Age."

external enhancements. These enhancements are over and above the completion of the huge range of labour-saving mechanistic cognitive tasks at which machines can excel.

First, external enhancements can be used for bringing otherwise inaccessible phenomena into the reach of human experience, to the point that the raw materials of human understanding can engage with them. For example, simulations of phenomena that would otherwise be inaccessible such as hypercubes or other higher dimensional objects can be made accessible to kinaesthetic perception and manipulation. Additionally, external enhancements can bring aspects of the physical world into human scrutiny that require superhuman speed to record such as the measurements taken by ATLAS at CERN. These significant advances benefits to understanding do not require either the creation of new capacities or direct modification to the human faculty of understanding but instead bring the world into reach of those capacities.

Second, there are many situations in which only the results matter of what would be obtained, by human beings, through acts of understanding. In these circumstances, it is irrelevant whether a machine actually understands, provided a computational process exists for generating the results. As one important example, classifying the shapes of galaxies was long believed to be an exclusively an act of human understanding, but algorithms exist today that generate faster results with a high degree of confidence.

Third, there are what may be termed missing insights that are valuable to understand which machines can aid humans in discovering. In the examples of AlphaGo and AlphaZero, the unexpected results generated from the algorithms have in fact given rise to understanding influential new strategies in Go and Chess. Other algorithms have assisted mathematicians in testing and verifying otherwise inaccessible insights, and there are some mathematical truths

that are non-surveyable, which means they can only be understood as a corollary to the successful to the successful operation of a machine.

Fourth, communication technology can facilitate interactions between geographically separated persons that approximates embodied second-person interactions, hence facilitating insights. Similarly, the users of complex systems often like to communicate with these systems using interfaces that simulate human conversation, such as chatbots. This mode of communication is easy and natural, improving efficiency and possibly opening up cognitive opportunities for new insights. Obviously, this facilitation is not without some risk, not least that the user can project an illusion of genuine personhood onto the machine and may also be mistaken in terms of some of the insights generated.

In conclusion, despite the poor grounding of transhumanist prospects for enhancing human insight directly, the transhumanists are not altogether wrong that contemporary technologies can and, in many cases, do exist for accelerating insights indirectly.

CHAPTER 6 Conclusions

6.1 Summary and Thesis Conclusions

This thesis has shown that, although the transhumanists are intensely interested in augmenting understanding and accelerating insight as part of their larger project of enhancing cognition, they show at best a partial awareness of the phenomena they seek to enhance.

The first and second chapters showed that the transhumanists acknowledge that what is commonly called understanding is ubiquitous in human cognition and that insight is inseparable from generating new ideas and solving problems innovatively. Nonetheless, the application of the phrase “I understand” to such a wide variety of concrete cognitive actions about which one can say no more than “I see it” or “I grasp it” is difficult for them to account.

Marvin Minsky acknowledges that the terms “understanding” and “insight” refer to cognitively important phenomena, but he denies that understanding is a distinct faculty or that an insight is an operation of this faculty. Instead, he uses these as terms of convenience while positing that many distinct processes and structures do the work ascribed to understanding.

Ray Kurzweil similarly identifies understanding as of fundamental importance to cognition but denies understanding is a distinct faculty, instead interpreting its fruits as the output of pattern recognition systems in the neocortex. “Insight” just denotes the successful operation of such a system. Kurzweil also argues that the outputs of a system in response to an input, a question, say, can warrant ascribing understanding to a system, if the responses it gives are sufficiently similar to a human response.

Nick Bostrom identifies understanding and insight as essential to the success of future augmented intelligences, specifically of transhumans and posthumans, but his focus is less on

what understanding and insight are than on what projections can be made about how they, along with the rest of human cognition, can expect to be enhanced through future technologies.

The transhumanist aspirations for enhancing understanding and insight, such as they are, are assumed to be rooted in some kind of augmentation of the brain. Minsky predicts opportunities to expand cognition by reverse engineering the general structure of the brain and learning how to encode the cognitive contents of particular brains in a form that can be simulated by computers. Kurzweil, in addition to predicting it will be possible to encode the content of a mind into a computer program, predicts specifically that there will be an artificial expansion of the neocortex. Nick Bostrom explores possibilities of creating superintelligence through multiple pathways: accelerating the rate of cognition for individual agents, increasing the number of thinking agents, and devising intelligences qualitatively superior to humans. All these prospective augmentations of intelligence assume that cognitive faculties, including understanding, can be fully represented in computer code, ultimately as finite bit strings. These accounts of understanding and insight, therefore, assume understanding and insight can be exhaustively captured in descriptions of data structures. They do not discuss how understanding and insight may be enhanced specifically rather than just as derivatives of a more general reverse engineering and re-engineering of intelligence.

The third chapter of the thesis showed that when there is a change in understanding there is a specific term used to describe the cognitive shock of new, sudden understanding, namely insight. Psychological research explores insight in the context of “Aha!” experiences, also called “Eureka!” experiences after the famous story of Archimedes’ discovery in the baths of Syracuse. Psychological research identifies insightfulness with the shifting of attention to what is relevant to solve a problem and recognising a coherent relationship between the materials of the test. Crucially, the insight is distinct from the problem-solving, although it can be elicited

by the need to solve a problem. More importantly, the sudden awareness that comes from an insight is arrived at suddenly rather than discursively. An insight is experientially similar to a gestalt shift, in which a new whole is perceived suddenly without intermediate steps.

Further investigation of the phenomena of insight from philosophical perspectives drew out five key themes concerning insights: insights are singular ascending operations; insights are grounded in embodiment; insights are found through metaphors; insights are communicated in a privileged way by narratives; and insights are catalysed by second-person relatedness.

Theology exhibits and supports these key themes. The singular, ascending operation of insight is portrayed in the metaphor of divine illumination and sudden seeing. The psalms of the Old Testament and Jesus' own use of parables illustrate the capacity of narrative to convey understanding from one person to another. The privileged communicative power of narratives vis-à-vis insight relies on and further illustrates the fundamental cognitive significance of embodiment and metaphor for understanding as exemplified by psalms and parables' use of images and descriptions to elicit new comprehension. Finally, depictions of personal encounters with Christ and transformations of the understanding, most famously the conversion of Paul, provide illustrative dramatic instances of the power of second-person encounters for effecting cognitive change in the form of sudden insight.

The fourth chapter examines the transhumanists' perspective on understanding and insight in the light of the philosophical, theological, and psychological analysis of chapter two. To begin, the transhumanists agree that the subjective experience of having understanding and gaining insights is that of having a simple, pervasive cognitive grasp, but they deny that this experience corresponds to an actual simplicity or distinctness to the faculty. Instead, they posit that to experience insight is just to experience the output of problem-solving algorithms in the brain, and the immediacy of understanding simply is the result of having forgotten or being

unaware of the processes that produced these solutions as the result of previous inquiry. Most especially, understanding and insight are both described in terms of representations. On this account, to understand something is just to have a sufficiently detailed and well-structured mental representation, and to have insight means simply to reformulate one's representations. They deny there is a distinct faculty of understanding or that insight is an operation of it.

The transhumanists also agree that embodiment is in some sense cognitively significant, for many concepts are defined in relation to the body, and much of the content of everyday cognition concerns the body. They even acknowledge that hypothetical posthuman entities might have some sort of embodiment. The transhumanists, however, do not tend to treat having one's particular body, one's particular *kind* of body, or even having a body at all, as being of intrinsic importance to human cognition. On their account, the body makes only an accidental contribution, for example, in the acquisition of local spatial properties. Indeed, the transhumanists consider that the cognitive benefits of embodiment could be supplied equally well by a kind of prosthesis, either in the form of a robot or, especially, through simulated experience of a virtual world, which they claim will eventually be indistinguishable from the physical world. Moreover, they believe one's own body may be something that needs to be jettisoned like the first stage of a rocket so as to exist as a being of pure information with purely virtual embodiment. The underlying assumption of their perspective on embodiment and its role in augmenting cognition is that all cognitive content and the world itself can be exhaustively represented, ultimately in finite bit strings, and capable of being modelled in a computer such that the model is interchangeable with the thing represented.

The transhumanists agree that metaphor is pervasive in human thought, even to the point of most human cognition being in some way metaphorical. The transhumanists, however, must conceptualise metaphor differently given their stance on embodiment. Rather than metaphor

being the link between utterance and embodied experience that makes language (and, by extension, other forms of representation) meaningful, metaphor is instead explained as a pattern recognition between different representations rather than an experience of similarity between things. Such accounts being, in effect, closed systems of representation that do not explain why the circles of reference are meaningful rather than arbitrary.

The transhumanists regard narrative as cognitively useful and show practical capacity for using it didactically, but their examination of narrative does not refer to its capacity to convey a sense of experience through time, something like an extended metaphor, but instead as a kind of pattern of information that is primarily useful for storing sequential information to highlight deviations from expectation. The communicative aspect of narrative is described as a sequence of utterances that activate representations in the mind of the listener such that it is a parsimonious form of communication that forgoes transmitting information in favour of pointing to already extant information in the listener. Absent from this account is a sense of shared engagement with the world as embodied beings.

The transhumanists have concepts that overlap with the notion of second-person relatedness or joint attention in the context of cognition, but they largely ignore the psychological state of shared awareness. This is well-illustrated by their discussions of pedagogy. The teacher is metaphorically described as transmitting programs and data to a student, but when the programming metaphor has been stripped away, nothing has been said about how the teacher conveys information. In one example, a machine highlighting text to read is described as doing the same thing a caregiver does when running their finger along text with a small child. The invitation to enter a shared stance towards what is being taught and the aligning of the pupil with the teacher is not discussed in their accounts. The second-person aspects of understanding concern that which is implicit and not amenable to abstraction.

6.2 Outcomes and Implications

The transhumanists' prospects for enhancing understanding directly are unfavourable. At present, transhumanists cannot even take account of many broadly acknowledged phenomena associated with understanding and insight. Instead, they generally attempt methodological reductions of such phenomena to the operations of components of information processing that do not themselves understand. I therefore conclude that the transhumanist accounts of understanding and insight are at present far too incomplete to offer specific, measurable, achievable, realistic, or timely prospects for enhancement. Consequently, transhumanist aspirations for accelerating insight directly are likely to remain of little value.

An Aristotelian-Thomist perspective brings the deficiencies of the transhumanist anthropology into sharp relief. The transhumanist aspiration to become posthumans relies on a computationalist interpretation of insight that requires them to find an equivalence between abstract representation and the implicit that leaves representation without adequate grounding. An Aristotelian-Thomist interpretation of insight, in contrast, views the human person as a unified being whose intellect is embodied, embedded in the actual and social world. This interpretation is more congruent with the ideas of insight in such diverse areas of human intellectual activity as psychology, studies of metaphor, and studies of narrative. While this does not constitute a strict proof of an Aristotelian-Thomist anthropology, it suggests that such an anthropology does match the human reality better than the transhumanists' anthropologies. It also indicates that the interpretations of insight given by those in a broadly Aristotelian-Thomist tradition are more persuasive, and will be more helpful when we come to evaluate what techniques of enhancement truly benefit humanity, and identify the limitations of such techniques.

This thesis has shown that the philosophical and theological perspectives drawn from Lonergan, Polanyi, and Aquinas better elaborates upon insight and understanding than the transhumanist perspectives. The philosophical and theological perspective sees insight as a unified act grounded in human embodiment but extending from linguistic and kinaesthetic sense of metaphor to the making and sharing of narrative. It embraces insight as personal, both in the first-person perspective and as relational in the second-person perspective. This account can offer grounded and concrete guidance as to how technology may be used to benefit human insight. It can also offer a picture of how character formation and the training of an implicit sense and taste for approaches to thinking and living can be allow technology to be used for flourishing.

More work remains to be done in this area, and future research can be done into human flourishing and the specific contexts in which technologies can be used for the flourishing of human beings. There are as many opportunities for developing accounts of the useful and virtuous use of technology as there are applications of technology in the world.

Fortunately, as discussed in chapter five, this assessment does not exclude all prospects for using technology, including that which is currently available, for enhancing understanding indirectly. For example, a prudent use of technology can benefit many concomitants of understanding such as by using tools and models to bring phenomena into the reach of sensible experience, accessible to metaphoric understanding. Communication technologies can also be designed to promote second-person interactions that can assist in catalysing insights. Modern advances in neural network-based AI have also shown another possibility. It is possible for a machine, without it being assumed to understand, to produce outputs with a great speed and reliability that a human would ordinarily require insight to create. As a consequence, in cases in which insight itself is not required directly, but only the fruits of insight, such as in the

classification of galaxies from astronomical imagery, it is possible to free the human intellect from such tasks and allow persons to devote themselves to more worthwhile endeavours.

In conclusion, the transhumanist aspirations for accelerating insight directly may remain unwarranted, at least at present. As this thesis has demonstrated, there is, a rapid contemporary expansion in the appreciation of understanding and insight across multiple disciplines, including psychology, neuroscience, philosophy, and theology. By freeing the human intellect from trivial but time-consuming acts in which only the results of understanding matter, combined with a greater appreciation of the faculty of understanding in general, all kinds of ways plausibly remain in which new technologies will enhance understanding and insight indirectly.

Deo Gratias

Sean Biggins, University of Oxford

October 2025

Bibliography

- Alter, Robert. *The Art of Biblical Narrative*. Ebook Central. Basic Books, a member of the Perseus Books Group, 2011. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=679754>.
- Aquinas, Thomas. *Catena Aurea: Commentary on the Four Gospels*. With John Henry Newman. III Saint Luke. The Saint Austin Press, 1997.
- Aquinas, Thomas. *Commentary on the Gospel of John*. Translated by James A. Weisheipl, O.P. Magi Books, Inc., 1998.
- Aquinas, Thomas. *New English Translation of St. Thomas Aquinas's Summa Theologiae*. Translated by Alfred J. Freddoso. n.d. Accessed September 28, 2025. <https://www.freddoso.com/summa-translation/TOC.htm>.
- Aquinas, Thomas. *On the Unity of the Intellect against the Averroists*. Translated by Beatrice H. Zedler. Mediaeval Philosophical Texts in Translation; No. 19. Marquette University Press, 1968.
- Aristotle. "Metaphysics." In *The Complete Works of Aristotle*, Revised edition, edited by Jonathan Barnes, translated by W. D. Ross, vol. 2. Princeton University Press, 1984.
- Aristotle. "Poetics." In *The Complete Works of Aristotle*, Revised edition, edited by Jonathan Barnes, vol. 2. Princeton University Press, 1984.
- Armstrong, Stuart, Anders Sandberg, and Nick Bostrom. "Thinking Inside the Box: Controlling and Using an Oracle AI." *Minds and Machines* 22, no. 4 (2012): 299–324. <https://doi.org/10.1007/s11023-012-9282-2>.
- Austen, Jane. *Mansfield Park*. Edited by James Kinsley. Oxford's World's Classics. Oxford University Press, 2003.
- Baker, David, Zoran Popovi, Janos Barbero, et al. "Predicting Protein Structures with a Multiplayer Online Game." *Nature* (London) 466, no. 7307 (2010): 756–60. <https://doi.org/10.1038/nature09304>.
- Beahm, George, ed. *I, Steve: Steve Jobs In His Own Words*. Later Printing edition. Agate B2, 2011.
- Bennett, M. R., and P. M. S. Hacker. *Philosophical Foundations of Neuroscience*. Blackwell Publishing Ltd, 2003.
- Block, Ned. "Psychologism and Behaviorism." *The Philosophical Review* 90, no. 1 (1981): 5–43. <https://doi.org/10.2307/2184371>.
- Bostrom, N. "The Fable of the Dragon Tyrant." *Journal of Medical Ethics* 31, no. 5 (2005): 273–77. <https://doi.org/10.1136/jme.2004.009035>.
- Bostrom, Nick. *Deep Utopia: Life and Meaning in a Solved World*. Ideapress Publishing, 2024.
- Bostrom, Nick. "In Defense of Posthuman Dignity." *Bioethics* 19, no. 3 (2005): 202–14. <https://doi.org/10.1111/j.1467-8519.2005.00437.x>.
- Bostrom, Nick. "Letter from Utopia." *Studies in Ethics, Law, and Technology* 2, no. 1 (2008). <https://doi.org/10.2202/1941-6008.1025>.
- Bostrom, Nick. "Quantity of Experience: Brain-Duplication and Degrees of Consciousness." *Minds and Machines* 16, no. 2 (2006): 185–200. <https://doi.org/10.1007/s11023-006-9036-0>.
- Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. First edition. With ProQuest. Ebook Central. Oxford University Press, 2014. <https://ezproxy->

- prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=1743975.
- Bostrom, Nick. "Transhumanist FAQ." Humanity+. Accessed October 16, 2019. <https://humanityplus.org/philosophy/transhumanist-faq/>.
- Bostrom, Nick. "Transhumanist Values." *Journal of Philosophical Research* 30 (2005): 3–14. https://doi.org/10.5840/jpr_2005_26.
- Bostrom, Nick, and Rebecca Roache. "Smart Policy: Cognitive Enhancement and the Public Interest." In *Enhancing Human Capacities*, 1st ed., edited by Julian Savulescu, Ruud Ter Meulen, and Guy Kahane. Wiley, 2011. <https://doi.org/10.1002/9781444393552.ch9>.
- Bostrom, Nick, and Anders Sandberg. "Converging Cognitive Enhancements." *Annals of the New York Academy of Sciences* 1093, no. 1 (2006): 201–27. <https://doi.org/10.1196/annals.1382.015>.
- Bostrom, Nick, and Anders Sandberg. *Whole Brain Emulation: A Roadmap (Technical Report 2008-3)*. The Future of Humanity Institute, 2008. www.fhi.ox.ac.uk/reports/2008-3.pdf.
- Bostrom, Nick, and Eliezer Yudkowsky. "The Ethics of Artificial Intelligence." In *The Cambridge Handbook of Artificial Intelligence*, edited by Keith Frankish and William M. Ramsey. Cambridge University Press, 2014.
- Bowden, Edward M., Mark Jung-Beeman, Jessica I. Fleck, and John Kounios. "New Approaches to Demystifying Insight." *Trends in Cognitive Sciences* 9, no. 7 (2005): 322–28. <https://doi.org/10.1016/j.tics.2005.05.012>.
- Brock, Stephen L. *The Philosophy of Saint Thomas Aquinas: A Sketch*. 1st ed. Cascade Books, 2015.
- Burdett, Michael S. "Contextualizing a Christian Perspective on Transcendence and Human Enhancement." In *Transhumanism and Transcendence: Christian Hope in an Age of Technological Enhancement*, edited by Ronald Cole-Turner. Georgetown University Press, 2011.
- Burdett, Michael S. *Eschatology and the Technological Future*. Routledge Studies in Religion ; 43. Routledge, 2015. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://www.taylorfrancis.com/books/9781315739397>.
- Callaway, Ewen. "'It Will Change Everything': DeepMind's AI Makes Gigantic Leap in Solving Protein Structures." *Nature* (BERLIN) 588, no. 7837 (2020): 203–4. <https://doi.org/10.1038/d41586-020-03348-4>.
- Callaway, Ewen. "'The Entire Protein Universe': AI Predicts Shape of Nearly Every Known Protein." *Nature (London)* (London) 608, no. 7921 (2022): 15–16. <https://doi.org/10.1038/d41586-022-02083-2>.
- Davies, Alex, Petar Veličković, Lars Buesing, et al. "Advancing Mathematics by Guiding Human Intuition with AI." *Nature* 600, no. 7887 (2021): 70–74. <https://doi.org/10.1038/s41586-021-04086-x>.
- Davis, Philip J., and Reuben Hersh. *The Mathematical Experience*. Birkhauser, 1982.
- De Haan, Daniel D. "Approaching Other Animals with Caution: Exploring Insights from Aquinas's Psychology." *New Blackfriars* 100, no. 1090 (2019): 715–37. <https://doi.org/10.1111/nbfr.12509>.
- Dennett, Daniel C. *Brainchildren: Essays on Designing Minds*. With Internet Archive. London: Penguin, 1998. <http://archive.org/details/brainchildreness0000denn>.

- Dennett, Daniel C. "Can Machines Think?" In *Language and Meaning in Cognitive Science: Cognitive Issues and Semantic Theory*. Artificial Intelligence and Cognitive Science, edited by Josef Toribio and Andy Clark, vol. 4. Garland Publishing, Inc., 1998.
- Dennett, Daniel C. *Kinds of Minds: Towards an Understanding of Consciousness*. Science Masters Series. Phoenix, 1997.
- Dreyfus, Hubert L. *What Computers Still Can't Do: A Critique of Artificial Reason*. Rev Ed edition. MIT Press, 1999.
- Eberl, Jason T. "Enhancing the Imago Dei: Can a Christian Be a Transhumanist?" *Christian Bioethics: Non-Ecumenical Studies in Medical Morality* 28, no. 1 (2022): 76–93. <https://doi.org/10.1093/cb/cbab016>.
- Eberl, Jason T. "Who Wants to Live Forever? Transhumanist Immortality and Christian Eternity." *Christian Bioethics*, ahead of print, 2025. <https://doi.org/10.1093/cb/cbaf004>.
- Eilan, Naomi, Christoph Hoerl, Teresa McCormack, and Johannes Roessler, eds. *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology*. 1st edition. Oxford University Press, 2005.
- Fleck, Jessica I., Mark Beeman, and John Kounios. "Insight." In *The Oxford Handbook of Cognitive Psychology*, 1 edition, edited by Daniel Reisberg. Oxford University Press, 2013.
- Google DeepMind. "AI Achieves Silver-Medal Standard Solving International Mathematical Olympiad Problems." May 14, 2024. <https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/>.
- Grimm, Stephen. "Understanding." In *The Stanford Encyclopedia of Philosophy*, Summer 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2021. <https://plato.stanford.edu/archives/sum2021/entries/understanding/>.
- Hadamard, Jacques. *The Psychology of Invention in the Mathematical Field*. Princeton University Press, 1945.
- Harrison, Harry. *The Turing Option*. With Marvin Minsky. Gateway, 2013.
- Hill, David J. "Exclusive Interview: Ray Kurzweil Discusses His First Two Months At Google." *Singularity Hub*, March 19, 2013. <https://singularityhub.com/2013/03/19/exclusive-interview-ray-kurzweil-discusses-his-first-two-months-at-google/>.
- Hobson, Peter. *The Cradle of Thought: Exploring the Origins of Thinking*. Reprints edition. Pan, 2004.
- Hobson, Peter. "What Puts Jointness Into Joint Attention?" In *Joint Attention: Communication and Other Minds*, edited by Naomi Eilan, Christoph Hoerl, Teresa McCormack, and Johannes Roessler. Issues in Philosophy and Psychology 3. Oxford University Press, 2005.
- Humanity+. "Transhumanist FAQ." Accessed September 20, 2022. <https://www.humanityplus.org/transhumanist-faq>.
- Huxley, Julian. *New Bottles for New Wine*. Chatto and Windus, 1957.
- Jackson, Frank. "Epiphenomenal Qualia." *The Philosophical Quarterly (1950-)* 32, no. 127 (1982): 127–36. <https://doi.org/10.2307/2960077>.
- Jackson, Frank. "Epiphenomenal Qualia." *The Philosophical Quarterly (1950-)* 32, no. 127 (1982): 127–36. <https://doi.org/10.2307/2960077>.
- James, Elaine T. *An Invitation to Biblical Poetry*. Essentials of Biblical Studies Series. Oxford University Press, 2022. <https://doi.org/10.1093/oso/9780190664923.001.0001>.

- Jones, Gregory L. "Rhetoric, Narrative, and the Rhetoric of Narratives: Exploring the Turns to Narrative in Recent Thought and Discourses." *Issues in Integrative Studies* 11, no. 1 (1993): 7–25.
- Jumper, John, Richard Evans, Alexander Pritzel, et al. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* (BERLIN) 596, no. 7873 (2021): 583–89.
<https://doi.org/10.1038/s41586-021-03819-2>.
- Keenan, Oliver. "Semantic Dogmatics in Outline: Cornelius Ernst's 'meaning of Meaning' as Michael Polanyi's Absent Doctrine of God." 2018.
- Kounios, John, Jessica I. Fleck, Deborah L. Green, et al. "The Origins of Insight in Resting-State Brain Activity." *Neuropsychologia* 46, no. 1 (2008): 281–91.
<https://doi.org/10.1016/j.neuropsychologia.2007.07.013>.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions: 50th Anniversary Edition*. Fourth edition. With Ian Hacking. University of Chicago Press, 2012.
- Kurzweil, Ray. "Deep Fritz Draws: Are Humans Getting Smarter, or Are Computers Getting Stupider? « Kurzweil." Kurzweilai. Accessed November 16, 2022.
<https://www.kurzweilai.net/deep-fritz-draws-are-humans-getting-smarter-or-are-computers-getting-stupider-2>.
- Kurzweil, Ray, dir. *Get Ready for Hybrid Thinking*. n.d. Accessed March 24, 2019.
https://www.ted.com/talks/ray_kurzweil_get_ready_for_hybrid_thinking.
- Kurzweil, Ray. *How to Create a Mind: The Secret of Human Thought Revealed*. Duckworth Overlook, 2013.
- Kurzweil, Ray. "Ray Kurzweil Response to Searle Review." October 31, 2005.
https://web.archive.org/web/20051031003301/http://www.kurzweiltech.com/Searle/searle_response_3.htm.
- Kurzweil, Ray. *The Age of Intelligent Machines*. MIT Press, 1990.
- Kurzweil, Ray. *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines*. Orion Business, 1999.
- Kurzweil, Ray. *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines*. Texere, 2001.
- Kurzweil, Ray. *The Singularity Is Near*. Gerald Duckworth & Co, 2010. Pro Quest.
<https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=2029217>.
- Kurzweil, Ray, and John R. Searle. "'I Married a Computer': An Exchange." *The New York Review of Books*, May 20, 1999.
- Lakoff, George, and Mark Johnson. *The Metaphors We Live By*. University of Chicago Press, 2003.
- Lonergan, Bernard. *Insight: A Study of Human Understanding, Volume 3*. 5th edition. Edited by Frederick Crowe S.J and Robert Doran S.J. University of Toronto Press, 1992.
- Ludwig Wittgenstein. *Philosophical Investigations*. 1958.
http://archive.org/details/philosophicalinvestigations_201911.
- McCabe, Herbert. *The Teaching of the Catholic Church: A New Catechism of Christian Doctrine*. Darton, Longman & Todd, 2000.
- McGilchrist, Iain. *The Master and His Emissary: The Divided Brain and the Making of the Western World*. Yale University Press, 2009.
- Minsky, Marvin. "A Framework for Representing Knowledge." In *The Psychology of Computer Vision*, edited by Patrick Henry Winston. McGraw-Hill Inc., 1975.

- Minsky, Marvin. "Afterword to Vernor Vinge's Novel, 'True Names'." Marvin Minsky's Homepage, October 1, 1984.
<https://web.media.mit.edu/~minsky/papers/TrueNames.Afterword.html>.
- Minsky, Marvin. *Inventive Minds: Marvin Minsky on Education*. With Cynthia Solomon and Xiao Xiao. Cambridge, Mass: MIT Press, 2019.
- Minsky, Marvin. *Semantic Information Processing*. MIT Press, 1968.
- Minsky, Marvin. "Society of Mind: A Response to Four Reviews." *Artificial Intelligence* 48, no. 3 (1991): 371–96. [https://doi.org/10.1016/0004-3702\(91\)90036-J](https://doi.org/10.1016/0004-3702(91)90036-J).
- Minsky, Marvin. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Reprint Edition. Simon & Schuster, 2007.
- Minsky, Marvin. *The Society of Mind*. Touchstone. Simon & Schuster, 1988.
- Minsky, Marvin. "Will Robots Inherit the Earth?" *Scientific American*, October 1994.
<https://web.media.mit.edu/~minsky/papers/sciam.inherit.html>.
- Mirowski, Philip. *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge University Press, 2002.
- Moravec, Hans. *Mind Children: The Future of Robot and Human Intelligence*. Reprint edition. Harvard University Press, 1990.
- Nussbaum, Martha C. *Poetic Justice: The Literary Imagination and Public Life*. With Internet Archive. Beacon Press, 1995. <http://archive.org/details/poeticjusticelit0000nuss>.
- Ochs, Elinor, and Lisa Capps. *Living Narrative: Creating Lives in Everyday Storytelling*. Harvard University Press, 2022. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://www.degruyter.com/isbn/9780674041592>.
- Park, Clara Claiborne. *The Siege: A Family's Journey Into the World of an Autistic Child*. New edition. Little, Brown and Company, 1982.
- Pentland, Alex. *Social Physics: How Social Networks Can Make Us Smarter*. Reissue edition. Penguin Books, 2015.
- Pinsent, Andrew. *The Second-Person Perspective in Aquinas's Ethics: Virtues and Gifts*. 1st ed. Routledge, 2013.
- Pinsent, Andrew, and Sean Biggins. "Catholic Perspectives on Human Biotechnological Enhancement." *Studies in Christian Ethics*, February 8, 2019. <https://ezproxy-prd.bodleian.ox.ac.uk:2246/doi/full/10.1177/0953946819826769>.
- Plato. *Republic*. Translated by G. M. A. Grube and C. D. C. Reeve. Ebook Central. Hackett PubCo, 1992. <https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=735561>.
- Poincaré, Henri. *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method*. Translated by George Bruce Halsted. 2012.
<http://www.gutenberg.org/ebooks/39713#download>.
- Polanyi, Michael. *Personal Knowledge: Towards a Post-Critical Philosophy*. Corrected Edition. University Of Chicago Press, 1974.
- "Ray Kurzweil - Director of Engineering, Google Inc., Alphabet Inc." *Boardroom Insiders Profiles; San Francisco*, July 10, 2017.
<https://search.proquest.com/docview/1933131584/abstract/6DB48C598DA44A84PQ/9>.
- Reddy, Vasudevi, and Sebo Uithol. "Engagement: Looking beyond the Mirror to Understand Action Understanding." *British Journal of Developmental Psychology* (England) 34, no. 1 (2016): 101–14. <https://doi.org/10.1111/bjdp.12106>.

- Reisberg, Daniel, ed. *The Oxford Handbook of Cognitive Psychology*. 1 edition. Oxford University Press, 2013.
- Ross, James. "Immaterial Aspects of Thought." *The Journal of Philosophy* 89, no. 3 (1992): 136–50. <https://doi.org/10.2307/2026790>.
- Sadler, Matthew, Natasha Regan, Garry Kasparov, and Demis Hassabis. *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*. New In Chess, 2019.
- Sand, George, Gustave Flaubert, and Aimée Leffingwell McKenzie. *The George Sand-Gustave Flaubert Letters*. With Robarts - University of Toronto. London, Duckworth, 1922. <http://archive.org/details/georgesandgustav00sand>.
- "Scientists' Open Letter on Cryonics – Biostasis." *Biostasis*, n.d. Accessed January 23, 2019. <https://www.biostasis.com/scientists-open-letter-on-cryonics/>.
- Searle, John R. "I Married a Computer." *The New York Review of Books*, April 8, 1999.
- Seemann, Axel. "What Is Joint Attention? Comparative, Developmental, and Definitional Issues." In *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, edited by Axel Seemann. MIT Press, 2012.
- Sertillanges, OP A. G., and SJ James V. Schall. *The Intellectual Life: Its Spirit, Conditions, Methods*. Reprint edition. Translated by Mary Ryan. The Catholic University of America Press, 1998.
- Sinha, Shiven, Ameya Prabhu, Ponnurangam Kumaraguru, Siddharth Bhat, and Matthias Bethge. "Wu's Method Can Boost Symbolic AI to Rival Silver Medalists and AlphaGeometry to Outperform Gold Medalists at IMO Geometry." arXiv:2404.06405. Preprint, arXiv, April 11, 2024. <https://doi.org/10.48550/arXiv.2404.06405>.
- Stump, Eleonore. *Wandering in Darkness: Narrative and the Problem of Suffering*. Oxford University Press, 2010.
- Ted Peters. "Progress and Provolution: Will Transhumanism Leave Sin Behind?" In *Transhumanism and Transcendence*, with Ronald Cole-Turner. Georgetown University Press, 2011.
- Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. "Understanding and Sharing Intentions: The Origins of Cultural Cognition." *The Behavioral and Brain Sciences* (New York, USA) 28, no. 5 (2005): 675–91. <https://doi.org/10.1017/S0140525X05000129>.
- Turkle, Sherry. *Reclaiming Conversation: The Power of Talk in a Digital Age*. Penguin Press, 2015.
- Tymoczko, Thomas. "The Four-Color Problem and Its Philosophical Significance." *The Journal of Philosophy* (New York) 76, no. 2 (1979): 57–83. <https://doi.org/10.2307/2025976>.
- "Understand, v.: Oxford English Dictionary." Accessed November 27, 2018. <http://ezproxy-prd.bodleian.ox.ac.uk:2355/view/Entry/212085?rskey=iUGDql&result=1&isAdvanced=false#eid>.
- Vogler, Candace. "The Intellectual Animal." *New Blackfriars* 100, no. 1090 (2019): 663–76. <https://doi.org/10.1111/nbfr.12503>.
- Walmsley, Mike, Lewis Smith, Chris Lintott, et al. *Galaxy Zoo: Probabilistic Morphology through Bayesian CNNs and Active Learning*. Oxford University Press, 2019.
- Wigner, Eugene. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." *Communications in Pure and Applied Mathematics* 13, no. 1 (1960). <https://www.dartmouth.edu/%7Ematc/MathDrama/reading/Wigner.html>.

Wolfe, Cary. *What Is Posthumanism?* Posthumanities 8. University of Minnesota Press, 2010.
<https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=557541>.