Custodians of Value: A Theory of Human Personhood

Philip Woodward

Abstract: According to a traditional view, humans are superior to their non-human terrestrial companions because they alone are ‘rational animals’. Although the traditional view is presupposed by our social and legal institutions, it has been called into question by modern science: Darwin himself claimed that humans differ in degree rather than in kind from animals, and recent discoveries in comparative animal cognition have seemed to confirm Darwin’s assertion. Sustaining the traditional view in light of these discoveries calls out for a careful comparison of the human mind with its precursors, the vertebrate mind and the mammalian mind. Psychological capacities shared across the animal kingdom are repurposed in humans, rendering humans uniquely capable of rationality. To be rational is to respond appropriately to value. Humans as such are not rational animals, but have the potential to become rational animals. This potential marks a morally relevant difference in kind between humans and non-human animals.

**1. Introduction: The Boethian Definition vs. Contemporary Science**

Western philosophy is often criticized for its inability to ‘make progress’. After interminable arguing, it is claimed, philosophers get no closer to consensus on a topic. Whether the criticism is apt in general, there is at least one vivid counterexample: the philosophy of personhood.

The category of personhood is absent in the work of Plato and Aristotle. It shows up in Cicero, but becomes a topic of philosophical reflection in its own right only in the context of the early Christian theological controversies. In the 6th century, Boethius provided the canonical definition of personhood in Latin: *an individual substance of a rational nature*. It was already standard in antiquity to distinguish between ‘theoretical rationality’, which aims at truth, and ‘practical rationality’, which aims at appropriate action. Presumably Boethius meant, at the very least, that persons as such exhibit rationality of both types.

Various alternative formulations of the Boethian definition were discussed throughout the medieval period, and various accounts of rationality were put forward. Discussion of the concept only increased in the early modern period. The culminating discussion is Kant’s. For Kant, having a rational nature is the same as exhibiting *moral autonomy*—the ability to understand moral ‘oughts’ and to regulate one’s actions accordingly. (Thus, for Kant, practical rationality takes precedence over theoretical.) Importantly, Kant draws a connection between *personhood*, understood in terms of moral responsibility, and *dignity*—having a distinctive kind of worth that commands respect from others. That is, being capable of recognizing moral worth is itself a source of moral worth. Or, at any rate, the two are conceptually linked in some way.

The Kantian conception of the nature and worth of persons is absolutely central to liberal democratic thinking. It has been exported to virtually the whole world via official venues (such as the United Nations’ Universal Declaration of Human Rights) and informal onces, becoming the moral ‘Lingua Franca’ of international law. Let us not forget to pay our respects to the countless philosophers, known and unknown to history, who labored over the centuries toward this rich understanding of ourselves—an understanding without which our most cherished social and political institutions would be inconceivable.

The emphasis of the school of thought I have been describing is on the rational nature and high moral status of persons, not on the *non*-rational natures and *lower* moral statuses of *non*-persons. But the contrast with non-human animals has always been implied and often made explicit. Aristotle, no doubt, has had an enduring influence here. For Aristotle, the formof animal bodies— i.e, the suite of powers that make them what they are—is their sensate soul, whereas the form of human bodies is the rational soul. (The human soul includes sensate powers, of course, but because these powers do not operate independently of reason, they are analogous rather than identical to animal powers.[[1]](#footnote-1)) Moreover, Aristotle claims that rationality is the most ‘godlike’ thing in nature. Thus, Aristotle bequeaths to virtually all subsequent thinkers in the Western philosophical tradition up until the 19th century[[2]](#footnote-2) a framework according to which humans, because rational, are as different from all animals as all animals are from plants, and, in virtue of this difference in kind enjoy a difference in rank.[[3]](#footnote-3) I’ll call this ‘the traditional view’.

The traditional view endures as moral common sense.[[4]](#footnote-4) At the same time, challenges to its intellectual foundations have mounted from within the modern scientific worldview, to the point where the advocate of the traditional view now shoulders the burden of proof. Whence this shift? A watershed was Darwin’s famous assertion in *The Descent of Man* that humans differ from animals only in *degree* and not in *kind*. Darwin meant by this that every psychology capacity found in humans (a) has precursors in non-human animals and (b) came into being via gradual rather abrupt evolutionary steps. Darwin’s view of the human/animal difference stands in stark contrast to Aristotle’s. Insofar as you need something like an Aristotelian kind-difference to underwrite a difference in rank, Darwin thus repudiates the Traditional View in both its metaphysical and moral dimensions. Contemporary animal rights advocates certainly have drawn this lesson from Darwin.

It is important to notice that Darwin was not so much announcing an empirical discovery as launching a research program. It was for very general reasons that he, and others like him, have believed that the received view is not tenable within a scientific worldview. In particular, there are two assumptions of that worldview that are apparently at odds with the traditional view. The first is *evolutionary gradualism*: the idea that stages in the evolutionary progression of life are always incremental alterations of what came before. Second is *functional reductionism*: the idea that the functioning of all systems, however complex, is reducible to basic physical and chemical causes. Either assumption has the implication that the differences between allegedly ‘higher’ and ‘lower’ organisms are just a matter of greater or lesser *complexity.*

I reject both assumptions. I deny that evolutionary gradualism ought to be taken for granted, not because of any particular evidence against it (though plenty has been adduced), but for a methodological reason. It is the job of the evolutionary theorist to tell us how human nature (along with the rest of the biosphere) came to be, not to tell us what human beings are like. So, if our best account of human nature is at odds with evolutionary gradualism, then evolutionary theorists need to adjust their paradigm to accommodate the fact.

I deny functional reduction because evidently there are psychological phenomena that are not functionally reducible (a point to which I return below).

But just because there are no good *general* reasons to think that modern science is inconsistent with the metaphysical and moral uniqueness of humans, there are, in fact, particular scientific discoveries—fruits of the research program Darwin launched—that put pressure on the traditional view. Take any psychological capacity that is reputed to be uniquely human, and it can be found, in an attenuated form at least, somewhere in the animal kingdom. Communicating vocally? So do vervet monkeys.[[5]](#footnote-5) Parsing syntax? So do dolphins.[[6]](#footnote-6) Making tools? So do crows.[[7]](#footnote-7) Teaching the young? So do meerkats.[[8]](#footnote-8) Forming abstract concepts (such as sameness and difference)? So do pigeons.[[9]](#footnote-9) Remembering particular past episodes? So do rats.[[10]](#footnote-10) Discerning other’s mental states? Coordinating activities in complex ways? Responding empathetically to suffering? So do chimpanzees.[[11]](#footnote-11) Some of these findings are disputed, but given how many are well-established, the advocate of the traditional view shouldn’t bet on their being overturned. Apparently, there is such a thing as rudimentary rationality. Animals have it, and it comes in degrees.

Not only has modern science shown that animals possess rudimentary rationality, it has also shown that rationality in human beings is far from ideal. Reasoning does not influence human decision-making very often. Rather, the reasons that come to our minds are most often those that justify our beliefs and actions after we have decided upon them, and we tend to overestimate the quality of these reasons.[[12]](#footnote-12) Furthermore, non-rational factors—such as whether one has just found a coin, or whether one is in a hurry—*do* influence human decision-making to a considerable degree.[[13]](#footnote-13) Thus, science has apparently narrowed the traditional gap between animals and humans by both ‘up-grading’ animal powers and ‘down-grading’ human powers.

My aim in what follows is to defend the *metaphysical* dimension of the traditional view, with a brief comment on the moral dimension in the concluding section. Contra Darwin, and consistent with the findings just described, being a human means being a different *kind* of thing from any animal, for it means being a person—or rather, having a capacity for personhood that no animal possesses.

I begin by clarifying the notion of ‘rational nature’ that is guiding my discussion, which turns out to be a normative ideal. Then I describe minds of three very general types found in nature: the Vertebrate Mind, the Mammalian Mind, and the Human Mind. I show how the latter two types re-purposes resources inherited from the type prior. I characterize the crucial differences between the Human Mind from the Mammalian mind in four stages: (1) the *accumulative* stage, (2) the *ampliative* stage (both of which involve only differences in degree), (3) the *additive* stage, and (4) the *transformative* stage (at which point we arrive at a difference in kind). Finally, I compare the resulting picture of the mature Human Mind with the normative ideal of rationality. I conclude that humans are endowed with the unique capacity to be ‘custodians of value’—to be persons.

**2. The Personal Ideal**

‘Rationality’ is used in many ways. As indicated above, the oldest division is between theoretical and practical rationality. These categories require further explication. ‘Theoretical rationality’ most generally construed is cognitive activity aimed at (propositional) knowledge. More narrowly, it can mean engaging in processes of *reasoning*, i.e., making inferences in one’s mind. (Note that inferences come in many varieties and exploit many different kinds of rational connections among thoughts.) Famously, for Aristotle, it is primarily in virtue of our theoretical rationality that we are ‘godlike’. But he emphatically did not mean by this the making of inferences. For Aristotle, the pinnacle of theoretical rationality is *‘teoria’*, which is not so much a matter of seeking the truth but of apprehending it, holding it in mind, contemplating it.

We can thus make distinctions between theoretical rationality as (a) knowing and (b) coming-to-know; and within the latter, between (b1) the quest for knowledge in general and (b2) rule-governed inferences in particular. A final, orthogonal distinction is between (c) isolable episodes of knowing (or coming to know) and (d) systematic understanding (or the pursuit of such).

‘Practical rationality’ has to do not with knowledge but with action. But it can have to do with action in multiple ways. *Procedural* practical rationality is means-ends reasoning, or more simply, ‘strategizing’. *Substantive* practical rationality is the intellectual apprehension of appropriate ends of acting. (Some thinkers classify this sense of rationality as theoretical rather than practical.) A further sense has to do with the assessing possible actions according to a normative standard, which is Kant’s notion.

Which of these notions of ‘rationality’ is definitive of personhood? Clearly, not all are equally so. For example, it would be very strange to suggest that persons as such are masters of deductive logic. The reason it would be strange, I suggest, is that drawing out deductive consequences is not particularly *valuable,* at least not intrinsically. But the relevant notion of ‘rationality’ ought to have something to do with what is valuable. A compelling alternative comes into view when we bring together Aristotle’s privileging of *teoria* and Kant’s privileging of moral autonomy (a kind of practical rationality). It appears at first glance that Kant and Aristotle are opposed to one another when it comes to the relevant ranking of theoretical and practical rationality: Aristotle says that theoretical rationality is highest, whereas Kant makes practical rationality definitive of personhood. And yet, it is hard to understand the appeal of *teoria* if the contemplated object is not itself something of great value. *Sensitivity to value* is thus the pinnacle of rationality for both thinkers.

What characterizes persons is their responsiveness to reasons, where the relevant reasons have to do with what is *valuable*. Persons give to things their due; they are *custodians of value*.

This is, of course, a normative ideal. I now want to cash out the psychological prerequisites of attaining this normative ideal. Let’s imagine the ideal (finite) person, a being that is maximally responsive to that which is valuable. What would such a being have to be like? There are, I want to suggest, three families of capacities that would need to be actualized in such a being.

First, such a being would have to be endowed with certain cognitive capacities. For she would need to know (1) what is valuable, and (2) how to respond.

Knowing what is valuable would involve knowledge of all the ‘final values’, the bearers of ultimate, non-instrumental value. Moreover, it would involve knowledge of the relative prioritization of the final values, insofar as they are commensurable with each other.

Knowing how to respond would involve several components. Different values call for different kinds of response (awe, respect, protection, promotion, imitation, union, consumption, and so forth)—some involving a kind of distancing from the valued thing, some involving a kind of closeness to it, and some a blend. So, the ideal person would need to understand how different final values count as reasons for these different sorts of responses. Furthermore, where the proper response calls for some goal-directed activity—e.g., the consumption of food, the promotion of others’ interests—she will need to know two more things: first, the variety of actions at her disposal; second, which of these will have the desired result vis-à-vis the final value(s) at stake. Knowing both these things will require causal knowledge of her world—ideally, exhaustive and systematic understanding.

Such are the cognitive endowments of the ideal person. Note that so described there is nothing essentially inferential or procedural about the rationality of the ideal person. She has before her mind ‘the space of reasons’. Only as a person falls away from this rational ideal will she need the ability to *reason*, where this is understood as a psychological process rather than a state. ‘Intelligence’—the ability to engage in reasoning processes well—is no essential ingredient of ideal rationality.

The ideal person will need other psychological endowments beyond the cognitive. She will also need certain *affective* capacities. In some cases, the proper response to value has an affective dimension (awe and union come to mind). And even where it does not, the ideal person nevertheless *desires* to respond in the right sort of way—it is certainly not *contrary* to her inclinations to so respond—and she *delights* in so responding. It makes her happy to be a custodian of value.

And finally: the ideal person is not responsive to value in a mechanical way. She *chooses* it; she integrates her will around it. So she will also need certain *volitional* capacities—the capacity to freely and consistently choose actions that are appropriate with respect to the final values that she knows to be such and that move her affectively.

This, then, is our ideal of a ‘rational nature’: the orientation of certain cognitive, affective, and volitional capacities toward *the good*. Let us now draw our attention as far away from this ideal as we can, toward the most rudimentary animal mind. I will then show how to ‘build up’ from such a mind to a *human* mind, a mind that has the capacity for rationality in the sense just described.

**3. The Vertebrate Mind**

I begin with what I’ll call the ‘Vertebrate Mind’. (Let me be clear that this is a toy model of a rudimentary mind. Some but not all vertebrates will exemplify it.) The Vertebrate Mind orients an animal in the world, helping the animal to find food, avoid threats, and so on. Significantly, the Vertebrate Mind exemplifies a tri-part structure analogous to the one exemplified the ideally rational mind. First, it has rudimentary ‘cognitive’ capacities, or rather, ‘C-capacities’ (a category wide enough to span the rudimentary and ideal versions, and everything in between). In particular, the Vertebrate Mind includes *sensation*, the ability to sense relevant stimuli in the environment, and *associative learning*, the ability to adjust behavioral dispositions in response to the pairing of stimuli and reward (or harm).[[14]](#footnote-14)

Second, it has rudimentary affections, or “A-capacities”: positively- and negatively-valanced feelings that attract or propel it with respect to certain stimuli, e.g. hunger, thirst, pain, and fear.[[15]](#footnote-15)

Third, it has a very rudimentary mechanisms for behavioral control, or “V-capacities”: it produces behaviors that are appropriate, given its perceptual stimuli and affective goads.

**4. The Mammalian Mind**

The Mammalian mind comprises the capacities of the Vertebrate Mind but with important augmentations. (Again, this is a toy model, and it is errs on the side of generosity—so probably closer to the mind of a chimpanzee than of a vole. Note, too, that certain avian and cephalopod minds are similarly sophisticated.) The most striking additions fall under the headings of ‘executive functions’ and ‘pro-social’ behavioral dispositions, if we use the prevailing taxonomy in empirical psychology.

Let’s begin with additions to Mammalian C-capacities. In the Mammalian Mind, *sensation* is transformed into *perception*, which involves the application of *concepts* to sensory inputs. The standard suite of basic perceptual concepts is known as ‘core cognition’, and it includes the concepts *object*, *quantity*, *space*, *animacy*, and *agency*.[[16]](#footnote-16) These concepts can also be deployed ‘offline,’ in the isolation from perceptual episodes. The Mammalian Mind thus includes the basic conceptual resources necessary for *thinking*. It also includes the basic functional structure needed for thinking. Here two of the three canonical ‘executive functions’ are relevant: *working memory* and *cognitive flexibility*. The former is the ability to hold information in mind and manipulate it. The second is the ability to consider alternative possibilities, both in the sense of (a) considering alternative underlying explanations of the same appearance in different contexts, and (b) considering rival plans of action that need to be weighed. Concept-possession, working memory, and cognitive flexibility together add up to a rudimentary form of *intelligence* in mammals. It’s not very impressive by human standards: disjunctive syllogism and hypothetical syllogism may constitute the upper bound.[[17]](#footnote-17) But it makes possible a whole new type of behavioral flexibility, one that is based not on associative learning but on learning by explicit trial-and-error. (Of course, mammals learn associatively, too.)

Regarding A-capacities: The Mammalian Mind adds a secondary emotional layer, while retaining the primary. These secondary emotions include grief, play, and care (or empathy).[[18]](#footnote-18) Importantly, these second-tier emotions are largely social. An animal’s affective state is responsive to conspecifics: it is disposed to feel well because a fellow animal is present (attachment), to feel badly because a fellow animal is absent (grief, or the animal analogue), and to feel well *or* badly because a fellow animal is displaying signs of itself feeling well or badly (empathy). This all amounts to a major modification of the animal’s motivational structure: other animals *matter* to it. Of course, they might not matter very much, when weighed against other affective motivators. Mammal sociality is rather callous by human standards.

The Mammalian Mind has updated V-capacities, too. In particular, it includes a rudimentary version of the third canonical executive function, ‘inhibitory control’. In humans, the term is used to cover a range of volitional activities. When applied to animals it usually just means *delayed gratification*. Most animals are very bad at delaying gratification, but a few can for a little while if the payoff is big enough.[[19]](#footnote-19) More generally, mammals need volitional control (not just ‘inhibition’) in order to bring their intelligence to bear on their behavior. If there is always a direct functional link between affect and behavior, the animal’s thinking won’t really matter at all; cognitive flexibility wouldn’t translate into behavioral flexibility. At the very least, volitional control interrupts this direct link, allowing a *thought* to chart the animal’s course in place of an urge. Additionally volitional control is involved in cognition itself, as an animal sustains attention on a particular task.

It is not easy to pin down the respects in which mammalian volitional control is of a piece with human volitional control. Minimally, one could hold that it is only *functionally* similar, in the sense of playing the same mediating role between cognition and behavior. Maximally, one could hold that it is the very same faculty of *free choice*, the difference lying in the nature of the inputs involved. (My view is intermediate between these extremes, as will become apparent.)

**5. Ontological Aside: Five Functionally Irreducible Features**

Before discussing the Human Mind, I want to pause to consider the ontology of the capacities I have just been discussing. Minds are realized in brains. Brains are complicated electrochemical systems. Is the transition from the Vertebrate to the Mammalian mind just a matter of more, and more complicated, electrochemical structure? For that matter, is the transition from non-minded organisms (presumably all plants, and perhaps some primitive animals) to Vertebrate minds just a matter of more, and more complicated, causal structure?

I don’t think so. There are five features of the Mammalian Mind that I think are ontologically emergent phenomena—phenomena that are not identical to, constituted by, or realized in neuronal structure and functioning, while nevertheless depending on that structure and functioning for their occurrence and nature. The first two features I discuss are probably shared with the Vertebrate Mind.

Arguments for the functional irreducibility of these features are philosophical rather than empirical, discussed in a vast literature. I have developed some of my own arguments elsewhere. Here I say enough to clarify and motivate my position.

*1. Subjectivity*. Although a brain is a distributed functional system with no ‘center’, one’s mental life is experienced from a unitary point of view. At least this is true of one’s *conscious* mental life, and I take the capacity for consciousness to be determinative of having a mental life at all. (Any system can be treated as though it were mental, of course, but only consciousness makes it the case that such treatment is reflective of the truth rather than merely useful.) What makes it the case that there are psychological individuals at all, rather than simply aggregates of processes occurring in tandem, is that psychological subjects are unitary in an irreducible way.[[20]](#footnote-20) There are multiple ways to conceive of such emergent unities, some ontologically dualistic and some not.[[21]](#footnote-21)

*2. Sentience*. Consciousness is populated with particular felt qualities that we call ‘sensations’. It is precisely because of the way sensations *feel*—the painfulness of pain, the ache of hunger, the terror of fear, the buzzing delight of pleasure—that they play their motivating role.

When it comes to sensations, there is no appearance/reality gap: we know by feeling them what they *are*, what their natures consist in. But the more we understand the structure and function of the brain, the less we understand how such feelings could be realized in or constituted by goings-on in the brain.[[22]](#footnote-22) The so-called ‘explanatory gap’ between brain functions and sensations is probably a matter of a straightforward ontological category difference: sensations are *qualities*, and brain functions are *causal episodes*.[[23]](#footnote-23) Thus, it cannot be true that sensations *just are* brain functions.

*3. Representation*. The difference between sensation and perception is the difference between having sensations in response to stimuli vs. being aware of the stimuli as such. In other worlds, perceptual episodes are not just *caused* by the world but are *about* the world; they *represent* the world. (Moreover, mammalian conceptual capacities allow the animal to *think* about the world, not just to perceive it.) One familiar type of representation is *symbolic* representation, and many researchers, both in philosophy and in empirical science, take mental representation to be a form of symbolic representation. But it is not. Symbols bear an extrinsic relationship to what they represent, so they must be interpreted to be understood. But mental episodes do not need to be interpreted to be understood. They ‘make present’ their representational objects in a way that symbols do not.[[24]](#footnote-24) So, rather than thinking of the phenomenon of mental representation along the lines of symbol-use, we should think of it as a form of conscious awareness.[[25]](#footnote-25) This form of conscious awareness—the ‘having in mind’ of something—is different from mere sensation, but, as a form of consciousness, it likewise resists reduction.

*4. Spontaneity.* The first three features I discussed are features of mental *states*, not mental *processes*. Accordingly, one might so far have the impression that, on my view, the *dynamics* of a mind is fully reducible to mechanical causes, even if the synchronic states linked by these causes are not reducible. But not so. I will now discuss two irreducible aspects of mental causation: spontaneity and rational insight.

To begin, let us return to the category of ‘executive function’. There are two ways to understand what is ‘executive’ about executive functions. One way is to treat them as architectonic but nevertheless mechanical functions within a mechanical system. Another way is to treat them as irreducible interventions *into* the mechanical systems. Volitional control, I maintain, must be understood in this latter way. Martine Nida-Rumelin makes the case in a particularly vivid and concise way:

If you observe a squirrel jumping from one branch of a tree to another, then the squirrel does not look to you like a mechanism that jumps as the result of some inner 'mechanical' process. It looks to you as though the squirrel itself, the subject of experience, does the jumping. … We see the movements of biological organisms that we implicitly accept to be conscious as being done by the conscious individual itself. A related claim is true for the way we experience our own doings. We experience our doings as brought about by ourselves. To assume that some inner processes cause our doings is incompatible with the content of the phenomenology of our experience. If these experiences of ourselves when we are active and our perceptions of others as being active are not illusionary, then conscious individuals are active in their doings.[[26]](#footnote-26)

In short, if ever a psychological subject is *active*—if it does more than ride the causal waves of the universe—then its behaviors are sometimes its very own, irreducible, spontaneous *doings*.

*5. Rational insight.* There are many varieties of rational insight, but the sort I am particularly interested here is an insight into how represented items ‘hang together’—a mental grasp of the explanatory and inferential relationships that thought-contents bear to one another.

Let me illustrate with an analogy. An episode of thinking consisting of the concepts *Elephants* and *walk* and *uphill* in that order is not, ipso facto, a case of thinking that elephants walk uphill. What is needed is *predicational structure*, and absent such structure, the episode is just a sequence of sub-propositional thoughts. Thinking whole thoughts, in other words, requires grasping semantic relationships that can hold among concepts—in particular, predicational relationships.

Analogously, an episode of thinking consisting of the thoughts *Elephants run uphill* and *Julie is an elephant* and *Julie runs uphill* is not ipso facto a case of inferring that Julie runs uphill, not even if the first two thoughts somehow *cause* the last one. Making an inference requires seeing that certain semantic relationships can hold among thoughts—in particular, inferential relationships.[[27]](#footnote-27) Of course, inferences can be *modelled* in a purely mechanical system. But the system would not be *drawing an inference*, strictly speaking. And intelligence, even of the most rudimentary sort, requires doing just that.

Here, then, is a second way in which the dynamics of mammalian minds involves more than mechanical causes. The dynamics of an intelligent mind must be governed by—at least sometimes, at least at some level—an awareness of irreducibly semantic relationships among thought-contents.[[28]](#footnote-28) Given that mammals are capable of rudimentary forms of inference (disjunctive syllogism, hypothetical syllogism), they must be capable of basic forms of rational insight—even if just the bare ability to tell when two grasped items are the same or different.[[29]](#footnote-29) Now, because problem-solving is slow and energy-costly, efficiency will dictate that frequently-deployed inferential processes be automated after all, as dedicated ‘modules’ (e.g., dedicated to the detection of predators or food-sources). But: that much apparently intelligent animal behavior is, in fact, the product of mechanical functioning does not mean that all of it is.

**5. The Human Mind.**

I will describe the transition from the Mammalian to the Human Mind in four ‘stages’:

1. *The accumulative stage*, in which Mammalian capacities are quantitively increased.
2. *The ampliative* stage, in which these accumulated capacities interact with each other in complex ways.
3. *The additive* stage, in which a qualitatively new capacity is superadded.
4. *The transformative* stage, in which this new capacity transforms the nature of the capacities shared with the Mammalian heritage.

I invoke these ‘stages’ heuristically, but I expect that they have corollaries both in phylogenetic and ontogenetic development. The first two stages are consistent with the Mammalian-Human difference being only one of degree. Whatever is controversial here will be a matter of detail, I suspect. But the third and fourth stages takes us into highly contested territory.

*1. The accumulative stage.* The uniqueness of the Human Mind is built on the foundation of increases of C-capacities on two fronts: executive functioning and social cognition.[[30]](#footnote-30)

Human executive function is more robust than mammalian executive function in all of its dimensions: humans can think longer and harder and about more things than can mammals. One particularly important advance is the ability to engage in in what cognitive scientists call ‘mental time travel’: recalling past episodes (‘episodic memory’, which is rare in the animal kingdom, maybe entirely absent[[31]](#footnote-31)) and imagining future episodes. The human *imagination*, in other words, is leaps and bounds better than the mammalian imagination, and this development—along with other quantitative improvements in executive function—opens up new possibilities for procedural rationality. Whereas mammals can solve the occasional problem, humans deliberate and create.

The basic social skill that humans have but mammals lack is, according to Michael Tomasello (2019), ‘joint attention’. A prototypical example of joint attention occurs in the context of a child playing with an adult. Each participant is aware of the object of their individual attention—some toy, say—but also of the other’s awareness of that object. Joint attention is crucial because it forms the basis of human communication. Every successful linguistic exchange is an invitation extended and received to attend together to some information or other. What makes joint attention possible is not a settled matter, but at least two ingredients seem essential. One is the mundane fact that humans have whites around their irises which makes possible gaze-following. But another (which exploits the first cognitive advancement) is a much-improved capacity for ‘mindreading’ (or ‘theory of mind’). While some mammals can track what another creature intends to do or knows, humans track enormous amounts of such information, about particular individuals, keeping it updated in real time.

Plausibly, there is an additional accumulation of A-capacities that is relevant here, viz., a greater motivational weighting of the social emotions. For example, while chimpanzees extend help to others, human children are much more eager to help, from an earlier age.[[32]](#footnote-32) It isn’t clear to me, though, that accumulations of C-capacities are insufficient to translate existing social emotions into such helpfulness.

The basic Mammalian-Human differences with respect to executive function and social cognition are all quantitative, and moreover, almost if not entirely restricted to C-capacities. But special things start to happen when these, at this advanced human quantitative threshold, begin to interact with each other and with the rest of the Mammalian inheritance.

*2. The ampliative stage.* Language, learned by exploiting enriched C-capacities and in the context of joint attention, is the catalyst for a plethora of mutually-enhancing feedback loops among the various capacities in the Human Mind.[[33]](#footnote-33)

Two features of language make its possession into something of a cognitive super-charger: it is *symbolic* and it is *communal*. Mental content, recall, is *not* symbolic. I can only entertain a mental content if I can consciously *grasp* it. But what can be symbolized does not have to be grasped. A symbol, once given a conventional meaning by a linguistic community, can be used by to mean that thing by any member of the community—whether or not that user is even capable of grasping of the meaning. Moreover, language involves systematic relationships among symbols, so a user can use language to begin mapping a whole body of knowledge prior to knowing what the symbols mean.

Why is this important? Language is the tool whereby the Mammalian inheritance of basic perceptual concepts (e.g. object, quantity, space, animacy, and agency) is leveraged into the full suite of concepts that humans use to make sense of the world. Susan Carey (2009) gives the canonical example of this process. Core cognition includes a system for tracking only very small quantities (up to 3 or 4 items), and yet humans learn to count. How? First, it turns out, by rote recitation of the number series, then by mapping magnitudes from core cognition onto the number series, and then—*eventually*—grasping the successor rule (successive numbers are the same as the previous number +1). More generally, words serve as place-holders for concepts that are acquired via such processes of ‘structure mapping’.[[34]](#footnote-34) And this leveraging of language does not just occasion the acquisition of discrete concepts. *Syntax* serves as a place-holder for various semantic relationships that contents can bear to each other. Language as a system of syntactic rules serves as an entre into a system of semantic relationships: the space of reasons.[[35]](#footnote-35)

Now, describing the developmental contexts wherein humans part company from mammals and acquire new concepts and new rational insight—and saying that language is central to it—does not yet tell us *what it is* to acquire a new concept or new rational insight. That last step in Carey’s process—from learning the number series to grasping a new rule—is unexplained.[[36]](#footnote-36) There are at least three different ways of understanding such conceptual expansions. First, the mind might acquire a new representational primitive, a conceptual content it could not grasp and now can. This sort of shift does seem to happen, and—it’s important to note—Carey’s developmental theory does not really tell us how it happens, just *when* it happens. Second, the mind *metaphorically extends* a conceptual domain it grasps, in order to map a domain it cannot grasp. For example, it has been suggested that the human concept of time, which the Mammalian Mind lacks, is a metaphorical extension of the human concept of space.[[37]](#footnote-37) Finally, at the other end of a continuum, one’s new concept could be an entirely *deferential* one. That is, one could use a term according to communal convention without grasping very much of its meaning (besides, perhaps, the very general category in which its meaning falls—*physical thing*, *physical* *stuff*, *social* *activity*, *mental state*, etc.). In such manner, a human can think about things she has never encountered and lacks the expertise to understand.

In sum: language ‘outsources’ meaning; 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. And the result is that the human can think about not just those categories that are built into the architecture of her mind but about anything at all.

A human’s greatly expanded C-capacities have a recursive, ampliative effect on A-capacities. In particular, humans develop a tertiary tier of emotions (without losing the primary and secondary tiers) that develop in response to a world richly conceptualized: feeling anxiety about an upcoming exam, feeling hopeful about reconciling with a friend, etc.[[38]](#footnote-38) Importantly, not only do these emotions have cognitive sources, they have representational contents as well. As Patricia Greenspan puts it: “Emotions that represent their objects in some positive or negative light (as most do) may be said to have a content expressible by an evaluative proposition.”[[39]](#footnote-39) Not only do such emotions *feel* good or bad, but such valences represent something else *as* good or bad. In other words, in the Human Mind, emotions begin to represent *value*.

How do they come to do so? Answering that question turns on the question of how humans acquire their normative concepts, and this is a disputed matter, but all seem to agree that social interaction plays a large role. Here is one account, from Carpendale & Lewis (2020):

[Moral obligation] is already implicit in the human developmental system as a result of the nature of early relationships. Infants are treated as persons, as participants in interaction. It is the product of treating others as persons and responding to them in everyday activity… Caring and mutual affection are embedded in the structure of the human developmental system. These strong emotional bonds are the seed for mutual respect, which is already there in communication, and develops increasingly into moral obligation. [[40]](#footnote-40)

The suggestion here is that interpersonal interaction transforms a certain kind of *feeling*—delighting in the presence of another—into a certain kind of *knowing*—that the other is valuable. The very feeling itself becomes a means of apprehending normative reality. In support of this idea, Dahl & Killen (2018) report that the helping behaviors of toddlers have much less to do with what others need and more to do with what will engage others interactively, but that this pattern is replaced by more need-centered helping in the next few years of development. That is, the development of human moral psychology starts with taking delight in interacting with others, proceeds to an awareness of others’ concerns, and culminates in the grasping of norms of rights and fairness—from attachment to partiality to impartiality.

Now, emotions are not the only way that humans represent value. We have coined the word ‘value’, for example. And this means that our capacity to represent value can come untethered from particular emotional experiences and appropriated into one’s cognitive economy. One can form beliefs about what matters, in other words. One can, furthermore, form such beliefs on the basis of explicit moral instruction. Subsequently, tertiary emotions can be generated that share representational contents with these moral beliefs. Thus, human emotions are to the world of value what sensory qualities are to the world of perceptibles. 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 hues and pitches and the like, and even our ‘offline’ concepts of physical objects carry traces of these experienced qualities. 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.

The effects of the foregoing on V-capacities are dramatic. Recall that inputs to mammalian V-capacities are (1) affective goading (what the animal wants or likes, including the company of other animals) and (2) procedural planning, of a rudimentary sort. With the explosion of language-induced cognition, the sheer quantity of such inputs, in both categories, is massively expanded. But there will also be a new *type* of input: awareness of the good, as such. In addition, an awareness of *time* will greatly expand the time-scale relevant to the exercise of control. Mammals navigate situations; humans pursue projects. Thus, the output of human V-capacities are not *behaviors*, but rather *intentions*, including temporally-extended varieties such as ‘standing’ intentions or ‘distal’ intentions.[[41]](#footnote-41)

This is an awful lot for a Human Mind to manage. Perhaps it could be managed with an exponentially larger working memory capacity, but our mammalian heritage has not bequeathed such resources to us. Rather, we manage our volitional life by pushing as many decision-making processes as possible out of consciousness entirely, automating them. One name for this is ‘secondary modularization’—the mind’s construction of automatic procedures dedicated to various cognitive tasks, analogous to the ‘primary modularization’ that underwrites core cognition and other mammalian cognitive tasks.[[42]](#footnote-42) Other names are ‘habit’ and ‘expertise’.

A final source of dramatic ampliative change comes in the form of social interaction. We have already mentioned two ways that social interaction shapes the Human Mind: it provides one with language, a cognitive super-charger; and it supplies the context in which *value* is first apprehended and conceptualized. Of course, the formative influence of one’s social world really kicks in once these initial contributions have already been made. Humans *teach* one another, broadly, extensively, explicitly, and universally, whereas mammalian teaching is extremely rare.[[43]](#footnote-43) One of the things that humans teach one another is *how to reason*. Left to themselves, individual humans are not great at thinking things through. Of course, we exhibit rational inight into semantic connections (as do mammals, as I claimed above), but we do not exploit those connections in deliberation particularly well. We make lots of mistakes, we fail to consider alternatives, and so forth. It is when we start giving reasons *to each other* that we get a better sense for what is relevant and what is not.[[44]](#footnote-44)

*3. Additive stage*. The Human Mind, as we have depicted it so far, is already very different from the Mammalian Mind. It thinks differently: about anything at all; it feels differently: evaluatively, in ways that are shaped by its beliefs; and it wills differently: it concerns itself with the good and with the future. It is debatable, I think, whether the introduction of normative elements into A-capacities and V-capacities amounts to a *kind*-difference rather than a mere *degree*-difference; probably sorting this out would require a more precise account of the transition than I have given. But it doesn’t matter, because what I have described so far is not yet a person, not a yet a fully rational nature. What still needs to be added is *self-consciousness*.

To motivate this claim, I begin with two evocative statements in support of a Mammalian-Human kind-difference. First, Robert Spaemann:

Human beings…exist by distinguishing their being from their specific way of being, their specific ‘nature’. Their nature is not what they *are*, pure and simple; their nature is something that they *have*. And this ‘having’ is their being.[[45]](#footnote-45)

The second is from Christine Korsgaard:

[We] exert a deeper level of control over own [sic] movements when we choose our ends as well as the means to them than that exhibited by an animal that pursues ends that are given to her by her affective states, even if she pursues them consciously and intelligently. Another way to put the point is to say that we do not merely *have* intentions, good or bad. We assess and adopt them. We have the capacity for normative self-government, or as Kant called it, “autonomy.” … The distinctive character of human action gives us a whole different way of being in the world.[[46]](#footnote-46)

The person, the rational human *agent* as such, has a ‘whole different way of being in the world’. What is this different way of being? Spaemann characterizes it as a kind of distance between the human and her nature. The crucial difference between a non-rational and a rational one is not so much the capacities that characterize each nature but the relationship that the rational animal has to those capacities—the relationship of ‘having’, rather than of *being*, of definition.

Likewise, Korsgaard claims that the human ability for ‘normative self-government’ is explained, not by a unique type of motivation, but by a unique relationship that she has to her motivations. They are not ‘hers’ automatically; she *adopts* them, or not.[[47]](#footnote-47) And she can do this because of the unique way she is aware of them.

Korsgaard calls the unique relationship ‘self-consciousness’. It is now standard to appeal to self-consciousness as a distinguishing feature of personhood, an emphasis that goes back to John Locke.[[48]](#footnote-48) But what can get lost in discussions of this distinguishing feature—which Korsgaard preserves—is the dramatic novelty of this psychological capacity. What Korsgaard has in mind is apt to be conflated with two lesser psychological phenomena: (1) introspection, in which a conscious state is directed toward itself; and (2) what Elisabeth Schechter calls ‘implicit self-awareness’, or the ability to have self-concerning attitudes. (The ‘mirror-test’ is a test of implicit self-awareness, argues Schechter.) Nor, I think, should self-consciousness be identified with the possession of an I-concept. Presumably no one can deploy, or perhaps even possess, an I-concept unless one is self-conscious, but self-consciousness is the more fundamental psychological phenomenon.

What, then, *is* self-consciousness? Scientists of consciousness study not only the contents of consciousness but the ‘level’ of consciousness. The paradigm here is the distinction between dreaming and wakefulness. I want to suggest that self-consciousness is as different from merely being conscious (even in sophisticated ways) as wakefulness is from dreaming. It is a new *level* of consciousness. (And it is one we drop in and out of, just as we drop in an out of being awake.) It is the dawning of an inner light, a ‘coming home’ to oneself as a self. Unless it is present, the subject is not self-governed, but rather nature-governed—even as she executes the deliverances of her capacities in spontaneous action.[[49]](#footnote-49) And if it is present, “it creates a new kind of psychological being,” as Schechter puts it.[[50]](#footnote-50)

*4. Transformative stage.* In “Additive Theories of Rationality: A Critique,” Matthew Boyle argues that the kind-difference between rational and non-rational animals cannot be accounted for simply in terms of the addition of a new capacity, because the rational part of the resulting mind would be at odds with the non-rational part. “What entitles us to hold that this reflects a fracture within a single subjective standpoint,” he asks, “rather than a struggle between two essentially distinct standpoints for control of a certain body?”[[51]](#footnote-51) Boyle mentions Korsgaard as someone whose theory of the rational/non-rational difference could lead to this problem.

I am not treating self-consciousness as a separate psychological *faculty* added on to the rest, so my proposal is already in a Boylean spirit. Moreover, while I do not think that self-consciousness is merely a result of ampliative interactions among human faculties, it is consistent with my picture that self-consciousness is a necessary catalyst of some of the ampliative processes I have described. For example, it is plausible that human social cognition requires a kind of self-other differentiation that is only possible for a self-conscious being.[[52]](#footnote-52) Thus, any capacities built on the foundation of social cognition—including the transition from affectively liking someone to valuing her/him—require self-consciousness.[[53]](#footnote-53)

Nevertheless, we can imagine that one who lacked self-consciousness and suddenly had ‘the lights turn on’, so to speak, could very well feel alienated from her own cognitive, affective, and volitional states. (This is, of course, the reaction typified by characters in science fiction who are freed from various kinds of mind-control.) This type of alienation is impossible for a being who lacks self-consciousness. There is, I suggest, a diachronic process of unification that occurs in a self-conscious agent. To use Korsgaard’s language, the self-aware agent *assesses* and *adopts* certain of her inclinations, rejecting others. She builds habits. She makes commitments. She constructs a self. Or tries to; the process can fail. If it succeeds, the psychological unity that results is, at least in part, of her own making.[[54]](#footnote-54)

**Conclusion: Becoming Custodians of Value**

According to the traditional view, we humans are *persons*: individual substances of a rational nature. Non-human animals are not persons; they lack a rational nature. In our endeavor to make sense of this traditional view, we started from the notion of ‘rational nature’ understood as an ideal: a *custodian of value*, one whose cognitive, affective, and volitional capacities are oriented toward the good. We then described rudimentary Vertebrate Minds and built up from there to the Human Mind, using only three resources.

1. Five primitive psychological features shared with animals: subjectivity, sentience, representation, spontaneity, and rational insight.
2. One primitive psychological capacity unique to humans: self-consciousness.
3. Functional complexity.

Are these resources sufficient to build a ‘rational nature’ that is different in kind from non-human animals? In a sense no, and in a sense, yes.

The sense in which it is not is that our notion of a rational nature is an *ideal*. Having a Human Mind does not guarantee that one meets the ideal. One is not born, but rather becomes, a person—and not simply by standard maturational stages but by free acts of normative self-governance.

But in another sense, yes: the Human Mind is rational in the sense of being *capable* of such normative self-governance. The ideally rational mind, recall, will (1) know what is valuable and according to what priority-ranking; (2) know how to respond appropriately; (3) desire to so respond and delight in doing so; (4) freely and consistent choose to so respond. The Human Mind can do these things, but (we might say) *just barely—*in peculiar and limited ways. Or, from another angle, in stunningly economical ways—repurposing every resource it can from its Mammalian heritage.

The Human Mind realizes (1) by repurposing Mammalian emotions (especially social ones) as representations of value.

It realizes (2) by repurposing Mammalian intelligence—in particular, (a) repurposing Mammalian understanding and problem-solving capabilities as general-purpose reasoning abilities, (b) expanding and calibrating these abilities by training automated cognitive modules (mental ‘habits’) while exchanging reasons socially; and (c) repurposing Mammalian core cognition in tandem with outsourced linguistic meanings, thereby expanding its representational repertoire indefinitely.

It realizes (3) by joining its affective system with the jerry-rigged cognitive system just described, creating a new ‘tertiary’ tier of emotional responses.

It realizes (4) by repurposing Mammalian spontaneous volitional control as bona fide free agency—empowered by self-consciousness, and expressed in the training of automated behavioral modules (that is, habits).

In short, though we humans are not, eo ipso, persons, we have the potential to become persons: we have the requisite psychological endowments to make the journey. Our capacity to make that journey means that we are different kinds of things from our non-human terrestrial companions, despite our sharing so many of their psychological endowments.

I have been defending the metaphysical dimension of the traditional view—viz., the claim that humans, qua persons, are creatures of a different kind from all non-human animals. It is worth returning now to the moral dimension of the traditional view. After all, it is not clear that the metaphysical dimension can even be framed in isolation from the moral dimension, since the relevant notion of different kind’ is a *morally* relevant notion. What is it about persons, then, that makes them uniquely morally inviolable? Why must we respect persons in ways and to degrees that do not apply to non-persons?

There may be different aspects of personhood that call out for different kinds of respect for different reasons. First of all, the summons to be a custodian of value is ennobling. Perhaps one who respects value is, for that reason, to be respected. This is at least one important sense in which persons have ‘dignity’.

Second, as beings who order their lives around what is valuable in temporally-extended ways, persons have *interests* in ways that non-persons do not. For this reason they also seem capable of *flourishing* in ways that non-persons are not. Humans pursue and delight in what is valuable as such. Perhaps the ‘pursuit of happiness’, so construed, calls out for unique respect.[[55]](#footnote-55)

Without denying the pull of these suggestions, I am inclined to think that there is a more basic ground of moral status exhibited by persons: their selfhood. Recall our discussion of the source of value-concepts in interpersonal interaction. When a person sees another self as a self—as a ‘you’, a self who looks back understandingly at oneself—the respect-worthiness of the other is immediately intuited. In such interpersonal recognition, a dual movement is initiated: the movement of affinity, toward connection and care, and the movement of deference, giving another space to be and to move. As persons, we intuit that persons are valuable and we intuit the proper response to their value, in the form of this double movement. I doubt there is much more to say about the matter. It is the basis on which we build our moral and political systems, but admits of no further accounting.

References

Asma, S. T. & Gabriel, R. (2019). *The emotional mind: Affective roots of culture and cognition*. Harvard University Press.

Asoulin, E. (2019). Phrase structure grammars as indicative of uniquely human thoughts. *Language Sciences*, *74*, 98-109.

Bluff, L. A., Troscianko, J., Weir, A. A., Kacelnik, A., & Rutz, C. (2010). Tool use by wild New Caledonian crows Corvus moneduloides at natural foraging sites. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1686), 1377-1385.

BonJour, L. (1998). *In defense of pure reason: A rationalist account of a priori justification*. Cambridge University Press.

Boyle, M. (2012). Essentially rational animals. In *Rethinking epistemology* (pp. 395-428). de Gruyter.

---. (2016). Additive theories of rationality: A critique. *European Journal of Philosophy*, *24*(3), 527-555.

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, *40*.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Carroll, L. (1895). What the tortoise said to Achilles. *Mind* 4 (14):278-280.

Carpendale, J. I., & Lewis, C. (2020). Tomasello's tin man of moral obligation needs a heart. *Behavioral and Brain Sciences*, *43*.

Chalmers, D. J. (2003). Consciousness and its place in nature. *Blackwell guide to the philosophy of mind*, 102-142.

Cheney, D. L., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour*, *28*(2), 362-367.

Crystal, J. D. (2010). Episodic-like memory in animals. *Behavioural brain research*, *215*(2), 235-243.

Dahl, A., & Killen, M. (2018). A developmental perspective on the origins of morality in infancy and early childhood. *Frontiers in psychology*, *9*, 1736.

De Waal, Frans. (2006). *Primates and philosophers: How morality evolved*. Princeton: Princeton University Press, 2006.

Doris, J, Stich, S., & Phillips, J. (2020). Moral Psychology: Empirical Approaches. *The Stanford Encyclopedia of Philosophy*(Spring 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/moral-psych-emp/>.

Eberl, J. T. (2017). The ontological and moral significance of persons. *Scientia et Fides*, *5*(2), 217-236.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752-775.

Greenspan, P. (2004). Practical reasoning and emotion. *The Oxford handbook of rationality*, 206-221.

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, *16*(2), 129-219.

Hoppitt, W. J., Brown, G. R., Kendal, R., Rendell, L., Thornton, A., Webster, M. M., & Laland, K. N. (2008). Lessons from animal teaching. *Trends in ecology & evolution*, *23*(9), 486-493.

Korsgaard, C. M. (2006). Morality and the distinctiveness of human action. In *Primates and philosophers* (pp. 98-119). Princeton University Press.

Laland, K., & Seed, A. (2021). Understanding human cognitive uniqueness. *Annual Review of Psychology*, *72*, 689-716.

Mele, A. R., & William, H. (2009). *Effective intentions: The power of conscious will*. Oxford University Press

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.

Nida-Rümelin, N. (2007). Dualist Emergentism. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), Contemporary Debates in Philosophy of Mind. Blackwell.

O'Connor, T., & Jacobs, J. D. (2003). Emergent individuals. *The Philosophical Quarterly*, *53*(213), 540-555.

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 8993-8999.

Schechter, E. (2018). *Self-consciousness and" split" brains: The minds' I*. Oxford University Press.

Shettleworth, S. J. (2013). *Fundamentals of comparative cognition*. Oxford University Press.

Spelke, E. S. (2000). Core knowledge. *American psychologist*, *55*(11), 1233.

Spaemann, R. (2006). *Persons: The difference between 'someone' and 'something'*. Oxford University Press.

Suddendorf, T. (2013). *The gap: The science of what separates us from other animals*. Constellation.

Tomasello, M. (2014). The ultra‐social animal. *European journal of social psychology*, *44*(3), 187-194.

Tomasello, M. (2019). *Becoming human*. Harvard University Press.

Wasserman, E. A., & Young, M. E. (2010). Same–different discrimination: The keel and backbone of thought and reasoning. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(1), 3.

Woodward, P. (2015). *The emergence of mental content: An essay in the metaphysics of mind* (Doctoral dissertation, Indiana University).

---. (2018). A Posteriori Physicalism and the Discrimination of Properties. *Acta Analytica*, *33*(1), 121-143.

---. (2019). Phenomenal intentionality: reductionism vs. primitivism. *Canadian Journal of Philosophy*, *49*(5), 606-627.

---. (2021). Woodward, P. (2021). The selection problem for constitutive panpsychism. *Australasian Journal of Philosophy*, *99*(3), 564-578.

---. (under review). Consciousness and Rationality: The Lesson from Artificial Intelligence.

1. See Boyle (2012) for a helpful explication of the relevant kind-difference. [↑](#footnote-ref-1)
2. An important exception is Porphyry, whose treatise in defense of vegetarianism proves the rule, written as it is in an iconoclastic tone of voice. [↑](#footnote-ref-2)
3. So, Jonathan Bennett: “It is commonly believed... that between a genius and a stupid man there is a smooth slide while between a stupid man and an ape there is a sharp drop, not just in the sense that there are no creatures intellectually half-way between apes and stupid men, but in the sense that there could not be such creatures. Any possible creature whose intellectual level was higher than that of normal apes and lower than that of normal men—so the common belief runs—either would or would not have that special something which puts humans importantly above other animals.” (Qtd. in Boyle (2012), p. 397). [↑](#footnote-ref-3)
4. Jaworska & Tannenbaum (2019, p. 342) characterize this commonsense view as follows: “all human beings, whether adults or infants, and whether cognitively (i.e., intellectually or emotionally) unimpaired or severely cognitive impaired have FMS [full moral status] and certainly a higher moral status than most animals.” [↑](#footnote-ref-4)
5. Cheney & Seyfarth (1980). [↑](#footnote-ref-5)
6. Herman et al (1984). [↑](#footnote-ref-6)
7. Bluff et al (2010). [↑](#footnote-ref-7)
8. Hoppitt et al (2008). [↑](#footnote-ref-8)
9. Wasserman & Young (2010). [↑](#footnote-ref-9)
10. Crystal (2010). [↑](#footnote-ref-10)
11. Tomasello (2019), de Waal (2006). [↑](#footnote-ref-11)
12. Mercier & Sperber (2017). [↑](#footnote-ref-12)
13. See Doris et al (2020). [↑](#footnote-ref-13)
14. Shettleworth (2012), p. 17. [↑](#footnote-ref-14)
15. See Asma & Gabriel (2019), p. 8. [↑](#footnote-ref-15)
16. The framework of ‘core cognition’ (Carey [2009]) or ‘core knowledge (Spelke [2000]) was first developed in the context of human developmental psychology, but was subsequently used to study other mammals. See Shettleworth (2012), p. 120ff. Different theorists provide slightly different lists. [↑](#footnote-ref-16)
17. Laland & Seed (2021). [↑](#footnote-ref-17)
18. Asma & Gabriel (2019), p. 9. ‘Care’ is apt to be understood anthropomorphically, so let me stress that we are still talking about a type of *feeling* that motivates. [↑](#footnote-ref-18)
19. “For a reward forty times larger than the immediate reward option, chimpanzees may wait up to eight minutes.” (Suddendorff 2013, p. 109). [↑](#footnote-ref-19)
20. See Woodward (2015), pp. 84-94. [↑](#footnote-ref-20)
21. For a dualistic version, see Nida-Rumelin (2007), and for a non-dualistic version, see O’Connor & Jacobs (2003). [↑](#footnote-ref-21)
22. See Woodward (2018). [↑](#footnote-ref-22)
23. This is a summary of the ‘Structure and Dynamics’ argument, so named by Chalmers (2003). Alternatively, one could maintain that there is more to the brain than its structural and dynamical features that neuroscientists study: perhaps it also has a non-structural, qualitative ‘interior’. I argue against this ‘panpsychist’ option in Woodward (2021). [↑](#footnote-ref-23)
24. The locus classicus for this point is Bonjour (1998), ch. 6. [↑](#footnote-ref-24)
25. See Woodward (2019). [↑](#footnote-ref-25)
26. Nida-Rumelin (2008), p. 208. [↑](#footnote-ref-26)
27. This is the lesson of Lewis Carroll’s famous (and famously brief) essay, ‘What the Tortoise Said to Achilles’. [↑](#footnote-ref-27)
28. See Woodward (under review). [↑](#footnote-ref-28)
29. See Wasserman & Young (2010) for a review of the empirical studies on this particular ability among non-human animals. [↑](#footnote-ref-29)
30. See Suddendorff (2013) and Tomasello (2019). [↑](#footnote-ref-30)
31. Laland & Seed (2021). [↑](#footnote-ref-31)
32. Tomasello (2014). [↑](#footnote-ref-32)
33. There are essentially two camps regarding the prerequisites for language-learning. One camp, associated with Noam Chomsky, maintains that there is a uniquely human language-learning module. Another holds that executive function and social cognition are together sufficient. In other words, language-learning is a form of social pattern-detection. See Tomasello (2019), p. 127ff. I will assume the latter, though of course it will do no violence to my aims if the former turns out to be correct. [↑](#footnote-ref-33)
34. See Gentner (2010). [↑](#footnote-ref-34)
35. See Tomasello (2019) pp. 120-121, Asoulin (2019). [↑](#footnote-ref-35)
36. She calls the process ‘bootstrapping,’ as in “pulling oneself up by one’s bootstraps”—which is to say that it remains a bit of a mystery. [↑](#footnote-ref-36)
37. Pinker (2010). [↑](#footnote-ref-37)
38. Asma & Gabriel (2019), p. 9. [↑](#footnote-ref-38)
39. Greenspan (2004), p. 204. [↑](#footnote-ref-39)
40. Carpendale & Lewis (2020), p. 43. [↑](#footnote-ref-40)
41. For a discussion of these categories see Mele (2009). [↑](#footnote-ref-41)
42. On ‘secondary modularization’, see Burkhart et al (2017). [↑](#footnote-ref-42)
43. LaLand & Seed (2021). [↑](#footnote-ref-43)
44. Mercier & Sperber (2017). [↑](#footnote-ref-44)
45. Spaemann (2006), p. 31. [↑](#footnote-ref-45)
46. Korsgaard (2006), pp. 112-113, 117. [↑](#footnote-ref-46)
47. In places, Korsgaard does suggest that what marks the difference is a unique type of motivation, viz., a moral one, for example: “A form of life governed by principles and values is a very different thing from a form of life governed by instinct, desire, and emotion” (117). No doubt this is right, but it cannot mark the non-rational / rational distinction, since (1) the *self* needs to be doing the governing, not the ‘principles and values’, and (2) I have suggested that in principle a non-rational nature could be motivated by values (see my discussion in the previous few paragraphs). There are, in other words, two ways a being could fail to exhibit ‘normative self-government’: if its motivations fail to be normative, and if the self fails to be what governs. [↑](#footnote-ref-47)
48. A person, according to Locke, is “a thinking intelligent being that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and as it seems to me, essential to it: it being impossible for any one to perceive without *perceiving* that he does perceive.” [↑](#footnote-ref-48)
49. Why do humans have so much more prefrontal cortex than primates? Not primarily, as is often claimed, to give humans more inhibitory control, which is said to happen in the prefrontal cortex. Rather, this most striking neuro-anatomical difference underwrites our most striking psychological difference, viz., our capacity for self-awareness. Or so I speculate. [↑](#footnote-ref-49)
50. Schechter (2018), p. 192. [↑](#footnote-ref-50)
51. Boyle (2016), p. 549. [↑](#footnote-ref-51)
52. As Schechter (2018) argues convincingly. [↑](#footnote-ref-52)
53. Boyle himself implies that the acquisition of value-concepts is only possible in an already-rational mind: “While it may be correct to say that a nonrational animal’s desires present their objects as attractive (for instance, as promising pleasure or promising to relieve some distress), it cannot be correct to say more specifically that they present their objects as desirable (i.e., as meriting desire): this way for something to be attractive lies beyond the scope of a nonrational mind.” Ibid., p. 539. [↑](#footnote-ref-53)
54. Compare Schechter (ibid., p. 186): “The executive aspect of self-consciousness provides the basis for the self-constructing agent, who is more autonomous than non-self-conscious agents in being able intentionally to guide and direct her own behavior by conforming it only to principles of action that she herself accepts.” [↑](#footnote-ref-54)
55. Compare Jason Eberl (2017, p.226), writing in a speculative rather than committal mode: “It is arguable that a person possesses the moral status she does by virtue of her capacity to have *significant interests*, the frustration of which would cause her to experience a degree of harm beyond the pain that merely sentient non-personal animals may experience.” [↑](#footnote-ref-55)