Philos. Technol. (2014) 27:31–46
DOI 10.1007/s13347-013-0132-9

SPECIAL ISSUE

Continuities and Discontinuities Between Humans, Intelligent Machines, and Other Entities

Johnny Hartz Søraker

Received: 10 January 2013 / Accepted: 10 September 2013 / Published online: 20 September 2013

Abstract When it comes to the question of what kind of moral claim an intelligent or autonomous machine might have, one way to answer this is by way of comparison with humans: Is there a fundamental difference between humans and other entities? If so, on what basis, and what are the implications for science and ethics? This question is inherently imprecise, however, because it presupposes that we can readily determine what it means for two types of entities to be *sufficiently* different—what I will refer to as being "discontinuous". In this paper, I will sketch a formal characterization of what it means for types of entities to be unique with regard to each other. This expands upon Bruce Mazlish's initial formulation of what he terms a continuity between humans and machines, Alan Turing's epistemological approach to the question of machine intelligence, and Sigmund Freud's notion of scientific revolutions dealing blows to the self-esteem of mankind. I will discuss on what basis we should regard entities as (dis-)continuous, the corresponding moral and scientific implications, as well as an important difference between what I term downgrading and upgrading continuities—a dramatic difference in *how* two previously discontinuous types of entities might *become* continuous. All of this will be phrased in terms of which scientific levels of explanation we need to presuppose, in principle or in practice, when we seek to explain a given type of entity. The ultimate purpose is to provide a framework that defines which questions we need to ask if we argue that two types of entities ought (not) to be explained (hence treated) in the same manner, as well as what it takes to reconsider scientific and ethical hierarchies imposed on the natural and artificial world.

Keywords Continuity · Moral status · Levels of explanation · Artificial intelligence · Turing · Scientific revolutions

1 Introduction

The uniqueness and corresponding moral status of humans compared to other types of entities has been a central philosophical topic since ancient Greece, and our implicit or

J. H. Søraker (✉)
Department of Philosophy, University of Twente, Postbox 217, 7500 AE Enschede, Netherlands
e-mail: j.h.soraker@utwente.nl

explicit views on the subject will strongly determine our stance in several of the most heated contemporary debates in applied ethics, including animal experimentation, abortion, autonomous agents, as well as several topics within environmental ethics. In many of these debates, the question of to what degree (adult) humans are unique and enjoy a privileged moral status rests on the idea that humans possess certain abilities and properties that other entities do not—that humans have a unique ontological mode of existence. Although I have argued along similar lines before (Søraker 2007), the purpose of this paper is to propose an alternative approach to this issue—one that is grounded in epistemology rather than ontology. This alternative approach does not ask which properties humans (typically) possess and whether these are unique or not, but rather asks what kinds of assumptions we presuppose when we explain humans, and whether these assumptions are unique to humans or not. In the terminology to be developed below: Is there a *discontinuity* between humans, on the one hand, and other types of entities, on the other? It is important to emphasize at the outset that I do not necessarily claim that this is a *better* approach than the more traditional, ontological ones, but it is a *different* approach that asks different questions and gives different insights into the relation between humans and other entities. In the remainder of this paper, I will try to show that it is important for several related reasons.

First, an epistemological notion of continuity allows us to specify new kinds of arguments to the effect that two types of entities should be treated and/or explained in the same manner—arguments that are grounded in scientific practices. The underlying idea is that when you regard humans as discontinuous from, for instance, intelligent machines, you ipso facto hold that the scientific theories required for explaining such machines could not fully explain a human being (either in principle or in practice, as I will discuss at length below). Second, it provides us with a way to conceptualize how new scientific explanations and discoveries may affect the way in which we treat and explain different types of entities, e.g., to what degree differential treatment of humans and other animals is justified, to what degree we can infer from animal experiments to humans, and to what degree we can make claims like "humans are nothing but machines." Third, the notion of discontinuity I will develop below raises several questions and will be in need of further clarification, but rather than being insurmountable problems with the theory, these are precisely the kinds of questions and clarifications that are needed in order to get a better picture of how and why humans, animals, machines, and other types of entities differ from each other. Fourth, an epistemological concept of continuity—since it is grounded in our understanding rather than our properties—allows us to more clearly understand how science sometimes conflict with common sense, or give rise to what Freud described as severe blows to the "narcissism of men … from the researches of science" (Freud 1953, p. 139). Fifth, although this approach has its own problems, which I will try to identify throughout, it avoids some of the problems that plague more ontological approaches, in particular the grounding of human uniqueness in properties that are inherently difficult to define precisely and may even be said to be "essentially contested" (Gallie 1955)—including such concepts as autonomy (Kant 1780/1997, 1788/1997), sentience (Singer 1990), having a conception of one's own life (Regan 2004), concept formation (Adler 1993), and having a will to live (Wetlesen 1999).[1]

[1] Cf. Søraker (2007) for an overview of the discussion of moral status from an ontological, property-based perspective.

Finally, the proposed theory provides a new way to specify one's stance regarding the relation between humans and other entities, a stance that will be partly determined by our often implicit assumptions regarding scientific realism and scientific reductionism. These are elements that are typically not included in more traditional approaches, which tend to neglect how important these assumptions are when we carve the world into categories and hierarchies. All of this will be elaborated below, but I will first outline the two approaches that originally inspired this paper and forms an important background: Bruce Mazlish's account of what he terms the "continuity" between humans and other entities as well as Alan Turing's epistemological approach (or so I will argue) to the question of machine intelligence.

2 Mazlish's Epistemological Notion of "Continuity" and Its Relation to the Turing Test

In his *The Fourth Discontinuity*, MIT historian Bruce Mazlish argues that progress in science and technology will result in a fourth *continuity* between man and machine (Mazlish 1993). Mazlish bases his work on Freud's discussion of scientific progress resulting in a decreasingly privileged role for humans (Freud 1953, p. 139), but Mazlish prefers to refer to scientific progress as the establishment of *continuities* instead of existentialist blows to our self-esteem. I will return to the Freudian aspect below, but according to Mazlish, there have been three dramatic scientific revolutions in the history of mankind, and these revolutions are best described as the establishment of *continuities*. Mankind has come to acknowledge that there is neither a sharp discontinuity between our planet and the rest of the universe (Copernican revolution), between humans and animals (Darwinian revolution), nor between rational and irrational humans (Freudian revolution). Mazlish argues that we should also overcome what he terms the *fourth* discontinuity that there is no sharp discontinuity between humans and machines. Mazlish's notion of continuity is helpful in many regards, primarily because it is grounded in scientific practice and, as I will explain shortly, makes the entities' uniqueness with regard to each other a question of epistemology rather than ontology. My goal in this paper is to make this notion more precise and useful, modify it so as to avoid some important problems, and extend it so as to include Freud's more existentialist notion of science undermining the uniqueness of humans. Before getting to all of this, let us start by looking closer at the advantages and problems with Mazlish's notion of continuity.

Mazlish is never explicit about his criteria for what it means to be continuous, but he seems to hold the view that a continuity is determined by whether or not the inner workings of two entities can be explained within the same scientific framework—that a discontinuity between two types of entities is bridged when "the same scientific concepts help explain the workings [of both]" (Mazlish 1993, p. 6). Using the continuity between humans and intelligent machines as example, Mazlish argues that the two will not need distinct scientific concepts, since they will both be explainable through some sufficiently advanced form of computationalism. His claim may seem immediately false, given the vast amount of criticism leveled against computationalism as a complete theory of mind by the likes of Searle (1984) and Putnam (1991), but Mazlish's point holds true if it is likely that we will (in the near future) develop a theory that can explain the functioning of both the brain and a computer. To use a different example, if Penrose (1999) is correct about the necessity of quantum phenomena for consciousness to occur, and such quantum phenomena could be used for both explaining and replicating the human mind, we would be continuous only with computers capable of replicating these phenomena. I will, henceforth, use “computationalism” in this weak sense of explaining human behavior as sets of operations that can, in principle, be replicated by a computer, whether this is Turing computable or hyper-computational (cf. Bringsjord and Arkoudas 2006; Copeland 2002).

There are clearly numerous problems with Mazlish’s approach, which I will return to below, but one of its advantages is that it takes an epistemological rather than ontological approach to the question of human uniqueness—an approach that in many ways mirrors Alan Turing’s well-known approach to the question of whether machines can be intelligent. In short, Turing argues that a computer is to be regarded as intelligent if a human judge cannot reliably distinguish the computer from the human in an imitation game (Turing 1950). What is important here is that Turing turns the question of intelligence from an ontological to an epistemological one. That is, Turing does not ask which properties a computer must possess in order to be deemed intelligent (its ontological mode of existence), but rather how an intelligent observer judges its behavior. The latter is a type of epistemological question, where we are really asking what kind of explanatory framework we need to presuppose in order to understand a particular type of behavior. If a computer were to pass the Turing test, this means that the judge had to *explain* its behavior as coming from an intelligent being, which says nothing about which properties that being must have (other than being able to display the behavior in question). Notice that this approach is radically different from the typical approach to questions of moral status and the like, where we typically discuss which properties an entity must possess in order to be regarded as a moral person.

In a similar manner, Mazlish argues (indirectly) that two types of entities should be regarded as continuous if they share scientific explanations; if the same framework of scientific concepts and models can fully explain the phenomenon under study. On this background, the Copernican revolution was really a realization that we do not need different scientific frameworks for the earth and the heavens (as was the case with the Aristotelian separation between the sublunar and supralunar realms), the Darwinian revolution was a realization that we do not need different scientific frameworks for humans and other animals, and the Freudian revolution was a realization that we do not need different scientific frameworks for the mentally ill and the mentally healthy. Mazlish’s prophesized fourth continuity, then, is the realization that we do not need different scientific frameworks for computers and humans either. Thus, all of these continuities amount to radical changes in how to *explain* different types of entities (epistemological), rather than claims about entities having shared properties or modes of existence (ontological). Turing convincingly argued that an epistemological approach to the question of machine intelligence was more fruitful than an ontological one, and I will take a similar approach to the question of continuity in the remainder of this paper. In doing so, I first need to make some important changes to Mazlish’s approach, which despite its advantages gives rise to some fundamental problems.

3 Problems with a Single-Level Approach to “Continuity”

As mentioned, Mazlish seems to claim that a continuity is determined by whether or not the inner workings of two entities can be explained within the same scientific framework, for instance the same physics being able to explain both the earth and the heavens, behaviorism being able to explain both humans and other animals, psychoanalysis being able to explain both mental health and illness, and some type of computationalism being able to explain both computers and the human brain.

This approach, which I will term the single-level approach, is problematic for two related reasons. First, anything can be explained within the same scientific framework. Disregarding supernatural and substance dualist accounts, it is probably possible *in principle* to explain the workings of the human brain and a computer by physics alone—and if we believe in scientific progress, our ability to do so will increase in time with progress in physics (I will return to this below).

Second, anything can be explained *as if* it belongs to a particular framework. For instance, as argued by Daniel Dennett, anything can be explained as if it is an intentional agent—what he refers to as taking an *intentional stance* (Dennett 1989). Since Mazlish does not specify how strict we need to be when claiming that the same scientific framework *can* explain two types of entities, his approach becomes inherently imprecise. If it is sufficient that it is *in principle* possible to explain something within the same framework, then every existing entity is continuous as long as there are no phenomena that cannot in principle be explained by some kind of physics. This would entail that humans are continuous with light bulbs, supernovae, and clouds, which leaves the concept of little use. It seems more reasonable, then, to refer to some kind of pragmatism where it must not only be in-principle possible but also pragmatically feasible to explain two entities within the same framework. But, this would require some kind of measure for what it means to be pragmatically feasible. Is it, for instance, pragmatically feasible to explain the brain fully in terms of physical processes, or do we (also) need to invoke chemistry, biology, psychology, and sociology?

Some of these problems are difficult to escape, since it is hard to provide objective criteria for when a particular scientific framework ceases to be feasible. However, we can try to remedy the problem of explanations at different levels by explicitly invoking this into the conception of continuity. In the next section, I will sketch such a multilevel account of continuity. Before outlining this multilevel account, allow me to emphasize that my main concern in this paper is to discuss the formal nature of these continuities, so it is important to note that the levels of explanation that I will use as examples below are to be seen as mere placeholders—the reader will inevitably find some of them problematic and/or imprecise. My goal is to first work out the formal schematics, and then return to a more substantial formulation in later work. Such a formulation will, among other things, require a defense of a particular type of both scientific and moral realism, along with a robust notion of scientific practice and development—both of which fall well beyond the scope of this paper.

4 A Multilevel Approach to "Continuity"

Rather than asking whether two types of entities can be explained within the same scientific framework, I believe it is better to approach this in terms of *sets* of scientific frameworks—or what I will refer to as sets of scientific levels of explanation. That is, rather than asking whether two entities can be explained within the same scientific framework, we should ask whether two entities require the same set of scientific levels of explanation. As Nagel (1979) rightly argues, there are (at least) four fundamentally different types of explanation—deductive, probabilistic, teleological, and genetic—which makes it difficult to precisely define what a scientific level of explanation (LoE) is. For present purposes, I will simply use the term in the more generic sense of a more or less coherent and mutually supportive set of principles, concepts, and models that attempt to provide an account of the relationships between cause and effect.[2]

As mentioned, one of the problems with Mazlish's single-level approach, where continuity is established on the basis of sharing *one* scientific framework, is that we often choose different levels of explanation (or, to use Luciano Floridi's term, levels of *abstraction* (Floridi 2008)) depending on what it is that we seek to explain. Even if it is in-principle possible to explain human behavior by physics alone, we typically employ higher-level explanations instead. For instance, at a behavioral level of explanation, we employ concepts like stimulus and response to explain behavior, without involving physics or chemistry. Even for entirely physicalist phenomena, such as an object moving through space, we often employ heuristics instead of explaining what is "really" going on, such as the complex interplay between electrons and various force fields. This, along with the other problems with a single-level approach mentioned above, entails that we cannot define a continuity in terms of *one* shared scientific framework. A more promising approach is to define continuity in terms of having a particular *set* of scientific levels of explanation in common.

To simplify things, if we take a single-celled organism, we may be able to explain its functioning at a physical level of explanation alone.[3] When we get to more complex forms of life, however, such explanations quickly become untenable. At some point, the chemistry involved becomes too complex to be described in physics terms alone. At even higher levels of complexity, chemistry also fails to provide a full explanation and we need to start talking about the *behavior* of a system, and employ principles from, say, behaviorism and comparative psychology instead of the actual physical and chemical processes. At even higher levels, we may need to involve the environment, higher-order cognitive processes, autonomy, phenomenal experience, and some notion of metacognition (Flavell 1979). At even higher supra-individual levels, we may also require social and cultural levels of explanation. Which levels we may need in order to fully explain a given entity is clearly controversial, and not my concern in this paper, but only the most radical and optimistic scientists maintain that we will in the foreseeable future be able to pragmatically explain everything by means of one unified theory. On the basis of all this, if we are to define continuity in terms of which types of explanation are required, we must talk about *sets of levels of explanation* (multilevel) instead of Mazlish's single-level scientific frameworks.

On this background, we can stipulate the following preliminary hypothesis: two types of entities are continuous if and only if a full understanding of their nature and behavior requires the same set of scientific levels of explanation; two types of entities are discontinuous if and only if a full understanding of their nature and behavior does *not* require the

[2] I am very aware that this is far from precise, but it should be sufficient for establishing the more formal nature of continuities, which is the limited purpose of this paper. As Mieke Boon has pointed out to me, it is probably better to speak of actual scientific "practices of explanation", but that will have to be reserved for future work.

[3] For such an attempt, see Princeton University's Laboratory for the Physics of Life (http://tglab.princeton.edu/).

same set of scientific levels of explanation.[4] This means that humans and other intelligent machines are continuous if and only if a full understanding of their nature and behavior requires the same set of scientific levels of explanation. These definitions still lack precision, however, and we need to first clarify what is meant by a particular level being "required".

5 Epistemic Versus Ontic Continuities

There are two radically different ways in which a LoE may be required for explaining a type of entity. On the one hand, we could for instance argue that the human brain works in such a way that we cannot fully understand its functioning without employing a chemical level of explanation. Perhaps the chemical properties of neurotransmitters and hormones function in a way that cannot possibly be accounted for by means of more mechanistic explanations. This would be an antireductionist view of chemistry. If such a chemical LoE is required *in-principle* because of the brain's unique mode of existence, then that LoE is required for what I refer to as *ontic* reasons. This does not mean that we are back to an ontological, essentialist approach, but this means that certain entities cannot be fully understood unless we presuppose a particular LoE, and that we infer that the reason for this is ultimately the entity's mode of existence.

Using the same example, we could also argue that the human brain works in such a way that it is much more pragmatic or "tractable" to use a chemical LoE, even if such an explanation can *in principle* be reduced to a more fundamental LoE. If we, despite this in-principle possibility, do require a chemical LoE in order to actually understand something, then that LoE is required for *epistemic* reasons. Thus, the distinction between ontic and epistemic continuity maps on to the distinction between realist and epistemic interpretations of explanation. The former "holds that the entities or processes an explanation posits actually exist—the explanation is a literal description of external reality. An epistemic interpretation, on the contrary, holds that such entities or processes do not necessarily exist in any literal sense but are simply useful for organizing human experience" (Mayes 2005). As specified by Atmanspacher (2002), the epistemic perspective relates to our *knowledge* of the states and observables of a system, whereas the ontic perspective relates to states and observables independent from such knowledge. In other words, the epistemic perspective takes into account our epistemological presuppositions and frameworks, whereas ontic perspectives attempt to specify intrinsic properties that are knowledge-independent (or transcendental). In light of the above, we can make the following distinction between an ontic and epistemic continuity as follows:

> Ontic continuity: two types of entities are ontically continuous if and only if a full understanding of their nature and behavior require the same set of scientific levels of explanation *in principle*, due to their mode of existence.

[4] One of the biggest and perhaps unavoidable problems with the proposed approach lies in defining what it means to have a "full" account of an entity, i.e., whether a set of levels of explanation is sufficient for explaining the structure and behavior of a system. I cannot possibly do this question justice in this paper, but it is perhaps best to define a "full account" negatively—that is, if you hold the view that humans cannot be explained without presupposing some form of phenomenal, indeterministic consciousness (what I will later refer to as a "cognitive" LoE), then you hold the view that a set of LoEs that does not include such a level, would not be able to provide a full account.

> Epistemic continuity: two types of entities are epistemically continuous if and only if a full understanding of their nature and behavior requires the same set of scientific levels of explanation *in practice*.

Humans and other animals are *ontically* continuous if and only if a full understanding of their nature and behavior requires the same set of scientific levels of explanation *in principle*. Humans and other animals are *epistemically* continuous if and only if a full understanding of their nature and behavior require the same set of scientific levels of explanation *in practice*. I will refer to the former as a set of LoAs that are "ontically necessary" and the latter as "epistemically necessary." The latter is "relative to what we, or some given set of people, know, or to what we believe … In ancient times, it was … epistemically necessary that the earth was flat. Its flatness followed from other beliefs then current" (Proudfoot and Lacey 2009, p. 260). The purpose of the term "epistemic continuity," and this approach in general, is in other words to explicitly invoke scientific practices, whether these are current or historical.

It is far from uncontroversial which LoEs are ontically or epistemically necessary for a full understanding (as well as what is to be meant by "full"), and it is far beyond the scope of this paper to discuss this for different types of entities. However, these controversies are not unique to this framework; the same problems occur in most reductionism debates—and one's stance with regard to the latter can be implemented in the "continuity" framework. For instance, in philosophy of mind, a property dualist would hold that consciousness is somehow irreducible to neurobiology and physics—which means that a "higher" LoE is ontically necessary for a full understanding of a conscious being. Eliminative materialism, on the other hand, holds that consciousness can and should be explained at a neuroscientific LoE, thus claiming that "higher" LoEs (folk psychological concepts, in particular) are neither ontically nor (at least in the future) epistemically necessary for a full understanding of conscious beings. If we compare humans and other animals, substance dualists would hold that conscious animals are ontically discontinuous from non-conscious animals, whereas eliminativists would hold that conscious animals are ontically continuous with non-conscious animals. Non-reductive physicalism, however, holds that conscious states really are the same as physical states and that the former can *in principle* be explained by the latter—but not *in practice*. In terms of "continuity," non-reductive physicalism would entail that entities with phenomenal experience would be *epistemically* discontinuous from entities without. Such a view is perfectly captured by Frank Jackson's famous thought experiment in "What Mary didn't know" (1986), which in essence argues that neuroscientific explanations are not sufficient for explaining the phenomenal experience of the color red. In this manner, the framework proposed in this paper can also be used to express one's stance on various philosophical questions, for instance by adopting a stance of "ontological discontinuity with regard to humans and intelligent machines" (meaning, in essence, the stance that machines will never be understood in the same way as humans, in principle and regardless of new scientific discoveries,). As a final example, the fact that there is a phylogenic continuity between humans and other animals strongly suggests that there is *at least* an ontic continuity as well. The only ways to establish an ontic discontinuity between humans and other animals is to either argue that natural selection has incrementally brought about a qualitative shift between humans and whichever animal we compare ourselves with, or to reject phylogenic continuity by resorting to some notion of intelligent design. On the latter view, human beings would be ontically discontinuous with other animals because they were designed differently—with the effect that they require different LoEs. In a manner of speaking, those who hold that humans are unique because we are created by an intelligent designer implicitly presuppose a "supernatural" level of explanation required to fully understand humans and only humans.

6 The Schematics of Continuities

I hope the considerations above clarify roughly what I mean by levels of explanation, and the difference between such levels being ontically (in-principle) necessary in contrast with epistemically (in-practice) necessary. I will return to more concrete examples of LoEs below, but in order to make the following schematization as simple as possible, I will temporarily employ four more abstract "placeholders" for the actual levels of explanation. Let us call these levels the "physical," "functional," "behavioral," and "cognitive" levels of explanation.

Let us assume, for the sake of the argument, that various aspects of what it means to be human *can* be explained at a physical, functional, behavioral, and cognitive level of explanation. The physical level includes any explanation in terms of the physical structure and processes of a system, and basically applies to anything that can be explained in terms of the language of physics, such as matter, mass, energy, magnitude, forces etc. For present purposes, I include chemical and biological explanations at this level, although there will in some instances be necessary to separate this level further, for instance, if discussing a discontinuity between living and nonliving things—where the latter would, for instance, not require a biological notion of autopoiesis (Maturana and Varela 1980). Add to this what I will refer to as a functional level, which includes explanations in terms of processes that take place between a physical structure's input and output. This includes brain processes, but also applies to any other system that carries out a function that cannot (again, in principle or in practice) be explained in terms of physical structure alone. We can also add a behavioral level, which includes any explanation in terms of the behavior of the system. The theory that comes closest to exhausting this level is "behaviorism," which explains a system entirely in terms of its observable behavior and includes anything that can be explained in terms of interactions between an individual entity and its environment—particularly those related to learning, reinforcement, and forming associations. For the same reasons that behaviorism came under fire in the 1960s, there is good reason to believe that, for humans, we need to add a level of explanation on top of the behavioral. If humans cannot be fully understood entirely in terms of behaviorist language, in particular due to exclusion of mental states, we need to add what I for present purposes refer to as a "cognitive" LoE.[5] Such a level would include any explanation in terms of the mental states of the system, in particular such concepts as attention, language use, complex problem solving, phenomenal experience, emotions, metacognition, empathy, and so forth. It is difficult to specify in precise terms how the cognitive differs from the behavioral, but a rule of thumb can be found in the

[5] The term "cognitive" may be somewhat misleading, since cognitive is often regarded as including computational and other low-level, non-conscious information processing (see, e.g., Marr (1982)). I am using it here in the sense of mental processes related to thinking and conscious experience. It may have been better to use "phenomenal LoE," but this comes with its own misleading connotations.

traditional distinction between *verstehen* (interpretive understanding) and *erklären* (causal explanation) (Dilthey 1989; Feest 2010), or the more analytical counterpart in the distinction between causes and reasons (Davidson 1963; Dretske 1991). Causal explanation "tries to make explanatory sense of the phenomenon by finding the laws that govern it, whereas (interpretive understanding) tries to make empathetic sense of the phenomenon by looking for *the perspective from which the phenomenon appears to be meaningful and appropriate*" (Bransen 2001, p. 16,165, my emphasis). It is important to emphasize that I do not claim that these four levels are the only levels of explanation we need, but I will use them for illustrative purposes below. Indeed, one of the features of the framework presented in this paper is that it *asks* which levels are required, an analysis that will necessarily depend on the type of continuity under investigation.

These levels are demarcated according to their explanatory "range." Some systems can be fully explained in terms of physical structures alone, some systems also require explanations in terms of the functions performed by the physical system, some systems also require explanations in terms of observable behavior, and, finally, some systems also require explanations in terms of subjective experiences and higher-order cognitive operations. If we, at least for the sake of illustration, agree that all these levels of explanations are required, if we are to fully understand humans, then these levels form a *set* of necessary LoEs for humans. The thesis to be defended below basically states that any entity that requires the same set will be continuous with humans, and any entity that requires a *different* set will be *dis*continuous with humans. This allows us to schematize what continuities might look like, according to this multilevel approach. Consider the following *hypothetical* necessity of LoEs for humans, other animals, intelligent machines, and inanimate objects (Table 1).

In this *hypothetical* example, humans require a physical, functional, behavioral, and a cognitive LoE for a full understanding, whereas other animals can be fully understood by physical, functional, and behavioral LoEs alone. If this is the case, then humans would be discontinuous with other animals. If the cognitive LoE is required *in principle*, this is an ontic discontinuity, if required only *in practice*, this is an epistemic discontinuity.

Now we are able to describe the seeming inconsistency of animal experimentation presupposing both a radical similarity (scientific validity) and radical difference (ethical justifiability) between humans and other animals. The scientific validity of such experiments can be grounded in the fact that the LoEs that are relevant for scientific validity (primarily the functional and behavioral) are shared as long as the experiments are not intended to explain an aspect of humans that can only be explained at a LoE that is unnecessary for the animals. The LoEs that are relevant for the *ethical* justifiability (cognitive) are not shared, however. That is, the non-shared cognitive LoE corresponds directly to a self-reflective ability that allows humans to be harmed in ways that other animals cannot.[6]

This further illustrates how one purpose of establishing discontinuities in this manner is to map their required LoE onto a classification of moral status. That is, there are different ways to harm entities corresponding to their required LoE. In a manner of speaking, the

[6] This reasoning accords with the principle of formal equality, one of the most fundamental and undisputed principles in ethics, which states that a difference in treatment or value between two kinds of entities can only be justified on the basis of a relevant and significant difference between the two (cf. Søraker (2007).

Table 1 Hypothetical sets of LoE for humans, other animals, intelligent machines and inanimate objects

Type of entity Required LoE	Humans	Other animals	Intelligent machines	Inanimate objects
Cognitive	X			
Behavioral	X	X		
Functional	X	X	X	
Physical	X	X	X	X

more LoEs that are required for understanding an entity, the more ways there are to harm that entity. Although we may speak of minimal harms at the physical and functional levels, such as Floridi's notion of informational entropy (cf. Floridi 2002), more conventional types of harm only occur at the behavioral level, in terms of rewards and punishment. At a cognitive level, the harms become much more complex, including harms related to offense, dignity, privacy, self-actualization, and so forth.

To more clearly show the difference with Mazlish's single-level approach, consider the following table in which we compare humans with simple, non-autonomous machines (I will leave out the "physical" level from now on, assuming that a physical LoE is necessary for a full understanding of any physical entity) (Table 2).

In this hypothetical example, humans are discontinuous with machines even if both were to require some form of functional explanation because humans require a behavioral and cognitive LoE, whereas machines can in general be fully understood without (I will return to intelligent machines below). Again, this would be an ontic discontinuity if the functional LoE is in-principle insufficient for a full understanding of humans. It would be an epistemic discontinuity if the cognitive and behavioral LoE are only required for pragmatic reasons. Note that Mazlish's single-level approach is unable to account for this, and would be committed to treating humans and machines as continuous as long as the functional LoE somehow applies to both types of entities.

7 Downgrading Versus Upgrading Continuities

Another important purpose of the schematization above is to re-conceptualize Freud's original insights about the more existentialist impact of scientific progress, an aspect that originally inspired Mazlish yet finds no place in his approach. In Freud's words, "the universal narcissism of men, their self-love, has up to the present suffered three severe blows from the researches of science" (Freud 1953, p. 139). I will refer to such blows as *downgrading* as opposed to upgrading continuities, which also further illustrates what is meant by LoEs being required in principle and in practice.

The notion of some LoEs being only epistemically necessary implies that scientific progress brings about changes in which levels that are necessary to explain a given entity—which is reflected in the scientific ideals of parsimony, unification, and reduction (Cat 2010). This means that two types of entities previously seen as discontinuous may *become* continuous—or vice versa. That is, two types of entities that previously required different sets of LoE may come to require the same set of LoEs in light of new

Table 2 Hypothetical sets of LoE for humans and machines

Type of entity Required LoE	Humans	Machines
Cognitive	X	
Behavioral	X	
Functional	X	X

scientific discoveries. In terms of these schematics, this can come about in two different ways—which correspond to two radically different ways in which science may change our worldview, and where we can more precisely conceptualize Freud's notion of blows to the self-esteem of mankind.

First, we may come to realize that a type of entity no longer requires a LoE that we previously thought to be necessary. This can happen through successfully reducing one LoE to a more fundamental one, or through the elimination of a LoE found to be false—as was the case with "vitalism," long thought to be a necessary LoE for living organisms (Williams 2003). When two types of entities come to share the same set of LoE because one type *loses* a LoE, this amounts to a *downgrading* continuity. More schematically: Table 3.

This was precisely the concern when Skinner's radical behaviorism aspired to explain both humans and other animals entirely in terms of behaviorist principles (Skinner 1974). This would, according to this line of reasoning, entail a continuity between humans and other animals because humans would no longer require any additional LoEs. This would *downgrade* humans to the level of animals. The same might also hold regarding intelligent machines, where humans would become downgraded to the level of intelligent machines if they come to share the same set of LoEs due to humans not really requiring a cognitive LoE at the same time as intelligent machines come to require a behavioral LoE. Such a downgrading continuity amounts to the claim that "humans are *nothing but* machines."

There is a converse way of becoming continuous, however. Consider first the following continuity between nonhuman animals and intelligent machines (Table 4).

In this case, nonhuman animals and intelligent machines become continuous because the latter *gain* new LoEs. That is, intelligent machines might become so complex that we can no longer explain their functioning by means of functional (computational) principles alone. We may need to adopt behavioral principles to explain intelligent machines as well, not only metaphorically but as an in-practice (epistemic) or even in-principle (ontic) requirement for fully explaining intelligent machine behavior. This is what I refer to as an *upgrading* continuity, where two types of entities come to share the same set of LoEs due to one *gaining* a new LoE—through the realization that the previously sufficient set of LoE is no longer sufficient. We may argue that intelligent machines have been upgraded in this manner, since highly advanced neural networks now require at least *behavioral* notions of stimuli and reward in order to be explained. That is, if we want to explain exactly how a complex neural network functions, a purely computational account of the weights of the nodes etc. will be insufficient for explaining exactly why the network generates its output. Thus, the necessity of a behavioral framework for explaining highly complex computers, neural networks, and embedded systems in particular entails that we can already now speak of an (epistemic) continuity between machines and many low-

Table 3 Hypothetical loss of required LoE (downgrading continuity)

Type of entity Required LoE	Humans	Other animals
Cognitive	- X	
Behavioral	X	X

level animals, depending on whether the latter require other LoEs or not.[7] They will then have come to share the same set of LoEs due to intelligent machines now requiring a behavioral LoE of equal complexity to animals of low complexity and, therefore, ought to be *upgraded* to the level of correspondingly complex living organisms—both scientifically and ethically.

If we now turn to the traditional question of "man versus machine," consider the following scheme (Table 5).

I stipulated above that intelligent machines may require a behavioral LoE, but they would still be discontinuous with humans as long as they do not require a cognitive LoE. However, intelligent machines might become so complex that they can no longer be explained by means of functional and behavioral language alone. At some point, we may need to adopt cognitive principles to explain intelligent machines as well, not only metaphorically but as an in-practice (epistemic) or even in-principle (ontic) requirement for explaining intelligent machine behavior. This further illustrates what I refer to as an *upgrading* continuity, where two types of entities come to share the same set of LoEs due to one *gaining* a new LoE through the realization that the previously sufficient set of LoE is no longer sufficient—either because its complexity eventually gives rise to phenomena that cannot be explained at a behavioral LoE *in principle*, or simply that it becomes practically unfeasible to do so. Rather than the downgrading sentiment of humans being "nothing but machines," such an upgrading continuity would correspond more closely to machines being *as intelligent as* humans.

8 Concluding Remarks

Needless to say, this approach is fraught with problems. Although it might sound like an evasive maneuver, the problems with the approach also indicate its advantage. I am certain that most readers have disagreed with the levels of explanation used for illustration above, but such disagreement will and *should* determine whether you regard two types of entities as continuous or not. If you think that a behavioral LoE is sufficient for all animals, including humans, then you should also hold that there is no sharp discontinuity between humans and animals. If you, *purely* for the sake of illustration, think that some type of supernatural LoA is required for a full understanding of humans but no other animals, then

[7] This actually illustrates how these schematics can lead us to ask the right kinds of questions, in the following manner: If (most) animals only require a physical, functional, and behavioral LoE, they are at least *epistemically* continuous with intelligent machines that are no longer explainable at a physical and functional level alone. Then the question becomes whether a "biological" LoE is necessary to understanding some types of entities, ipso facto, making them (epistemically or ontically) discontinuous with nonbiological entities.

Table 4 Hypothetical gain of new LoE (upgrading continuity)

Type of entity Required LoE	Non-Human animals	Intelligent machines
Behavioral	X	+ X
Functional	X	X

you should also hold that there is a sharp discontinuity between the two. If you think, for instance, that I missed a biological LoE and that this makes a difference, then that will and should determine how you carve up the world. If you believe that biological processes cannot be reduced to physical processes, then this entails a discontinuity between those who do and those who do not require a biological LoE. In other words, many of the problems with the multilevel notion of continuity sketched above can be related directly to equivalent controversies in the reductionist debate, and your stance on the latter will determine which levels you find to be epistemically or ontically necessary for which types of entities.

Allow me to repeat that my only concern in this paper has been to sketch one possible epistemological approach to the question of human uniqueness, and a lot of this has to be informed by similar debates in philosophy of science, philosophy of mind, epistemology, and meta-ethics. Indeed, the account sketched above presupposes some notion of scientific realism "typified by an epistemically positive attitude towards the outputs of scientific investigation, regarding both observable and unobservable aspects of the world" (Chakravartty 2013), but it should also be compatible with more pragmaticist notions of science. The difference between the two will basically be reflected in when and why one finds particular LoEs to be required for ontic or epistemic reasons. Although not necessary, my approach also presupposes some idea of scientific progress—i.e., that science, through the development and elimination of LoEs, is providing us with an increasingly accurate picture of reality. That said, I certainly do not rule out the possibility of dramatic paradigm shifts, but this can be accounted for within this conception of continuity as well. Indeed, a continuity between intelligent machines and humans is likely to require a paradigm shift that obliterates our current LoEs—for instance, if we arrive at some future LoE that is required in order to explain consciousness *and* to build conscious machines.

There is no doubt that the details, if we can agree on the formal nature, will require a lot of clarification. In the meantime, my hope for this paper is that the reader will find the notion of continuity an intuitively helpful concept—along with the distinctions between epistemic versus ontic and downgrading versus upgrading continuities. Furthermore, I hope and believe that the following questions that the reader is left with are precisely the kinds of questions that can lead to a better understanding of the two types of entities ought to be regarded as continuous: which levels of explanation are required in order to fully

Table 5 Hypothetically gaining set of LoEs equivalent to humans

Type of entity Required LoE	Humans	Intelligent machines
Cognitive	X	+ X
Behavioral	X	X
Functional	X	X

explain an entity, what does it mean to "fully" explain something, which of these are merely epistemically required, which levels do we share with other types of entities, which levels should be eliminated or reduced to other levels, and what are the implications when the corresponding downgrade or upgrade leads to a shared set of levels. In conclusion, I can only hope that this paper was read in the spirit intended—as an initial, exploratory and formal account of what it means for two types of entities to be (dis-)continuous, as seen from an epistemological perspective.

Acknowledgements Since this is an idea that has resisted precision, hence publication, for more than 10 years, I can no longer thank everyone who has given me advice over the years. Most importantly among them, my then-supervisor Magne Dybvig played a very important role in the initial development. More recently, I am indebted to the helpful comments from several colleagues from my department, in particular Marianne Boenink, Mieke Boon, Philip Brey, Mark Coeckelbergh, and Pak Hang Wong. I am also indebted to the feedback and encouragement from the participants at the AISB/IACAP 2012 conference symposium on 'The Machine Question', in particular Joanna Bryson, David Gunkel, Steve Torrance and Wendell Wallach. I would also like to acknowledge the very useful and constructive feedback from the journal's anonymous referees – in particular "reviewer #1" who provided an extraordinarily detailed and insightful analysis that was of immense help. The usual disclaimer applies.

References

Adler, M. J. (1993). *The difference of man and the difference it makes*. New York: Fordham University Press.
Atmanspacher, H. (2002). Determinism is ontic, determinability is epistemic. In H. Atmanspacher & R. Bishop (Eds.), *Between chance and choice: interdisciplinary perspectives on determinism* (pp. 49–74). Charlottesville: Imprint Academic.
Bransen, J. (Ed.). (2001). *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Elsevier.
Bringsjord, S., & Arkoudas, K. (2006). On the provability, veracity, and AI-relevance of the Church Turing Thesis. In A. Olszewski, J. Woleński & R. Janusz (Eds.), *Church's Thesis After 70 Years* (pp. 66–118).
Cat, J. (2010). The Unity of Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2010 ed.).
Chakravartty, A. (2013). Scientific Realism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (*Summer 2013 Edition*).
Copeland, J. (2002). Hypercomputation. *Minds and Machines, 12*, 461–502.
Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy, 60*(23), 685–700.
Dennett, D. C. (1989). *The Intentional Stance*. Cambridge: MIT Press.
Dilthey, W. (1989). *Introduction to the human sciences*. Princeton: Princeton University Press.
Dretske, F. I. (1991). *Explaining behavior: Reasons in a world of causes*. Cambridge: MIT Press.
Feest, U. (2010). *Historical perspectives on Erklären and Verstehen*. Berlin: Springer.
Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34*(10), 906–911.
Floridi, L. (2002). On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology, 4*(4), 287–304.
Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines, 18*(3), 303–329.
Freud, S. (1953). A difficulty in the path of psycho-analysis (Eine Schwierigkeit der Psychoanalyze). In J. Strachey (Ed.), *The standard edition of the complete psychological works* (Vol. XVII, pp. 135–145). London: Hogarth.
Gallie, W. B. (1955). Essentially contested concepts. *Proceedings of the Aristotelian Society, 56*, 167–198.
Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy, 83*(5), 291–295.
Kant, I. (1780/1997). *Lectures on Ethics* (P. Heath, Trans.). Cambridge: Cambridge University Press.
Kant, I. (1788/1997). *Critique of Practical Reason* (M. Gregor, Trans.). Cambridge: Cambridge University Press.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman and Company.
Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition*. Dordrecht: D. Reidel Publishing Company.
Mayes, G. R. (2005). Theories of Explanation. *Internet Encyclopedia of Philosophy* Retrieved January 20, 2012, from http://www.iep.utm.edu/explanat/.
Mazlish, B. (1993). *The fourth discontinuity*. New Haven and London: Yale University Press.

Nagel, E. (1979). *The structure of science: problems in the logic of scientific explanation*. Indianapolis: Hackett Publishing.

Penrose, R. (1999). *The emperor's new mind: concerning computers, minds, and the laws of physics*. Oxford: Oxford University Press.

Proudfoot, M., & Lacey, A. R. (2009). *The Routledge dictionary of philosophy* (4th ed.). New York: Routledge.

Putnam, H. (1991). *Representation and reality*. Cambridge: MIT Press.

Regan, T. (2004). *The case for animal rights*. Berkeley: University of California Press.

Searle, J. (1984). *Minds, brains, and science*. Cambridge: Harvard University Press.

Singer, P. (1990). *Animal Liberation* (2nd ed.). London: Thorsons.

Skinner, B. F. (1974). *About behaviorism*. New York: Knopf.

Søraker, J. H. (2007). The moral status of information and information technologies – a relational theory of moral status. In S. Hongladarom & C. Ess (Eds.), *Information Technology Ethics: Cultural Perspectives* (pp. 1–19). Hershey, PA: Idea Group Publishing.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 236*, 433–460.

Wetlesen, J. (1999). The moral status of beings who are not persons: a casuistic argument. *Environmental Values, 8*, 287–323.

Williams, E. A. (2003). *A cultural history of medical vitalism in enlightenment Montpellier*. Farnham: Ashgate.