Book Narrative

Human Persons: A Contemporary Philosophical-Scientific Synthesis

Timothy O'Connor & Philip Woodward

Contents:

Chapter 1: Introduction

Consider how fraught the following questions are in contemporary society: *Are any animals persons? Are fetuses persons? Are institutions ever persons? Could any AI systems be persons?*

These questions are fraught because of the moral force of the concept of personhood: to be a person is to enjoy certain moral prerogatives. The concept of personhood is not an exclusively prescriptive concept, however. It also picks out a psychological kind. Because the concept of personhood plays this dual psychological/normative role, it is arguably the central term in our legal, social, and geopolitical ethical vocabulary.

The concept of a person has a fascinating history in Western culture.[1] It was forged by explicit philosophical reflection over many centuries before taking hold of the popular imagination. It emerged in the ancient world as the coalescing of threads from several disparate intellectual traditions:

- From the Greeks (especially Aristotle) came the doctrine that the rational soul is the form (*morphe*) of the human body, in contrast to the merely 'sensate' souls of animals.
- From the Romans (especially Cicero and Seneca) came the use of the term 'persona' to express the uniquely dignified role that human beings play on the world's stage ('persona' originally denoted the thespian's mask).
- From the Christian theologians came an emphasis on the underlying, unshareable, and enduring individuality of the person, an emphasis which had been developed

[1] It would be even more fascinating to compare its development in Western culture with that taking place within the cultures of China, India, and beyond. But this would require a much lengthier synopsis, and it is in any case beyond the present authors' competencies.

> in the context of theological reflection on Christ's two natures and the shared nature of the three Trinitarian persons.

Pulling these strands together, Boethius declared in the 5th century that a person is 'an individual substance of a rational nature.' The idea took hold. Medieval thinkers worked with the concept and refined it, emphasizing especially the *dignity* of persons, so understood. (Aquinas, for example, insists that personhood is the *dignissimus*—the *most* dignified—mode of existence.) Despite the widespread disdain among modern philosophers for medieval categories, the category of personhood only became more central to Western philosophy in the modern era, preeminently in the work of Locke and Kant. Through the direct influence of such thinkers on modern social reformers, the concept came to have the cache it enjoys today. In short, one of the great substantive philosophical discoveries of the centuries—that we are persons—has proven its mettle in the political world. It is a deeply resonant idea.

In sum: modern Western societies are built on the idea that we are persons, an idea bequeathed by philosophers. But modern Western societies are also afflicted with an opposing idea, likewise bequeathed by philosophers: that science has shown us that we are not persons, or at any rate that being a person is not as the Western philosophical tradition would have had us believe.

Philosophical handwringing about personhood began with the 17th century rejection of Aristotelian science in favor of corpuscularianism. The broad outlines of this story of the development of revisionary conceptions of human nature throughout the modern era are well known:

- Galileo and Descartes advocate for a mechanical, mathematical physics, thereby scrubbing the material world of sensory qualities and creating the 'mind-body' problem—the problem of the place of the conscious self in a world of interacting particles.
- In the wake of the reception of Newtonian mechanics, Leibniz, Kant and others struggle to find a place for human freedom among apparently deterministic physical causes.
- Hume applies the principles of Newtonian mechanics to psychology. The result is skepticism about the 'powers of reason' that were long assumed to be the summit of human nature—a priori knowledge, moral knowledge, knowledge of natural essences, etc.
- In light of the spatial and temporal cosmic decentering of human beings, from Copernicus to Darwin, Nietzsche concludes that humans have been 'dethroned'—that they do not occupy a privileged moral position in the universe.

Thus, there is a popular narrative according to which modern science has come to show that we are material systems, rather than conscious selves; that we are causally determined, rather than free and responsible; that we are capable only of empirical knowledge, rather than conceptual or moral knowledge; and that we are morally inconsequential, rather than dignified. Wilfred Sellars famously described the apparent tension between the traditional view and the modern challenge to it in terms of two 'images' of human beings: the 'Manifest Image' and the 'Scientific Image.' Where the fundamental ontology of the Manifest Image is the person, the fundamental ontology of the Scientific Image is the particle. Unless the traditional view of the person can be accounted for – somehow – in terms of the behavior of particles, the traditional view has been debunked.

Given that the Manifest Image is foundational to modern societies globally, it would be socially catastrophic if the popular narrative (which sees an essential tension between the Manifest and Scientific images) were true. Fortunately, the popular narrative is but a nasty rumor, based on a misunderstanding of the Manifest Image and a misrepresentation of science in the terms of the Scientific Image:

- The misconception about the Manifest Image is that it is akin to a *theory* of what we are, and an antiquated one at that. But it is not. It functions not as an explanation of the appearances but as a codification of the appearances—the canonical body of knowledge about ourselves to which any plausible scientific theory must be responsive. The Manifest Image is thus not negotiable, at least in its general outlines; for *what is manifest is known*, even if only at the level of the appearances.
- The Scientific Image misrepresents science as wedded to global reductionism. Those who identify science with reduction – and these include some scientists – thereby adopt a metaphysical dogma contrary to the spirit of science itself. Science does provide us with local reductions as a matter of empirical discovery. But even these discoveries are rarer than is commonly supposed, and global reductionism will never be among them.

The central negative thesis of this book, then, is that there is no essential tension between the Manifest Image and the many achievements and lessons of science. But more important is our constructive thesis: that *science helps us understand how it is possible that we are persons.* The deliverances of contemporary sciences are not only consistent with the Manifest Image, they provide it with a theoretical underpinning. The Manifest Image supplies the appearances which science must save, and science succeeds in saving them, in interesting and often surprising ways.

But to substantiate this thesis, we must move beyond the rumored tension between the Manifest Image and science and look at the details of what the sciences have actually discovered, especially the human sciences. To do this, we anatomize the Manifest Image into its several dimensions. Persons are conscious (ch.2) unified subjects (ch.3), graspers of meaning (ch.4) theoretically rational (ch.5), purposive and free agents (ch.6), bearers of moral knowledge (ch.7), social creatures bound by love (ch.8), and bearers of dignity (ch.9). With respect to each of these dimensions, we explore what contemporary science does and does not tell us; and then we propose an explanatory story that harmonizes each dimension of the Manifest Image with our best relevant science. In a final chapter, we grapple with our history of re-fashioning our environment and ourselves through technology, by asking whether there are metaphysical boundaries to human genetic and artificial enhancement, and whether, within the bounds of the possible, there are limits to what would be valuable to us.

Many of the chapters to follow a basic template: a brief statement of the manifest image with respect to the chapter's theme; a select overview of highlights from recent scientific work on that theme; a critical discussion of physicalist reductionist or eliminativist gambits with respect to it, allegedly rooted in the science; and finally, our own way of synthesis, on which the basic form of the manifest image is taken as given and science may be seen as clarifying and deepening our understanding of it.

Chapter 2: Consciousness

According to the Manifest Image, human persons are psychological individuals: our minds are central to what we are. Mental phenomena are seemingly set apart from other phenomena by being *conscious*—feeling somehow or other—and their being *intentional*—being ‘about’, or ‘directed at’, something or other. We explore the first of these so-called marks of the mental, consciousness, in the present chapter.

In the last thirty years, a new subfield of neuroscience has emerged that seeks to identify the ‘neural correlates of consciousness.’ Though the field is still in its early stages, it has made the following discoveries:

- Specific features presented in consciousness have local correlates.
- There is no dedicated consciousness region or circuit in the brain.
- Which regions of the brain modulate consciousness at any one time is a matter of those regions’ network-properties—in particular, differentiated regions communicating with each other.

‘Physicalists’ about consciousness claim that consciousness is ontologically reducible to neural states in the sense that conscious states are wholly constituted by neural states. More precisely, the specific contents of conscious states (e.g., ‘red apple on the table in front of me’) are ontologically reducible to their localized neural correlates and consciousness itself to the activation of a form of neural network among such localized states (yet to be definitively identified), or perhaps to a kind of neural state that represents the content encoded in a first-order neural-representational state. Philosophers often state physicalism rather abstractly. But over the past four decades, as an offshoot of the search for neural correlates program, scientists and science-engaged philosophers have proposed more than twenty different reductive frameworks, or proto-theories, of consciousness. We consider the four most prominent of these theories: Higher-Order Theory, Global Neuronal Workspace Theory, Integrated Information Theory, and

Recurrent Processing Theory. Differences aside, all four employ a general reductive strategy of dubious intelligibility: reducing the qualitative features of consciousness to specified structural-dynamical features of the brain. Other attempts to reduce consciousness to its neural correlates likewise try to fit the square qualitative peg of consciousness into the round structural-dynamical hole of neural dynamics. In light of this seemingly intractable barrier to reductive theories, a natural intellectual successor is eliminativist, under the guise of *illusionism*—the claim that neural activity produces not consciousness but rather the illusion of consciousness. (There is no qualitative peg.) But this seems a desperate gambit.

An alternative to orthodox physicalism, but likewise intended to be consistent with a reductive picture of the brain, is panpsychism. Panpsychism does not reduce the qualitative features of consciousness to the structural-dynamical features of the brain, but to further qualitative features posited as the intrinsic natures of neural structures. But panpsychism turns out to be inconsistent with what we know about the neural correlates of consciousness: for it does not explain how *the dynamics of consciousness* are a function of shifting neural dynamics.

The way of synthesis sees consciousness as ontologically distinct but causally dependent in a systematic way on neural processes (in contemporary lingo, 'strongly emergent').

Chapter 3: Unity

We are *individuals*: each of us is an ontological unity in a determinate, non-conventional sense – a center of conscious awareness, knowledge, and agency.

Philosophers and scientists who take the brain merely as a complex system have one of two things to say about the unity of the psychological subject. The first camp rejects subjective unity as a fiction, given that the brain has no functional center (or rather, given that it has multiple). The other camp affirms subjective unity by treating it as a feature of the whole brain. Members of this camp explain the unity of the brain by appeal to its biological unity or its cognitive unity. But both these phenomena are vague, and thus neither can explain why it is a determinate, non-conventional fact of the matter that each brain houses one psychological subject (in normal circumstances, anyway).

Revisionary theses such as illusionism about consciousness, elimination of the subject, or treating subject-talk as conventional are by no means forced by our best science, but they do seem inevitable if a reductive framework is imposed on our best science. We propose, instead, a modest emergent dualism, on which the locus of subjectivity is an ontologically emergent individual bearing a manifold of conscious qualities and a suite of intellectual and active powers. But this emergent individual is too intertwined in its functioning and features with corresponding functions and features of the brain from which it emerges for it to be wholly identified with the *person*. The person, instead, is a psych-physical entity whose psychology, both conscious and non-conscious, is continuously constituted by the interactivity of the brain and the conscious subject.

Chapter 4: Intentionality

It is part of the manifest image that we are able to represent the world, to form beliefs about it and aims with respect to it. There is some controversy among cognitive scientists about the good standing of consciousness-talk in their disciplines. Not so with talk of representation. The neuroscience of mental representation, for example, has been an ongoing enterprise for about 150 years. Neuroscientists have identified patterns of firing in motor cortex that correspond to different motor movements and firing patterns in visual cortex that correspond to features of the environment. Such ‘feature-detectors’ in the visual cortex are found in a hierarchy of increasingly abstract pattern-detection. Recent developments in brain-imaging have allowed researchers to map the firing-patterns associated with word-meanings. In addition, developmental psychology and comparative linguistics have shown that human conceptual capacities are founded on a small set of basic categories (aka ‘core cognition’), categories which are highly abstract.

Because cognitive science (not to mention common sense!) is committed to talk of representations, philosophers under the spell of reductionism have endeavored to reduce the relation between representation and represented to the relation of *tracking* (which they have cashed out in various ways). The trouble with all such theories is that they omit *the subject’s awareness of what her own thoughts mean.*

We propose a non-reductive alternative, which treats the fundamental building blocks of intentionality as a species of conscious awareness. This theory is more consonant with the findings of cognitive science than reductionist alternatives. Here’s why: the categories that comprise core cognition are not easily understood in terms of tracking, yet these are the starting points for human conceptualization. These, we propose, are presented in consciousness, and they constrain the development of the hierarchy of feature-detectors.

Chapter 5: Rationality

According to the Manifest Image, human psychology is especially rich: human persons not only represent the world around them, they are *rational* with respect to it. Per longstanding philosophical tradition, rationality subdivides into theoretical rationality and practical rationality. To say that human persons have a rational nature is at least to say that human beings have the capacity to believe for reasons and to act for reasons. This we might call a ‘descriptive’ conception of rationality; it is the conception of ourselves that cognitive scientists call ‘folk psychology’. But the Western philosophical tradition has tended to affirm a stronger, normative conception of rationality: we have the capacity to believe the truth for good reasons and to act rightly for good reasons. We aim to preserve both conceptions.

To be rational in the descriptive sense requires capacities for at least the following:

- Intentionality, i.e., being able to represent the world, to form beliefs about it and aims with respect to it (discussed in the previous chapter).
- Inference, i.e., taking one intentional state as a reason to form another intentional state.
- Integration, i.e., forming a coherent set of beliefs and aims.

The normative conception of rationality adds a substantive connection to what is objectively valuable:

- Normative awareness, i.e., apprehension of alethic and moral value, and the proper response to the bearers of these values in cognition and action.

Our view is that the descriptive conception grounds the normative, because beliefs and aims are constitutively connected to alethic and moral values by which they are assessed. The present chapters focuses almost exclusively on theoretical rationality; our discussion of practical rationality begins here but spans chapter 6 as well.

Inference. There are three families of accounts of what inferences are: the semantic, syntactic, and procedural accounts. According to the semantic account, inferences are grasped connections among the contents of intentional states. According to the syntactic account, inferences are the applications of rules to representations in virtue of their syntactic structure.

According to the procedural account, inferences are activations of reliable belief-forming dispositions.

We advocate for the semantic approach. We are motivated in part by the shortcomings of the other two. But we also note that the semantic approach is favored by one of the leading empirical theories of human reasoning, the mental model theory. Our consciousness-based theory of intentionality shows how model-based inference is possible—and more generally, how a priori knowledge, as traditionally attributed to humans, is possible.

Integration. We use the phrase 'point of view' narrowly to denote a set of perceptual appearances at a time, but we also use it broadly to denote an integrated set of beliefs. Rationality, it seems, precludes widespread inconsistency among a person's beliefs. Whence this requirement? It would be hard to act on the basis of a discordant belief-set, so there are practical reasons to pursue integration. But these practical reasons cannot by themselves explain why it is infelicitious, from the perspective of *theoretical* rationality, to maintain a discordant belief-set.

We suggest that the norm of integration arises from the inherently normative nature of belief itself. Philosophers have long noted a puzzle about self-ascriptions of belief, a puzzle known as Moore's paradox (after G.E. Moore, who first drew philosophical attention to it). What is wrong with saying, "*p* is false, and I believe that *p*"? The utterance is a manifest absurdity, despite the fact it contains no formal contradiction. (Compare: "*p* is false, and James believes that *p*," which has the same logical form but lacks the absurdity.) The answer is that *believe the truth* is the constitutive norm of rationality. It is not merely that one *ought not* believe falsehoods; it is that one cannot do so knowingly if one is to play the belief game at all. The requirement to form an integrated belief-set—a cognitive point of view—is a function of this constitutive norm of belief in tandem with our capacity to see rational connections among contents. The constitutive norm *believe the truth*, a norm that applies to cognitive states in a piecemeal way, gives rise to a holistic norm: *have a cognitive perspective*.

There are documented cases of cognitive fragmentation: Associative Identity Disorder, schizophrenia, dementia, and so forth. These phenomena might be thought to put pressure on the appeals to subjective and rational unity in this chapter and Ch.3. We will take up these questions in Ch.10, where we will argue that these are the exceptions that prove the rule: they are pathologies of integration. In each case, a unified psychological subject is prevented from attaining the rational unity toward which her very own psychology propels her.

Chapter 6: Agency

According to the Manifest Image, human persons are *purposive agents*. We are (often) aware of a variety of means to a given end and freely choose from them. Sometimes we have a plurality of aims that cannot be jointly pursued on a given occasion or in general, and it is up to us to decide which of them to pursue, or which to prioritize. Finally, we can decide to take up new ends altogether or abandon ones previously adhered to. It seems, then, that we have a great deal of control over what we do in the moment, which motivations will guide these actions, and the wider structure of our longer-range goals and plans for achieving them.

Neuroscience has begun to identify brain centers associated with the encoding of plans for voluntary movement and their execution. Psychology has shed light on human action both via study of various pathologies of action as well as through studies of causal determinants of action, both conscious and unconscious.

Here, too, some have seen in these emerging sciences the grounds for skepticism. Ever since the success of Newton's physics, thinkers have alleged that a properly scientific attitude assumes that human behavior, along with every other process in the universe, has sufficient causes, i.e., causes that determine their occurrence on that very occasion, precluding apparent alternatives that could have come about only if prior circumstances had differed in some way. The sense that we might have chosen and acted differently is perhaps the result of the incompleteness of our knowledge of the internal motivations and external stimuli acting upon us, factors that cumulatively 'tip the scales' in a particular direction. But, as we observe, in the 21st century, the sciences are rife with merely statistical models of processes, right down to our most fundamental physical science. Human action is no less scientifically comprehensible if its causal influences are merely 'probabilistic', as it often seems to us, rather than being always strictly determined. The postulate that the present state of things is pregnant with a particular, fully determinate future is neither theoretically required nor supported by evidence.

Other free will skeptics, allowing that science leaves very much open whether human behavior is causally determined, contend instead that science commits us to supposing that human goal-directed activity is entirely grounded in underlying purposeless mechanisms (in the first instance, of neuronal interactions, which are grounded in turn in physical chemistry, and on to particle-field interactions of basic physics). The specter of reductionism re-appears. Sometimes the challenge is expressed as the thesis that physics is 'causally closed (or complete)'. A more impressive-sounding version invokes the 'law of conservation of mass-energy,' maintains that irreducible free will would violate it, and finds such 'violation' incredible. However, a careful examination of the status of conservation laws in modern physics reveals that they are rooted in 'local' conservations observed in carefully controlled experimental conditions, rather than functioning as a priori postulates whose failure in other contexts would amount to a falsification of basic physical theory. And we argue that support for energy conservation in the purely mechanistic contexts of physics and chemistry – where it obtains – does not give us strong inductive reason for supposing that conservation also obtains in contexts influenced by partly non-physical minds.

Some of the findings of recent psychology and the nascent neuroscience of human action do pose interesting challenges to the extent of our control over, and even awareness of, all the causal influences on the choices we make. But seeing these findings as 'evidence against free will' depends on thinking of having free will as all-or-nothing. We argue that they are in fact simply sharpened observations of what reflective humans have always recognized, and point us to the constructive scientific project of understanding freedom as a *degreed* phenomenon: something that varies in and across individuals over time in its range and magnitude, and depending on whether we are talking about synchronic or diachronic control (the latter being relevant to the ability to carry out mid- and long-range aims). We also need to better understand the ways in which our direct actional control is intermittently manifested within the larger flow of highly automated neural processes. Finally, it is plausible that philosophical thinking about free will has been excessively individualistic. We are intensely social creatures, and the cultural and technological context we inhabit profoundly shapes our understanding of the options available to us. We also act collectively as well as individually. All of this suggests a further, social dimension shaping the extent and character of our freedom.

Chapter 7: Morality

According to the Manifest Image, human persons are *morally responsible*. This means that our practical rationality is guided by a sense of what we ought and ought not to do, and, moreover, that this ‘moral sense’ tracks reality, at least in some cases. Human persons can know what is good and right and can act accordingly.

Moral psychology (both cognitive and evolutionary) is a vibrant science today. We rehearse its most important findings and its open questions. Although the field has made interesting discoveries about the nature and etiology of human moral judgments, it does not tell us whether those judgments track moral reality. A number of prominent philosophers have argued from moral psychology to moral skepticism, on the grounds that the best explanation for why (either cognitively or evolutionarily) we make the moral judgments we make does not mention moral reality, and thus that we have no reason to think that our judgments track moral reality.

The thesis that we cannot know what is good or right is repugnant to most theorists. In consequence, many moral philosophers have preferred a deflationary account (of one variety or another) of the moral domain, according to which moral facts are not independent of the psychological mechanisms used to form judgments about them. To our mind, these deflationary moves fail to preserve the Manifest Image. We cannot avail ourselves of this ‘easy way out’ of moral skepticism.

It is not unreasonable, in our view, to respond to skeptical worries about moral knowledge by making a ‘Moorean’ shift. If cognitive or evolutionary moral psychology implies that we cannot know that cruelty is wrong, then so much the worse for those theories!

But we acknowledge that doubts about our pretentions to moral knowledge will not be fully dispelled without an alternative psychological theory that explains how our judgments come to reliably track the truth. Empirical moral psychology points us in the directions of such a theory, but philosophy is needed to complete it. We suggest that the ingredients for such a theory are as follows:

- Basic emotions, inherited from the mammalian evolutionary past, as the vehicles of our representations of value;
- Hedonic experience as *direct psychological contact* with one species of objective intrinsic value;
- Explicit reflection on the elements of well-being (which includes but is not restricted to hedonic experience), and the individual and societal activities conducive to well-being;
- The cultural transmission of the outputs of this explicit reflection;
- Non-basic emotions, constructed on the basis of this cultural transmission.

The consequence of all this is the capacity to be moved, emotionally and consequently behaviorally, by objective intrinsic value. The moral sense is somewhat analogous to sense-perception, where emotions are to the world of value what sensory qualities are to the world of perceptibles, as follows. We know there is a world out there even when we are not sensing it, but the world is made manifest to us only when our consciousness is saturated with visual hues and sonic pitches and the like. Likewise, we know that we inhabit a world of goods and bads and rights and wrongs, and we reason about these matters, but their reality is made manifest to us only when our consciousness is saturated with longings and revulsions and enchantings and so on—affective states that have been attuned by nature and culture to what is good and bad.

Believe the truth is the constitutive norm of belief. Thus, beliefs are subject to standards of evaluation by their very nature. But just as there is a constitute norm of belief, so there is a constitutive norm of volitional states: *pursue the good*. Of course, there are many ways that an aim might present itself as choice-worthy: it might seem pleasant, interesting, curious, convenient, diverting, beautiful, morally required, heroic. It might be the least bad of a set of bad options. It might seem good only in the sense that it is *an aim*. ("I'm tired of sitting here with nothing to do. Let's go for a walk.") But just as '*p* is false; and I believe that *p*' is not possibly a sincere utterance, so '*o* repulses me in every way; and I willingly aim for *o*' is not possibly a sincere utterance. It is not merely that one ought not pursue the utterly bad; it is that one cannot do so, if one is to play the willing game at all.

Just as the norm of belief gives rise to a holistic norm of theoretical integration (have a cognitive perspective), so the norm of volition gives rise to a holistic norm of practical integration: *have a practical identity*. This dimension of the psychology of a person is best identified with the notion of the 'self'—a practical rather than a metaphysical category.

Thus, volitional states are subject to standards of evaluation by their very nature, the standard of goodness. Not only are human persons capable of representing value, not only are they capable of knowing what is valuable and what is not, they additionally have an *essential interest* in ordering their choices according to such knowledge. The notion of a person is thus a teleological notion, as the old Aristotelian tradition insisted: persons are 'by nature' ordered toward the true and the good.

Chapter 8: Love

According to the Manifest Image, human persons are *social creatures*. This dimension of the Manifest Image is both psychological and normative. At one level, to say we are social creatures is to say that our aversion to isolation is hard-wired. But at another level, it is to say that this hardwiring is a good guide to our well-being. Moreover, to say we are social is also to indicate what we ought, morally, to do—viz., take care of those to whom we are socially related. Sociality is not just a human need but the duty and flourishing of persons as such. That's the traditional view.

Much of modern philosophy and psychology has focused on the cognitive and agential capacities *of the individual*. This emphasis was in some cases an explicit corrective to the collectivism of an old, pre-modern Aristotelian tradition. But recent decades have seen a renaissance, within all the human sciences, of the idea that human persons are essentially social. The following are some of the most important recent scientific claims made about human social life, many of which subvert earlier scientific consensus:

- The competitive advantage of the human species over other primates stems from our ability to cooperate with one another.
- The evolution of the human brain was driven in large part by the need to keep track of and manage complex social relationships.
- Human evolution cannot be understood apart from the evolution of human culture.
- The foundation of infant and child development is a secure attachment between child and caregiver.
- Social interactions drive the ontogenetic development of every distinctively human capacity.
- Aside from baseline human needs, the single biggest ingredient in human health and happiness is social connection.

As far as the social dimension of human nature goes, it thus appears that the Scientific Image is moving in the direction of the Manifest Image. Recent science has resoundingly endorsed the idea that social connection is central to human well-being. It provides qualified endorsement of the idea that human social behavior is connected to the moral life. For example, social psychologists often speak of the ‘pro-social behavior’ that sets humans apart, behavior that is motivated by mutual care rather than by reciprocity. Such endorsement is qualified by the observation that our craving for belonging also results in social dynamics that are repugnant from the moral point of view: intense pressure to conform, disdain for, or even cruelty toward, out-group members, and so forth.

Our own view is that human social instincts do propel human persons toward their well-being and toward their moral duties. But the good of human sociality cannot be understood apart from our rationality, especially our capacity for moral awareness. Qua rational, human persons are capable not just of ‘pro-social behavior’, but of *love*. Love, we contend, depends on rationality and is inherently normative. We draw on recent work on the philosophy and science of love to show how to connect the dots between social instincts, on the one hand, and human flourishing and moral agency, on the other.

Greek-speaking antiquity has bequeathed to us a familiar taxonomy of the varieties of love: *eros* (romantic love), *philia* (love between friends and family members), and *agape* (disinterested compassion). Contemporary scientific work on love employs this taxonomy as well, under the headings ‘romantic love’, ‘attachment’, and ‘altruism’. Contemporary philosophers use a related taxonomy: attraction-love, attachment-love, and benevolence-love. These categories map roughly onto the original taxonomy, although contemporary philosophers tend to think of these as three *dimensions* of love rather than three *species* of love. Most philosophers think of paradigmatic loving relationships (romantic, friendly, or familial) as including all three dimensions.

Recent work on the neuroscience and physiology of love has set that field on a clear trajectory:

- From thinking of love as a simple emotion to thinking of it as a psychologically complex state that includes cognitive elements;
- From treating love as a simple physiological response, the hallmark of which is the release of neurotransmitters such as oxytocin, serotonin, and/or dopamine, to treating such physiological responses as pieces of a much larger puzzle;
- From treating romantic love vs. friendship/family love as psychologically distinct, to treating them as significantly overlapping.

Love, as contemporary philosophers and scientists are coming to understand it, is not best thought of as a desire or feeling. It does include affective elements, but it also includes cognitive and volitional elements, e.g., the making of commitments to other people in response to an awareness of their value. This complex state is the proper locus of study of human social 'behavior' (or better, social *activity*).

Whether or not 'social behavior' as such is a great good in human life, recent philosophical work has shown that love *is* a great good (intrinsically, not just as the engine of reciprocal beneficence): it gives content to one's will, thus orienting a person in the world; it structures one's practical rationality, thus providing the sense of certain actions as *meaningful*; and it integrates one's psychology, harmonizing volitional, conative, and cognitive states with one another. Manifesting our social nature by committing to love others is thus an important component in human flourishing.

Moral philosophers have not been friendly to the idea that love, at least in the sense of *eros* or *philia,* is constitutive of moral duty. Their suspicions of partiality (as opposed to the supposed impartiality of the ethical point of view) are of a piece with social psychologists' worries about in-group bias. But we argue that these suspicions are misplaced. Love, as a rational attitude (that is, an appropriate response to value), is constitutive of duties of partiality *and* of duties of impartiality, in the following manner. The goodness of each person, as such, calls out for a loving response; this is a requirement of practical rationality. But the shape that this loving response takes, practically speaking, is a function of the nature of one's relationship to each person. In short, *to each person one owes exactly what one's relationship with each demands.* To those with whom one's relationship consists exclusively in shared humanity, one owes recognition. To those with whom one shares something further—an economic system, a society, a culture—one owes a more active respect for their rights. A few highly-demanding relationships generate the bulk of one's duties of beneficence. One's most intensive duties of beneficence fall out of one's most intimate relationships.

Thus it turns out that human social proclivities, when drawn up into and modulated by practical rationality, are properly expressed as interpersonal love, and not merely as 'pro-social behavior' operating instinctively. And love does not shun out-group members but attends dutifully to them in ways befitting one's relationship to them.

Chapter 9: Dignity

According to the Manifest Image, human persons are *bearers of dignity*. Dignity is a complex moral status, including at least four dimensions of worth:

- unique worth: worth that is greater, qualitatively and/or quantitatively, than the worth of non-human terrestrial creatures;
- inherent worth: worth that is objective and impartial, grounded in one's natural properties;
- inviolable worth: worth that commands respect for a suite of rights;
- irreplaceable worth: worth such that the loss of an individual human is not mitigated by replacement with a duplicate.

Charles Darwin made the famous claim in *The Descent of Man* that humans differ from animals only in degree and not in kind. Darwin meant by this that every psychological capacity found in humans (a) has precursors in non-human animals and (b) came into being via gradual rather abrupt evolutionary steps. Many subsequent thinkers have taken these claims to threaten the unique worth of human beings with respect to the rest of the animal kingdom. Other thinkers have pushed back, suggesting that the degree/kind difference is obscure, and that differences of degree, over a certain threshold, could account for differences in moral status.

We begin by reviewing what is known and what is not known about human evolution and some of the most recent findings in comparative animal psychology. Then, in light of these findings, we proceed to account for the moral uniqueness of persons in terms of three stages of psychological transitions that separate persons from their primate ancestors:

1. The accumulative stage, in which primate capacities are quantitively increased. Most significant are increases in executive function and in social cognition. These increases make language-learning possible.

2. The ampliative stage, in which these accumulated capacities interact with each other in complex ways. Language is the catalyst for a plethora of mutually-enhancing feedback loops among the various capacities in the human mind. Primate intentional capacities are massively expanded by the power of language, which is symbolic and communal; it puts at one's disposal all of the representational resources of one's community, prior to, or even in the total absence of, grasping meanings for oneself. The result is the full suite of human conceptual capacities.

3. The additive stage, in which a qualitatively new capacity is superadded, viz. self-consciousness. This novel feature of human psychology is apt to be conflated with meta-cognition, in the sense of having mental states that represent other mental states, or the ability to have self-concerning attitudes, or the possession of an I-concept, but it is distinct from all of these. We propose that the transition from consciousness to self-consciousness is akin to the transition from dreaming to waking consciousness, i.e., a transition in the *level* of consciousness. Self-consciousness is the dawning of an inner light, a coming home to oneself as a self. It provides the necessary cognitive distance on one's own psychology for one to be truly autonomous.

The foregoing describes the rise of personhood out of the primate past. The question remains why persons, in virtue of their categorically different kind of psychology, enjoy a difference in worth within the several dimensions listed above. Persons, we have urged, are capable of responding appropriately to value. Their worth, in the first instance, is a function of this capacity, because that which can respond appropriately to value is inherently worthy of respect. The shape that this respect must take (the suite of rights to which humans are entitled) is a function, further, of the unique vulnerability that results from that capacity (to respond appropriately to value). Because our interests include apprehending value, and because we are self-conscious and thus reflective about those interests, we are capable of a different kind of flourishing (when we apprehend the good) and a different kind of suffering (when we are deprived of the good) than other creatures. Our *dignity* is thus ultimately grounded in a special kind of *vulnerability*.

Chapter 10: Technology

This culminating chapter moves in a different direction. So far, we have provided a synthesis of contemporary philosophical and scientific work on human persons that preserves the Manifest Image. But we now acknowledge that the Manifest Image is something of a moving target. Could we become something very different from what we are now? If we could, ought we to pursue the possibility? Should we respect the integrity of our personhood, as bequeathed to us by evolution, or might our personhood fall within the purview of our creative control?

According to the 'natural law' ethical tradition that dominated ethical theorizing from Roman times until the dawn of the modern era, nature is the starting point for practical deliberation. Creatures of each kind have a fixed essence that determines what is good for them, and they flourish when their essence is fully actualized. The 20th century existentialists rejected every element of the natural law framework. For the existentialists, persons, being free, have no fixed essence, and suggesting that they do have an essence is to demean them. We think it plain that neither of these positions (natural law or existentialism) has proven satisfactory. Contra the natural lawyers, how we happen to be is no infallible guide to the conditions of our flourishing; contra the existentialists, some facts about our flourishing are grounded in the kinds of things we are. In particular, we are persons: we are conscious, rational selves built to believe the truth and to will the good, summoned to meaningful living in the context of relationships with other persons, and owed respect and care. What we are establishes what our interests are, to a certain extent, and these interests must be respected by any proposed technological modification of the human world.

We discuss three types of technological intervention on persons: creating artificial persons, natural persons becoming artificial persons, and enhancing natural persons artificially.

1. Creating artificial persons. It is more common to speak of artificial 'intelligence' than artificial 'persons'. The term 'intelligence' is ambiguous (as Alan Turing pointed out so many years ago): there are senses of 'intelligent' on which we already have produced artificial intelligences. But if intelligent means *rational*, in the sense in which persons are rational, then it is much harder to determine whether it is even possible to create artificial persons. For our purposes, the crucial point is that there are no purely *mechanical* persons, since being a person is more than instantiating a causal structure. Being a person entails consciousness, intentionality, agency, and a constitutive vulnerability to truth and goodness.

If it were possible to create artificial persons, ought we to do so? Not simply for fun, for profit, or for power, any more than one ought to beget a natural person for those reasons. And these seem to be the only reasons currently on offer by the relevant parties.

2. Becoming artificial persons. Humans have various natural vulnerabilities (e.g. to the elements and to disease) that have been mitigated technologically. Since our biological nature seems to entail such vulnerabilities, it has been proposed by so-called 'transhumanists' that we should contrive to become non-biological. On the leading practical proposal, the unique informational pattern of a person's brain might be uploaded to a digital platform where she could continue to live in perpetuity.

If there are (of necessity) no purely mechanical persons, then it is not possible to become one. Moreover, there are metaphysical constraints on the persistence of persons that would be violated in such cases. We take a brief detour into this vast topic (the persistence conditions of persons).

Less radical ways of overcoming our vulnerabilities (which bleed into the scenarios discussed in the next section) might not run afoul of metaphysical necessities, but this would not render them desirable. After all, some vulnerabilities (i.e., to truth and goodness) are essential to what we are, and others (to other people, in the context of loving relationships) are essential to our flourishing. Great care must be taken to determine which vulnerabilities ought to be safeguarded and which remediated.

3. Enhancing natural persons artificially. The category of 'enhancement' is quite broad—covering everything from wearing running shoes to using a calculator to (hypothetically) modifying the size and power of the brain via gene-editing. There is no simple calculus for adjudicating various proposals along these lines, but the following two principles can take us quite far.

- Beyond baseline human needs, the most important ingredient in human flourishing is social connection.
- Vulnerabilities engender interdependence between persons, and interdependence provides the infrastructure for most human social connection.

In light of these principles, much of the allure of enhancement—becoming bigger, faster, and stronger!—evaporates.

Chapter 11: Coda: Personhood and the Drama of Life

We end with some brief comments about why the foregoing matters.

One of us, in the context of a philosophy of mind course one of us taught, recounted to the class the reductionist cosmic narrative—a story consisting entirely of the causally-necessitated, or at least probabilified, exchanges of energy among particles clumped in various ways. A student responded: “But that is not the only story there is to be told about the universe!” We concur. Indeed, the reductionist ‘story’ is not a story at all. But our universe *is* a universe in which dramatic stories play out because it is a universe that includes persons. Persons are aware of the world, not just as a nexus of causes but as an arena of value, in which great goods can be attained or lost, honored or violated, depending (among many other things) on the intentions they are free to form.

Because we ourselves our persons, our lives are occasions of comedy and tragedy. This then, is the final payoff of a synthesized philosophical-scientific conception of ourselves: an invitation to meaningful engagement with what matters—to the drama of life.