How Biology Shapes Philosophy

New Foundations for Naturalism

Edited by

DAVID LIVINGSTONE SMITH

University of New England, Biddeford, Maine

CAMBRIDGE
UNIVERSITY PRESS

1 *Darwin and the Overdue Demise of Essentialism*

DANIEL C. DENNETT

Now that Darwinian thinking has replaced essentialist thinking in biology, should "essence" be considered a dirty word, banished from the working vocabulary of philosophers in all but historical contexts? The term gets some protective coloration from its innocent, nontechnical use by nonphilosophers. For instance, Douglas Hofstadter, queried by me about its use in the title of his recently co-authored book, *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking* (with Emmanuel Sander, New York: Basic Books, 2013) responded

> I don't react negatively or fearfully when I hear the word "essence", because in my mind the word has not been tainted by a wearying set of arcane debates in philosophical circles. For me, it's just an informal, everyday word, not a technical term. I doubt that you and I have any disagreement about what the word "essence" means when it's used informally, as a synonym for "gist:", "crux", "core", etc. And that's how it's used in the book. (Personal correspondence, January 3, 2014)

He is right that I have no disagreement with him regarding his (familiar, nontechnical) sense of the term. I worry, however, that the excellent insights he reaps from his celebration of essences in *this* sense will lend false respectability to the philosophers' "arcane debates" when they use the term – and, more recently, when they use thinly veiled substitutes (euphemisms, in effect), now that its credentials have been challenged.

Ever since Socrates pioneered the demand to know what all *F*s have in common, in virtue of which they are *F*s, the ideal of clear, sharp boundaries has been one of the founding principles of philosophy. Plato's *forms* begat Aristotle's *essences*, which begat a host of ways of asking for *necessary and sufficient conditions*, which begat *natural kinds*, which begat *difference-makers* and other ways of tidying up the borders of all the sets of things in the world. When Darwin came

Thanks to Diana Raffman for illuminating discussion and advice on an earlier draft.

along with the revolutionary discovery that the sets of *living* things were not eternal, hard-edged, in-or-out classes but historical populations with fuzzy boundaries, islands historically connected to other islands by vanishing isthmuses, the main reactions of philosophers were to either ignore this hard-to-deny fact or treat it as a challenge: Now how should we impose our cookie-cutter set theory on this vague and meandering portion of reality?

"Define your terms!" is a frequent preamble to discussions in philosophy, and in some quarters it counts as Step One in all serious investigations. It is not hard to see why. The techniques of argumentation inaugurated by Socrates and Plato and first systematized by Aristotle are not just intuitively satisfying ("self-evident" on reflection) but demonstrably powerful tools of discovery, indispensable for answering difficult questions and resolving contentious disagreements, often with an undeniable finality. Shouldn't the goal of all inquiry be the triumphant coda "*Quad erat demonstrandum*, which was to be demonstrated"? Euclid's plane geometry was the first parade case, with its crisp isolation of definitions and axioms, inference rules, and theorems. If only all topics could be tamed as thoroughly as Euclid had tamed geometry! The hope of distilling everything down to the purity of Euclid has motivated many philosophical enterprises over the years, different attempts to *euclidify* all the topics and thereby impose classical logic on the world. These attempts continue to this day and have often proceeded as if Darwin never existed. Philip Kitcher points to a glaring example:

> Consider, for example, what Ernst Mayr has called Darwin's replacement of "typological thinking" by "population thinking." Darwin's recognition of a vast amount of intraspecific variation often goes unappreciated today in philosophical discussions, even though it has been uncontroversial for well over a century. Recent discussions of natural kinds, prompted by the seminal ideas of Saul Kripke and Hilary Putnam, often assume that one can revive essentialism. Yet, if species are natural kinds, no such revival is in prospect. Kripke and Putnam largely restricted their discussions to the cases of elements and compounds, and with good reason, for given the insights of neo-Darwinism, it is clear that the search for some analog of the microstructural essences can't be found. No genetic or karyotypic property will play for species the role that atomic number does for the elements. (Kitcher 2009)

It was Quine (1969) who reintroduced the term "natural kinds" to philosophy, and he, at least, appreciated that only a few of the kinds found in nature are natural kinds considered as modern-day essences, and he probably regretted the way his imprimatur was interpreted as a naturalist's blessing for some kind of return to carefree essentialism. "Green things, or at least green emeralds, are a kind," Quine observed (p. 116), manifesting his own appreciation of the fact that while emeralds may be a natural kind, green things are not. Colors are not natural kinds precisely because they are a product of biological evolution, which has a tolerance for sloppy boundaries when making categories that would horrify any philosopher bent on achieving good, clean definitions. If some creature's life depended on lumping together the moon, blue cheese, and bicycles, you can be pretty sure that Mother Nature would find a way for it to "see" these as "intuitively just the same kind of thing."[1]

The common unspoken presumption that somehow essentialism can be made to work outside the abstract realm of mathematics is, I suggest, a methodological, not metaphysical, prejudice. Spelled out, the presumption is this: we have this wonderful tool, classical bivalent logic, and we've invested our lives in mastering its use. Without sharp boundaries, as sharp as those of Euclid's classes of geometric elements, it is disabled, so we will take on as a working assumption that there is some way of euclidifying all the vagueness and fuzziness out of our terms. In this way, we can go back to business as usual, tolerating Darwinian population thinking among those with a taste for such practices but denying its application to our chosen topics.

The prejudice is widespread among philosophers, and not without reason. For instance, one of the most popular practices threatened by the Darwinian perspective is the tactic of confronting opponents with disjunction-elimination arguments of what might be called the fish-or-cut-bait variety. First, let me demonstrate why it is so often favored by philosophers and then show why they are wise to foreswear it in many – almost all – naturalistic contexts.

[1] This paragraph is adapted from Dennett, *Consciousness Explained* (1991, p. 381, n. 2). The heterogeneous set could mark an idiosyncratic kind for any organism that had three detector/sensor systems yoked with OR gates that were impenetrable to introspective analysis, a case of radical synesthesia. It could be lumpy with crisp edges or lumpy and fuzzy at the same time. What counts as a bicycle or blue cheese is negotiable in our world.

Consider the following simple proof that there are irrational numbers *A* and *B* such that *A* to the *B* power is rational. It depends on the assumption that every real number is either rational or it isn't.

Let *A* be $\sqrt{2}$.
Let *B* be $\sqrt{2}$.
Then what about *A* to the *B* power? Is $(\sqrt{2})^{\sqrt{2}}$ rational? *I don't know.*

But I do know that either it is or it isn't:

($(\sqrt{2})^{\sqrt{2}}$ is rational) v. ~($(\sqrt{2})^{\sqrt{2}}$ is rational).
If it is rational, QED.
If it isn't, then keeping *B* as $\sqrt{2}$, let *A* be $(\sqrt{2})^{\sqrt{2}}$.
Now is $((\sqrt{2})^{\sqrt{2}})^{\sqrt{2}}$ rational? Yes, because it is $(\sqrt{2})^2$, which is 2.

One way or another – and I don't have any idea which way it is – there are such a pair of numbers.[2]

The great benefit of this form of argument is that it permits you to finesse your ignorance about difficult matters and still achieve a demonstration. But you really can't use this delicious form of argument when the topic is dogs, say, instead of numbers. Is it true that every animal either is a dog or isn't a dog? What about coydogs and wolf hybrids? The boundaries of the *dog* concept are vague, and so are the boundaries of *coyote* and *wolf* and many other important concepts. These undeniable borderline cases are not just a nuisance to anyone intent on framing a fish-or-cut-bait argument; they typically disable the argument form altogether.[3]

[2] It is often noted that this is an example of an intuitively satisfying proof that proponents of intuitionist logic such as Brouwer cannot accept – a stiff price for intuitionism.

[3] In a challenge to the standard conception of borderline cases, Raffman (2005, 2014) argues that a borderline case for a vague term Φ lies between Φ and an incompatible (contrary) category Ψ but is neither Φ nor Ψ; for example, a borderline case of a dog lies between a dog and a coyote but is neither a dog nor a coyote or between a dog and a wolf but is neither a dog nor a wolf, etc. Thus bivalence and excluded middle are safe: the sentence "*x* is Φ" is false, and the sentence "*x* is not Φ" is true in a borderline case. (*Aside:* There are no higher-order borderline cases between Φ and borderline Φ because borderline Φ items are not Φ and borderline cases are not defined between contradictories – only between incompatibles.) For present purposes, the important point is that on Raffman's view, the class of not-Φ things *includes the borderline cases*; the nondogs include the coydogs. Hence, if she is right, my criticism of the fish-or-cut-bait strategy will require reformulation (but only reformulation): the problem will be that the negated disjunct, now covering a heterogeneous class containing the borderline cases as well as the polar opposites – the coydogs as well as the coyotes, the wolf hybrids as well as the wolves – does not generally support the sort of conclusion that euclideans wish to draw. (Raffman's approach does not entail sharp boundaries for vague words; see Raffman [2014], especially chaps. 2 and 4.)

An argument that exposes the impact of Darwinian thinking is David Sanford's (1975) nice "proof" that there aren't any mammals:

1. Every mammal has a mammal for a mother.
2. If there have been any mammals at all, there have been only a finite number of mammals.
3. But if there has been even one mammal, then by (1), there have been an infinity of mammals, which contradicts (2), so there can't have been any mammals. It's a contradiction in terms.

Because we know perfectly well that there are mammals, we take this argument seriously only as a challenge to discover what fallacy is lurking within it. And we know, in a general way, what has to give: if you go back far enough in the family tree of any mammal, you will eventually get to the therapsids, those strange, extinct bridge species between the reptiles and the mammals. (Technically, mammals are also classified as therapsids, the only surviving therapsids, but usually the term is used to refer to the premammalian nonreptilian species from which mammals descended.) A gradual transition occurred over millions of years from clear reptiles to clear mammals, with a lot of intermediaries filling in the gaps. What should we do about drawing the lines across this spectrum of gradual change? Can we identify a mammal, the Prime Mammal, that didn't have a mammal for a mother, thus negating premise (1)? On what grounds? Whatever the grounds are, they will compete with the grounds we could use to support the verdict that that animal was *not* a mammal – after all, its mother was a therapsid. What could be a better test of therapsid-hood than that? Suppose that we list ten major differences used to distinguish therapsids from mammals and declare that having five or more of the mammal marks makes an animal a mammal. Aside from being arbitrary – why ten instead of six or twenty, and shouldn't they be ordered in importance? – any such dividing line will generate lots of unwanted verdicts because during the long, long period of transition between obvious therapsids and obvious mammals there will be plenty of instances in which mammals (by our five+ rule) mated with therapsids

(fewer than five mammal marks) and had offspring that were therapsids born of mammals, mammals born of therapsids born of mammals, and so forth! Of course, we would need a "time machine" to see all these "anomalies" because the details are undetectable after all those millions of years. Just as well, since the details don't really matter in the long run. What should we do? We should quell our desire to draw lines. We can live with the quite unshocking and unmysterious fact that, you see, there were all these gradual changes that accumulated over many millions of years and eventually produced undeniable mammals.

The insistence that there *must* be a Prime Mammal, even if we can never know when and where it existed, is an example of *hysterical realism*. It invites us to reflect that if we just knew enough, we'd see – we'd *have* to see – that there is a special property of mammal-hood – the *essence* of mammal-hood – that defines mammals once and for all. To deny that there is such an essence, philosophers sometimes say, is to confuse metaphysics with epistemology: the study of what there (really) *is* with the study of what we can *know* about what there is. I reply that there *may* be occasions when thinkers do go off the rails by confusing a metaphysical question with a (merely) epistemological question, but this must be shown, not just asserted.[4] In this instance, the charge of confusing metaphysics with epistemology is just a question-begging way of clinging to one's crypto essentialism in the face of difficulties.

Richard Dawkins, in his recent essay recommending the retirement of the concept of *essence* (2014), writes

> Paleontologists will argue passionately about whether a particular fossil is, say, *Australopithecus* or *Homo*. But any evolutionist knows there must have existed individuals who were exactly intermediate. It's essentialist folly to insist on the necessity of shoehorning your fossil into one genus or the other. There never was an *Australopithecus* mother who gave birth to a *Homo* child, for every child ever born belonged to the same species as its mother. The whole system of labelling species with discontinuous names is geared to a time slice, the present, in which ancestors have been conveniently expunged from our awareness (and "ring species" tactfully ignored). If by some miracle every ancestor were preserved as a fossil,

[4] Passages in the last few paragraphs have been drawn, with minor revisions, from Dennett (2013, pp. 240–3).

> discontinuous naming would be impossible. Creationists are misguidedly fond of citing "gaps" as embarrassing for evolutionists, but gaps are a fortuitous boon for taxonomists who, with good reason, want to give species discrete names. Quarrelling about whether a fossil is "really" *Australopithecus* or *Homo* is like quarrelling over whether George should be called "tall". He's five foot ten, doesn't that tell you what you need to know?

So it isn't just philosophers who have trouble breaking the habit of presupposing essences. As Dawkins notes, there are good reasons for having tidy, "discrete" names for lineages, agreed-upon landmarks to work with, but then we mustn't mistake our convenient agreements for discoveries. Plato unforgettably recommends we carve nature at its joints, but there just aren't enough real, objective joints to suit our communicative purposes. We don't *need* to draw lines, but we *may* draw lines, arbitrarily, in the interest of practical taxonomy. Even if we make this move, disjunction elimination is pretty much disabled as a tool for demonstrating anything because wherever we've drawn our line, we are left with variations on one or both sides of our line that defy the sorts of generalizations that are needed to run the elimination arguments. But that is a good thing because the conclusions typically drawn from such arguments are apt to mislead us away from important truths, as we shall see.

In particular, the demand for essences with sharp boundaries blinds thinkers to the prospect of gradualist theories of complex phenomena, such as life, intentions, natural selection itself, moral responsibility, and consciousness.

If you hold that there can be no borderline cases of being alive (such as, perhaps, viruses or even viroids or motor proteins), you are more than halfway to *élan vital* before you start thinking about it. If no *proper part* of a bacterium, say, is alive, what "truth maker" gets added that tips the balance in favor of the bacterium's being alive? The three more or less standard candidates are having a metabolism, the capacity to reproduce, and a protective membrane, but since each of these phenomena, in turn, has apparent borderline cases, the need for an arbitrary cutoff doesn't evaporate. And if single-celled "organisms" (if they deserve to be called that!) aren't alive, how could two single-celled entities yoked together with no other ingredients be alive? And if not two, what would be special about a three-cell coalition? And so forth.

If, as Fodor (2008) insists, the frog either does or does not have the intention to catch a fly, you end up claiming that natural selection cannot account for adaptations:

I suppose it is likewise plausible that frogs catch flies with the intention of doing so. (If you are unprepared to swallow the attribution of intentions to frogs, please feel free to proceed up the phylogenetic ladder until you find a kind of creature to which such attributions are, in your view, permissible.) Now, intentions-to-act have intentional objects, which may serve to distinguish among them. A frog's intention to catch a fly, for example, is an intention to catch a fly, and is ipso facto distinct from, say, the frog's intention to sun itself on the leaf of a lily. (p. 2)

Now the intention to catch a fly is distinct from the intention to catch an "ambient black nuisance" even if, in the selective environment "fly" and "ambient black nuisance" are coextensive. Because natural selection "can't, as it were, 'see' the difference between intentional states that are extensionally equivalent" (p. 4), it cannot select *for* one intention rather than the other. This conclusion soon leads, by a cascade of disjunctions, to a killer disjunction: either natural selection has a mind or it doesn't. And, surprise, surprise, it doesn't. And because it doesn't, it cannot explain adaptations. The dubiety of this conclusion vies for top honors with the mythic aerodynamic "proof" that bumblebees can't fly.[5]

Fodor is well aware that he's legislating from an essentialist position by insisting that we all stick to "literal" readings of every term and either confirm or deny each proposition:

Surely, you may say, nobody could *really* hold that genes are *literally* concerned to replicate themselves? Or that natural selection literally has goals in mind when it selects what it does? Or that it's literally run by an intentional system? Maybe. Admittedly, the tactic of resorting to scare quotes when push comes to shove (as in 'what natural selection "prefers "', 'what Mother Nature "designs"', 'what the selfish genes "want"' and so forth) can make it hard to tell just what is being claimed in some of the canonical texts. Still, there are plenty of apparently unequivocal passages. Thus Pinker (1997, p. 93):

[5] The story appears to have some foundation in fact, though it has been transformed through retelling, a meme with quite a distinguished history, dating back to the 1930s, when August Magnan, a famous French entomologist, and his lab assistant, M. Saint-Lague, did the engineering calculations, as reported in Magnan's book, *Les Vols des Insects* (1934). Of course, Magnan realized that this was a *reductio* of current thinking in aeronautical engineering. See also John McMasters (1989). This footnote is drawn from Dennett and Plantinga (2011).

Was the human mind ultimately designed to create beauty? To discover truth? To love and to work? To harmonize with other human beings and with nature? The logic of natural selection gives the answer. The ultimate goal that the mind was designed to attain is maximizing the number of copies of the genes that created it. Natural selection cares only about the long-term fate of entities that replicate . . .

Fiddlesticks. The human mind wasn't created, and it wasn't designed and there is nothing that natural selection cares about; it just happens. This isn't Kansas, Toto. (p. 7, n. 12)

And if, like John Searle, you deny that there is any room for gradations of consciousness, or gradations of understanding, you end up declaring that "Strong AI" is impossible, or that consciousness is inexplicable, or both.

What we need to break through these self-imposed straitjackets of theoretical imagination is an appreciation of what I call the *sorta* operator, and a good way to see it in action is by putting Turing's revolutionary idea about computation in juxtaposition with Darwin's revolutionary idea about evolution. The pre-Darwinian world was held together not by science but by tradition: all things in the universe, from the most exalted ("man") to the most humble (the ant, the pebble, the raindrop), were the creations of a still more exalted thing, God, an omnipotent and omniscient intelligent creator – who bore a striking resemblance to the second-most exalted thing. Call this the trickle-down theory of creation. Darwin replaced it with the bubble-up theory of creation. One of Darwin's nineteenth-century critics put it vividly:

In the theory with which we have to deal, Absolute Ignorance is the artificer; so that we may enunciate as the fundamental principle of the whole system, that, IN ORDER TO MAKE A PERFECT AND BEAUTIFUL MACHINE, IT IS NOT REQUISITE TO KNOW HOW TO MAKE IT. This proposition will be found, on careful examination, to express, in condensed form, the essential purport of the Theory, and to express in a few words all Mr. Darwin's meaning; who, by a strange inversion of reasoning, seems to think Absolute Ignorance fully qualified to take the place of Absolute Wisdom in all the achievements of creative skill. (MacKenzie 1868)

It was, indeed, a strange inversion of reasoning. To this day, many people cannot get their heads around the unsettling idea that

a purposeless, mindless process can crank away through the eons, generating ever more subtle, efficient, and complex organisms without having the slightest whiff of understanding of what it is doing.

Turing's idea was a similar – in fact, remarkably similar – strange inversion of reasoning. The pre-Turing world was one in which computers were people, who had to understand mathematics in order to do their jobs. Turing realized that this was just not necessary: you could take the tasks they performed and squeeze out the last tiny smidgens of understanding, leaving nothing but brute, mechanical actions. IN ORDER TO BE A PERFECT AND BEAUTIFUL COMPUTING MACHINE, IT IS NOT REQUISITE TO KNOW WHAT ARITHMETIC IS.

What Darwin and Turing had both discovered, in their different ways, was the existence of competence without comprehension (Dennett 2009, from which material in the preceding paragraphs has been drawn, with revisions). This inverted the deeply plausible assumption that comprehension is in fact the source of all advanced competence. Why, after all, do we insist on sending our children to school, and why do we frown on the old-fashioned methods of rote learning? We expect our children's growing competence to flow from their growing comprehension; the motto of modern education might be "Comprehend in order to be competent." And for us members of *Homo sapiens*, this is almost always the right way to look at, and strive for, competence. I suspect that this much-loved principle of education is one of the primary motivators of skepticism about both evolution and its cousin in Turing's world, artificial intelligence. The very idea that mindless mechanicity can generate human-level – or divine-level! – competence strikes many as philistine, repugnant, an insult to our minds and to the mind of God.

Turing, like Darwin, broke down the mystery of intelligence (or intelligent design) into what we might call atomic steps of dumb happenstance, which, when accumulated by the millions, added up to a sort of pseudointelligence. The central processing unit of a computer doesn't *really* know what arithmetic is or understand what addition is, but it "understands" the "command" to add two numbers and put their sum in a register – in the minimal sense that it reliably adds when thus called on to add and puts the sum in the right place. Let's say it *sorta* understands addition. A few levels higher, the operating system doesn't *really* understand that it is checking for errors of transmission

and fixing them, but it *sorta* understands this and reliably does this work when called on to do so. A few further levels higher, when the building blocks are stacked up by the billions and trillions, the chess-playing program doesn't *really* understand that its queen is in jeopardy, but it *sorta* understands this, and IBM's Watson on *Jeopardy sorta* understands the questions it answers.

Why indulge in this *sorta* talk? Because when we analyze – or synthesize – this stack of ever more competent levels, we need to keep track of two facts about each level: what it *is* and what it *does*. What it *is* can be described in terms of the structural organization of the parts from which it is made – so long as we can assume that the parts function as they are supposed to function. What it *does* is some (cognitive) function that it (*sorta*) performs – well enough so that at the next level up we can make the assumption that we have in our inventory a smarter building block that performs just that function – *sorta* good enough to use. This is the key to breaking the back of the mind-bogglingly complex question of how a mind could ever be composed of material mechanisms. The *sorta* operator is, in cognitive science, the parallel of Darwin's gradualism in evolutionary processes. Before there were bacteria, there were *sorta* bacteria, and before there were mammals, there were *sorta* mammals, and before there were dogs, there were *sorta* dogs, and so forth. We need Darwin's gradualism to explain the huge difference between an ape and an apple, and we need Turing's gradualism to explain the huge difference between a humanoid robot and hand calculator. The ape and the apple are made of the same basic ingredients, differently structured and exploited in a many-level cascade of different functional competences. There is no principled dividing line between a *sorta* ape and an ape. The humanoid robot and the hand calculator are both made of the same basic, unthinking, unfeeling Turing bricks, but as we compose them into larger, more competent structures, which then become the elements of still more competent structures at higher levels, we eventually arrive at parts so (*sorta*) intelligent that they can be assembled into competences that deserve to be called *comprehending*. We use the intentional stance (Dennett, 1971, 1987) to keep track of the beliefs and desires (or "beliefs" and "desires" or *sorta* beliefs and *sorta* desires) of the (*sorta*)rational agents at every level from the simplest bacterium through all the discriminating, signaling, comparing, remembering circuits that compose the brains of animals from starfish to

astronomers. There is no principled line above which true comprehension is to be found – even in our own case. The small child *sorta* understands her own sentence, "Daddy is a doctor," and I *sorta* understand "$E = mc^2$." Some philosophers resist this antiessentialism: either you believe that snow is white or you don't; either you are conscious or you aren't; nothing counts as an approximation of any mental phenomenon – it's all or nothing. And to such thinkers, the powers of minds are insoluble mysteries because they are "perfect" and perfectly unlike anything to be found in mere material mechanisms.

When we turn to moral responsibility, consider the influential argument by Galen Strawson (2010):

1. You do what you do, in any given situation, because of the way you are.
2. So in order to be ultimately responsible for what you do, you have to be ultimately responsible for the way you are – at least in certain crucial mental respects.
3. But you cannot be ultimately responsible for the way you are in any respect at all.
4. So you cannot be ultimately responsible for what you do.

The first premise is undeniable: "the way you are" is meant to include your total state at the time, however you got into it. Whatever state it is, your action flows from it nonmiraculously. The second premise observes that you couldn't be "ultimately" responsible for what you do unless you were "ultimately" responsible for getting yourself into that state – at least in some regards. But according to step 3, this is impossible.

So step 4, the conclusion, does seem to follow logically. But let's look more closely at step 3. Why can't you be (ultimately) responsible for *some* respects, at least, of the way you are? In everyday life we make exactly this distinction, and it matters morally. Suppose that you design and build a robot and send it out into the world unattended and unsupervised and knowing full well the sorts of activities it might engage in, and suppose that it seriously injures somebody. Aren't you responsible for this, at least in some respects? Most people would say so. You made it; you should have foreseen the dangers – indeed, you *did* foresee some of the dangers – and now you are to blame, at least in part, for the damage

done. Few would have any sympathy for you if you insisted that you weren't responsible *at all* for the harm done by your robot.

Now consider a slightly different case: you design and build a person (yourself at a later time) and send yourself out into the risky world knowing full well the possible dangers you would encounter. You get yourself drunk in a bar and then get in your car and drive off. Aren't you responsible, at least in part, for the "way you were" when you crashed into a school bus? Common sense says of course. (The bartender or your compliant host may share the responsibility.) But how could this be, in the face of Strawson's knockdown argument? Well, remember that Strawson says that you can't be *absolutely* responsible for the way you are. But so what? Who would think it was important to be *absolutely* responsible? Here is what Strawson (2010) says:

> To be absolutely responsible for what one does, one would have to be *causa sui*, the cause of oneself, and this is impossible (it certainly wouldn't be more possible if we had immaterial souls rather than being wholly material).

The burden falls on Strawson and others to show why we ought to care about ultimate or absolute responsibility. I think it is just as obvious that people can gradually become morally responsible – *sorta* responsible – during their passage from infancy to adulthood as it is that lineages of reptiles and then therapsids can gradually become a lineage of mammals over the eons. You don't have to be an *absolute mammal* to be a mammal, and you don't have to be *absolutely responsible* to be responsible. So the constructive way of reading Strawson's argument is that, like Sanford's argument that there are no mammals, it is a *reductio ad absurdum* of the concept of absolute responsibility. The law may oblige us to draw a line (like the line for minimal age for a driver's license or for voting) but will understand that it is arbitrary, an imposed boundary, not a discovered joint in nature.

There are other philosophical puzzles that can benefit, I suspect, from exploring the no-longer-forbidden territory opened up by Darwin's critique of essentialism. Might there be important precursor grades of *semi-quasi-proto-sorta*-altruism, from which

we could get a better vantage point to look at "real" or "pure" altruism? Are there interesting epistemic states that are *almost* genuine knowledge? Once we give up essentialism for good, we can perhaps begin to reconstruct the most elevated philosophical concepts from more modest ingredients.