

A Brief History of Mind

I Introduction

My aim is to give an overview of what minds are and how they came to be. Minds are a product of billions of years of evolution so it is a daunting task to summarize this history in 45 minutes. My attempt will involve vast oversimplifications, highly speculative and condensed “just so” stories, and a great amount of hand waving. In particular, I will presuppose the theory of evolution and will not attempt to either explain it or justify it.

I will start with a few comments on the role of historical understanding, though the significance of history for the understanding of mind will not become clear until later. I will then sketch out four stages in the history of mind: the pre-mental, mental, conscious and selfhood stages. I will conclude with a few general remarks about the significance of this approach.

II Methodology: Cause and History

When we try to understand things, sometimes our explanations are timeless. My explanation for why a triangle has a sum of angles equal to two right angles, or for why heat causes a piece of metal to expand, is just as valid five billion years ago as it is today. But if I explain higher prices for a product as caused by reduced supply, the explanation would not apply five billion years ago since prices had not yet been invented. The causal laws of economics depend for their validity on the historically contingent development of our system of money, capital, investment and free market. While we may be able to use economic laws for pragmatic purposes without referring to their history, a full understanding of why those laws work requires that we grasp their historical genesis. Unless we look at how trading, exchange and ownership have been constructed over the centuries, economic laws present themselves as brute facts, regularities for which we have no

understanding.

The study of mind is sometimes approached by psychologists, cognitive scientists or even philosophers like Descartes in a non-historical manner, but I maintain that mind is more analogous to economics than to geometry or physics and can only be fully understood by investigating its historical construction. Mental processes are not eternal like triangles or physical causation, but are essentially historical constructs, constructed largely by biological evolution, with a little recent help from society. I will argue that mental processes encapsulate the biological history of the species and so the story of the genesis of mind is not just an interesting academic curiosity, but is the only way to really understand what minds are.

III The Evolution of Mind

By the term “mind,” I just mean all the mental processes which go on in someone like me: my perceiving, my thinking, my feeling, and so on. I take it for granted that they go on in you too; otherwise, this colloquium would be a monologue. I will not be talking about some global Geist, an absolute ego, a world soul, or anything else beyond the individual. I've enough trouble understanding people like you and me.

Once upon a time, say five billion years ago, there were no minds on the earth. Today there are. My claim is that we got here from there by means of gradual steps in an evolutionary process. The nature of any gradual process of becoming is such that, although the beginning and the endpoints may be radically different, we are not likely to find clear-cut divisions along the way, so there may be many intermediate stages about which there is no definitive answer as to whether they are mental or not.

Five billion years ago, all that existed were processes governed by the laws of physics and

chemistry. Let me call this *stage zero* or the *stage of direct causation*. At this stage causal laws describe regularities between objects which belong to natural kinds. For instance, heating a piece of iron causes it to expand. This is a general natural law that applies to all pieces of iron; that is what it means to say that iron is a natural kind. Such a law is timeless.

Stage One: The Pre-Mental

Initially when life appeared, primitive organisms continued to be governed solely by direct causation. Perhaps glucose caused bacteria to grow, but iron caused them to die. But some organisms developed a more sophisticated pattern of reaction to their environment. They developed the means to use the same mechanisms of response for dealing with a number of different stimuli; that is they came to treat different physical stimuli as *the same*. Imagine for instance a simple organism, let me call it an amoeba, that connects its receptors for glucose, starch, triglycerides and amino acids to the same unifying input node which then releases the mechanism for the approach response. That is, it comes to treat these four chemicals in the environment as the same, as belonging to the one category, let us call it “food.” Similarly it may treat various other chemicals, such as iron, ammonia or methane as “poison” to be avoided. Since developing detectors for neutral substances, such as silicon dioxide, would be a waste of valuable biological energy, it is unlikely such detectors would evolve. Assuming that these developments give the amoeba an evolutionary advantage, then after many millions of generations, all surviving amoebas will operate in this way.

There are a number of different ways of analyzing this new development. First, we can see it as the emergence of a new supervenient causal law; let me call it the Amoeba Food-Approach Law. I call it supervenient to make it clear that there is nothing mystical about it. It depends for its force on the subvenient physical

laws, which lose none of their validity in the new situation. Indeed in any token case of the application of the Food-Approach Law, e.g., approaching a piece of sugar, we can explain completely what happens in terms of the underlying physical laws. The supervenient, Food-Approach Law, is nevertheless, a real law. We can distinguish a real, type-based law from a superficial re-description of token events on the basis of its counterfactual force. The Food-Approach Law not only explains those events which actually happen; it also allows us to say what would happen if, contrary to the facts, silicon dioxide could be digested by the amoeba. Once we understand that it is evolutionary fitness that brought about the law in the first place, we can see that silicon dioxide would have been included in the category of food if, contrary to fact, there had been a survival advantage in doing so. Nevertheless it is a truly emergent law into the sense that, given its counterfactual scope, it cannot be reduced to the subvenient physical processes. The law is a contingent law that only emerges because of the particular biological history of the earth during its first billion years. If history had been different, as it appears to have been on Mars, then a Food-Approach Law would not have come into being. Since its emergence is not necessary, but contingent, it could not have been predicted by, nor reduced to, physical laws alone.

The second way of understanding the new evolutionary situation is to see food as a new kind of object constituted by the organism. That food is constituted by the organism does not mean that food is *in* the organism. The food is in the world, otherwise the amoeba would not be approaching it! In this context, however, “world” refers to the world as it is for the organism. The world-in-itself, that is, the world directly governed by physical laws, is made up of objects such as glucose and starch; the world-for-the-organism is made up of objects categorized as food (or as poison, or whatever.) Already at stage one we have a clear distinction between the in-itself and the for-the-organism.

Uxhill talks of the *Umwelt*, the surrounding world of an organism. This world is made up of objects of significance for the organism, objects whose very essence is defined by the pragmatic needs of the organism. It is the organism which determines which objects in the world are granted the power to cause which responses. For these objects, the criteria for *sameness* and *difference* come, not from physics, but from biological needs. Objects are possible components of the organism's world in so far as they have a value for the organism.

Another way to interpret a situation is to see it as the genesis of a function. The node which unites the various receptors doesn't just detect food; its *function* is to detect food. In the first amoeba that detected food, perhaps by chance genetic mutation, the central node just happened to detect food, but after many generations, when later amoebas inherit this capacity precisely because their ancestors survived in competition with others because they had this new capability, the node takes on the function of being a food detector. That is its goal or *telos*. The arrival of functionality is also the arrival of normativity. Once the function of the node within the species is to detect food, then if, in the case of an individual amoeba, some accident leads to the node detecting a poison such as ammonia, this action is a malfunction, a mistake, an error, not just a neutral happening. Once there are functions, a process can be said to be right or wrong. It is only the evolutionary history which has established a norm against which the supervenient causal process can be judged to be correct or incorrect. This normativity cannot be seen if we look exclusively at the present physical events: only if we look at the events in the context of their evolutionary history does their functional normativity become visible. In the purely direct causal world of stage zero, the most we can say is that things happens; we cannot describe what happens as a failure. The food detector is like a crystallized embodiment of the previous history of the species. It is as if, in responding to a stimulus, a function is taking into account not only the current situation but

also the evolutionary history of the species. The organism's response is no longer just to the present stimulus, as is the case in stage zero. Already with our simple amoeba we have moved from the purely physical world of facts to a world which, due to its history, is defined by value.

It is important to recognize the systematic nature of this new development. It is not simply that, within the organism, a food detector has been invented. Nor is it simply that there is a new kind of object, food, in the world, in the *Umwelt*. It is not even enough to say that these two events are correlated. What we need to grasp is the evolutionary creation of a new system that sets the context for its parts. Only within the holistic structure of the organism-environment system does the object "food" exist; but it is also only within that same system that the node within the organism can be understood as a "food detector." Just as money, prices, supply and demand, and so on are only defined within an economic system, so both organism and environment, detector and object are defined within an ecological system that sets the context for both.

What distinguishes the system of stage one from the direct causal relationships of stage zero is what one might call *indirectness*. The direct coupling of stimulus to response has been interrupted by the insertion of a detector that encapsulates part of the history of the species and allows the resultant structure a measure of control over the input before the response is released. This is an evolutionary trick that we will see repeated again and again at different levels throughout our brief history of mind.

So where are we? Are we dealing with mind yet? Well, a dualist might say, following a principle explicitly stated by Descartes, that if we cannot account for this system on the basis of purely physical principles, then we must attribute it to mind. But I am no more a dualist than I am a physicalist. There are many different levels of reality and whether we care to label stage one "mental" is arbitrary. The issue

has more or less the nature of a disciplinary dispute: should this stage be studied by biologists or by psychologists? To which agency should we apply for funding for our investigation? From the viewpoint of the ontological status of the amoeba, the question is empty. If we already understand how the stage one system operates, labelling it “mind” or “non-mind” helps us no further.

Stage Two: Mental Processes

Let me now skip forward two or three billion years, in this long Odyssey. By then some organisms have developed the ability to keep track of one individual object through various sense modalities. Let me call this new mechanism an Object Tracker. This kind of organism integrates inputs from various sense modalities and treats all of them as indicators of the one selfsame object. Thus, a sabre-toothed tiger may detect scents, sounds and sights and treat all these inputs as modes of detecting the one antelope. It is a *tracker* in so far as this neural structure covaries with changes in the world: if the antelope runs left or right, speeds up or slows down, etc., then the Object Tracker is updated accordingly. If information by vision is unavailable, it is updated by smell, or whatever other information channel is available. Even if no current information is available (perhaps the antelope has run behind a clump of bushes), the Object Tracker may use dead reckoning to estimate the current position of the antelope. That is, the Object Tracker may operate even in the absence of an actual object.

Once again, we have the interruption of a relatively direct process and the introjection of a new point of flexibility, which allows the integration of information from other sources. The Object Tracker permits the organism to allow information from other sensory input systems, from earlier experiences, and from the history of the species to influence the unleashing of any response. These evolutionary developments partly insulate the organism from the environment. Instead of direct causation, we now have delayed and indirect reactions. How

the world is categorized, and how the organism reacts to it, depend not solely on the stimulus from the world, but also on the internal structure of the organism, that is, its history and values. We can now distinguish the kind of system in which the reaction is a direct effect of an environmental cause from the organism-controlled system in which it is the internal, historically generated, structures of the organism which control the response. The organism has developed, as it were, a certain autonomy from its environment, a control which depends on its accumulated history.

One can describe this stage as the constitution of enduring, or substantial objects, objects which persist in the world, even when they are not being detected. I am, of course, still speaking of objects in the *Umwelt*: objects as they are for-the-organism, not as they are in-themselves. These are pragmatic, value-laden constitutions: antelopes are all categorized as the same in that they are prey, that is, potential food, and categorized as different from, say, trees, which must be avoided but not eaten.

The *constitution of stable objects* requires that the organism keep track of its own situation. The amoeba swimming towards food needs no information about its own position. The tiger, however, can only constitute a stable antelope if it keeps track of information about its own perspective. If the tiger turns its head so that it can no longer see the antelope, it must discount the absence of visual input and attribute it to the position of its own head, and not to the vanishing of the antelope. If the tiger is to intercept the antelope when it reappears from behind a clump of bushes, it needs to be able to predict its own position at that time. A world of stable objects requires a certain minimum knowledge by the organism about its own state.

Are we here yet? Is this mind? If by mind we mean a conscious, self reflective subject aware of its own existence, then surely not. Yet the kind of processes involved in the constitution of stable objects are the kind studied by psychologists and cognitive

scientists, so they might naturally be called mental processes, even though unconscious. Sterelny, who refers to what I call Object Trackers as “representations,” therefore considers this stage the appearance of mind. My worry about this way of putting it is that the Cartesian dualist might misunderstand this designation as a claim that we have here a conscious subject, a danger exacerbated by the ambiguities of the word “representation.” For centuries philosophers have used the word “representation” to translate *Vorstellung*, that which stands out as present before a conscious subject. In the Cartesian perspective, such “representations” are entities within the mind, and it is those entities that we are conscious of. But what Sterelny means by a “representation,” what I prefer to call an Object Tracker, is a neural structure in the brain, and almost no one claims that we are aware of our brain structures. My tiger may perceive *by means of* an Object Tracker in its brain, but *what* it perceives is an antelope, which is an object in its world, in its *Umwelt*, not in its brain. So while I still think the question is an empty one and involves arbitrary labelling, with these provisos I see little harm in going along with Sterelny's designation and accepting that what we have at stage two is indeed a mind, though not yet a conscious self.

Stage Three: Reflective Consciousness of Mind

So far I've discussed the input side of the organism and claimed that, as evolution develops, the Object Tracker mechanism becomes more and more decoupled from the direct causal sequence of stage zero. A similar decoupling of direct links occurs on the output, or response side. In stage two, once the tracker has registered a certain kind of object in the world, such as a piece of food, then an *instinctive* response of approach is automatically released by a direct causal process. In stage three, this automatic release is interrupted and

the tracker findings are redirected to a way station which may or may not release the approach mechanism depending on other factors. Let me call this way station an *inclination*. How, or if, an inclination gets implemented depends on the existence of other, possibly competing, inclinations, on information from other trackers and, again, on the history of the species. Further, once an inclination has been activated, unlike the case of instinct, the possibility of selecting between alternative ways of fulfilling the inclination arises. The *selection* mechanism therefore is not hardwired to any specific response, but takes on the task of selecting whatever response will most effectively carry out the goal towards which the inclination is directed. An *instinct* unleashes a specific response. An *inclination* sets a more abstract goal and allows for flexibility of response and the integration of whatever other information is available to the organism including information. Even after it perceives an antelope, our tiger may follow his inclination to lie in the sun and postpone working for food until tomorrow. With the substitution of inclination for instinct we have one more step toward self-control, and the overcoming of the direct domination of the input.

Sterelny refers to any Object Tracker that is decoupled from a single, rigid response, as a *belief*. In so far as an Object Tracker may remain activated for a period of time, even after the object itself ceases to be present, and in so far as the tracker no longer automatically triggers a response, this designation appears to me to be plausible. We can say that our tiger lying in the sun has a belief that there is an antelope, though it does not act on the inclination this belief induces. If we were now to interpret the inclination module as a *desire*, which is also plausible, we have an organism whose actions are to be understood in terms of belief and desire, which for many is the hallmark of mind. In that case the selection module should be understood as *decision*.

So far in our story, what organisms react to in

the world are objects, including animals, and their behaviour. A monkey may see a snake crawling towards it and so rapidly climbs a tree. But once animals act on the basis of beliefs and desires, there is an evolutionary advantage to being able to recognize and respond to what other organisms, especially conspecifics, believe and desire. There is likely to be an evolutionary advantage for a primate that is able to recognize that another, bigger and stronger primate, believes that the first has stolen his food, whether it is true or not. A primate who fails to distinguish between another's anger and its desire to be friendly may not survive long. Some authors referred to the capacity to attribute beliefs and desires to others as the possession of a *Theory of Mind*. I think this way of putting it is somewhat misleading. First, we are talking of an unconscious skill rather than the kind of explicit intellectual notion that we call a theory. Secondly, I think it very unlikely that some primate or hominid had the Eureka experience, "Ah! They have minds!" It is much more likely that there is a piecemeal evolution, so that one might develop the ability to attribute mental states of anger, while still remaining unable to recognize, say, jealousy. Since evolution proceeds pragmatically, the ability to notice specific mental processes in others probably developed gradually on the basis of the survival utility of identifying that specific process. Furthermore, since we are dealing with needs-based categorizations of the *Umwelt*, the way these mental processes are categorized depends on practical considerations. For example, the different states that we would today describe as anger, rage, irritation or grumpiness might all perhaps have been categorized together as "the kind of hominid to stay away from."

Learning to attribute some states of belief, desire or emotion to others opens up the possibility of attributing them to ourselves. This possibly is reinforced by the fact that those I'm dealing with have beliefs about my mental states, and I can hardly take these beliefs into account without attributing some mental states

to myself. In other words, just as the organism tracks significant objects in its external environment, it may come to track significant elements in its own inner environment. To call this tracking of one's own states "reflection" is to invite misinterpretations. There was not one global moment of enlightenment when one came to think that they were a thinker. Rather the ability to track specific thoughts or desires of one's own developed insofar as there was some evolutionary advantage to each of these trackings. Hence I might develop the ability to reflect on one process without the ability to reflect on many others. "Reflection" also misleadingly suggests some unified method of introspection. But there are various different bases on which I might learn to attribute mental states to myself: I discover I'm irritable, for example, when someone else points it out to me; I find out that I'm hungry, by input from some inner sense; my jealousy is revealed to me when I noticed my tendencies to act in certain ways. I suspect each of these reflective capacities developed gradually, independently of each other over a period of millions of years. Even today, our capacity for self reflection remains spotty; how else would psychoanalysts make a living? A sudden awakening to the truth of "Cogito ergo sum" is the unlikely figment of some philosopher's imagination.

In any case, however it came about, the ability to track our own inner states is an essential step in the development of mind. Indeed many might ask whether this is not *the* essential step. Is it not reflection on our own inner states which constitutes consciousness and how could there be "mind" without consciousness? This is another empty question. Label things whatever way you like! The important thing is to understand that at the end of stage three we have hominids who keep track not only of objects-for-the-organism in the *Umwelt* but also of some of their own inner states, as well as of the inner states of some other organisms. For convenience, I would like to label this the stage of conscious minds.

Stage Four: Selfhood

But our evolution is not yet over. We are more than conscious minds. There are yet further levels of decoupling and indirection to come. In stage three, Object Trackers activate an inclination or desire that then sets the goal for which the selection module chooses the best means. But once desire can be represented, it becomes possible to decouple the direct link from Object Tracker to desire so that the desire itself can be evaluated before being acted on. Charles Taylor distinguishes between weak evaluators and strong evaluators: my stage three organisms are weak evaluators, whose values are set by their first-order desires; a strong evaluator, however, evaluates these first-order desires themselves. Once again, we can see this step as the interruption of a direct process and the opening up of an opportunity for the integration of considerations of wider concern. If a stage three (conscious mind) organism recognizes that someone has slapped him on the cheek, then a desire for vengeance is triggered and he starts to select the best means of getting his own back. A strong evaluator, however, may interrupt the automatic triggering of the vengeance desire provoked by the slapping and consider whether he wishes to be kind of person who will respond with vengeance to aggression. That is, a strong evaluator makes his first-order desires themselves objects to be tracked, analyzed and modified.

Let me call this fourth stage of the evolution of mind, selfhood. One of the most central features of being a self is integration over time. At stage three, we had a mind which was aware of itself at the current moment but had no concern or awareness for its past or for what might happen to it in the future. A self, however, understands its current mental states as related to its past and to its future. Many have claimed that this integrating relationship is one of narrative. They say that, in so far as the mind tells itself a story about its past, it attributes these past states to an enduring self, which it

thereby constitutes. In my opinion, this account fails if by narrative we understand just a simple chronicle of events laid out in a sequence. If each event or mental state is simply juxtaposed in time to the next one, I see little basis for the strong unity we attribute to a self. I think the relationship of simple, temporal juxtaposition must be replaced by the relationships of responsibility and commitment. It is by accepting responsibility for actions in the past, and by making commitments in the present that would be binding on the future, that mental states distributed over time are constituted as episodes in the life of one, self-identical, enduring self. I'm speaking here not just of moral responsibility, but of the kind of basic responsibility which unites over time *all* actions and mental states, moral or amoral, as *mine* and so constitutes personal identity from birth to death.

I'm not suggesting that there is no temporal unity before the establishment of the self. I've argued that even an amoeba's current state is a kind of crystallized history. Indeed the essential feature that distinguishes each of my evolutionary stages from stage zero, from purely physical direct causation, is precisely the integration of history so that the past can have a structuring effect on the present. What is distinctive about the temporal integration of stage four is that it is governed by a higher order or strong evaluation, a set of norms on the basis of which I monitor my ongoing life and interpret past actions and future hopes in terms of an ideal I set for myself, an ideal of the kind of person I wish to be. This kind of self-monitoring requires language and symbols. However, these are not enough. Something more is needed for selfhood. Although human language has been around for perhaps 100,000 years, and may even be a requirement for the consciousness of stage three, the idea of creating and caring for the self is probably a more recent development, maybe only a few thousand years old.

Of course none of this is possible for a solitary individual. Languages and symbols are

social constructs, and so is selfhood. Only a culture with the ideal of responsibility, with the meme of selfhood, can give rise to individual selves. Responsibilities and commitments are conventional norms that can only exist within a social context. Indeed if the self is characterized essentially by responsibility, as I suggest, then the self is essentially normative and social. Just as my amoeba's food detector, as well as the food it detects, make sense only within the context of an ecological system proper to its species, so selfhood only makes sense within a certain kind of cultural environment, one which attributes responsibility to enduring selves.

And so we arrive finally at who we are today.

V Conclusion

My brief history of these superventions is ludicrously oversimplified. There are not four stages in the evolution of mind; over 4 billion years there have been an indefinitely large number of stages, without defined boundaries. Nor does every stage involve the decoupling of direct processes: evolution is opportunistic and proceeds by whatever devices it can discover. Every stage, however, is "naturalistic" in that it builds upon the earlier stages in a transparent manner that requires no appeal to mysterious, miraculous or other outside interventions. Just as the category-*Umwelt* structure of the amoeba supervenes on the direct causality of the physical world, so my higher-level stages of the mind/self presuppose the continued operation, albeit decoupled, of the lower levels. If direct causality were abolished, the rest of the structure would collapse.

This position is neither monistic physicalism nor dualism. It is not even a quadralism, for my

four stages are arbitrary snapshots of the gradual evolution which in reality is a continuum. There are many levels, an indefinite number of levels, of reality, not just one or two. That is why it makes little sense to ask, in this story, where the mind essentially begins. Only a dualist would ask for a precise boundary between the purely physical and the mind, between the pre-mental and the mental, or between the mind and the self. If one meant by mind that which does not operate on a purely direct, causal, physical basis, then even an amoeba has mental processes. On the other hand if one were to insist that the mind is essentially conscious of itself as being-towards-death, then only my stage four would qualify. But these questions are all empty ones. Once we understand the process of gradual evolutionary change through many stages over the course of evolutionary history, the decision about which stage should be designated the birth of mind is an arbitrary decision that, while useful for labelling and disciplinary communication, provides no further philosophical enlightenment.

Understanding who we are is an essentially historical enterprise. I don't just mean that each individual is defined by their own history, although that is true; I mean that the essential nature of selfhood can be understood only by grasping how we became what we are. We must understand ourselves historically because, like every other stage of the development of mind, the very essence of the self is to be a crystallized history.

David L. Thompson
Philosophy, Memorial University
2006-01-15