Presidential Addresses of the American Philosophical Association

2000-2001

Daniel C. Dennett

In Darwin's Wake, Where am I?

Brian Skyrms

The Stag Hunt

Lawrence Sklar

Naturalism and the Interpretation of Theories

In Darwin's Wake, Where am I?

Daniel C. Dennett, Tufts University

Parfois je pense; et parfois, je suis.

Paul Valéry[1]

Valéry's "*Variation sur Descartes*" excellently evokes the vanishing act that has haunted philosophy ever since Darwin overturned the Cartesian tradition. If *my body* is composed of nothing but a team of a few trillion robotic cells, mindlessly interacting to produce all the large-scale patterns that tradition would attribute to the non-mechanical workings of my mind, there seems to be nothing left over to be *me*. Lurking in Darwin's shadow there is a bugbear: the incredible Disappearing Self.[2] One of Darwin's earliest critics saw what was coming and could scarcely contain his outrage:

> 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.[3]

This "strange inversion of reasoning" promises–or threatens–to dissolve the Cartesian *res cogitans* as the wellspring of creativity, and then where will we be? Nowhere, it seems. It *seems* that if creativity gets "reduced" to "mere mechanism" *we* will be shown not to exist at all. Or, we will exist, but we won't be thinkers, we won't manifest genuine "Wisdom in all the achievements of creative skill." The individual as Author of works and deeds will be demoted: a person, it seems, is a barely salient nexus, a mere slub in the fabric of causation.

Whenever we zoom in on the act of creation, it seems we lose sight of it. The genius we thought we could see from a distance gets replaced at the last instant by stupid machinery, an echo of Darwin's shocking substitution of Absolute Ignorance for Absolute Wisdom in the creation of the biosphere. Many people dislike Darwinism in their guts, and of all the ill-lit, murky reasons for antipathy to Darwinism, this one has always struck me as the deepest, but only in the sense of being the most entrenched, the least accessible to rational criticism. There are thoughtful people who scoff at Creationism, dismiss dualism out of hand, pledge allegiance to academic humanism–and then get quite squirrelly when somebody proposes a Darwinian theory of creative intelligence. The very idea that all the works of human genius can be understood *in the end* to be mechanistically generated products of a cascade of generate-and-test algorithms arouses deep revulsion in many otherwise quite insightful, open-minded people.

Absolute Ignorance? Fie on anybody who would thus put "A" and "I" together! Serendipity is the wellspring of evolution, so it is fitting that an evolutionist such as I should adapt MacKenzie's happy capitalization for a purpose he could hardly have imagined. His outraged scoffing at the powers of Absolute Ignorance has an uncannily similar echo more than a century later in the equally outraged scoffing at those who believe in what John Searle[4] has called "strong AI," the thesis that *real* intelligence can be made by artifice, that the difference between a mindless mechanism and a mindful one is a difference of design (or *program*–since whatever you can design in hardware you can implement in a virtual machine that has the same competence).[5]

Darwin's "strange inversion of reasoning" turns an ancient idea upside-down. The "top-down" perspective on creative intelligence supposes that it always takes a big, fancy, smart thing to create a lesser thing. No horseshoe has ever made a blacksmith; no pot has fashioned a potter. Hence we–and all the other fancy things we see around us–must have been created by something still fancier, something like us only more so. To many–perhaps most–people, this idea is *just obvious*. Consider this page from a creationist propaganda mailing:

> 1. Do you know of any building that didn't have a builder? YES/NO
> 2. Do you know of any painting that didn't have a painter? YES/NO
> 3. Do you know of any car that didn't have a maker? YES/NO
>
> If you answered "YES" for any of the above give details:

But however strongly the idea appeals to common sense, Darwin shows us how it can be, in a word, false. Darwin shows us that a bottom-up theory of creation is, indeed, not only imaginable but empirically demonstrable. Absolute Ignorance *is* fully qualified to take the place of Absolute Wisdom in all the achievements of creative skill–*all* of them.

John Searle's Chinese Room thought experiment is a variation on the desperate joke of the creationists:

> Do you know of any machine that can understand Chinese? YES/NO
>
> If you answered "YES" give details!______________________

While the creationists' rhetorical questions merely gesture towards the presumed embarrassments facing anybody who tries to "give details" of an instance of bottom-up creation, Searle's challenge offers a survey of possible avenues the believers in strong AI might take in their attempts to "give details" and purports to rebut them one and all. The believers in strong AI have been remarkably unmoved by Searle's attempts at refutation, and the comparison of Searle's position with creationism shows why. Biologists who cannot *yet* explain some particular puzzle about the non-miraculous path that led to one marvel of nature or another, who cannot *yet* "give details" to satisfy the particular critic, nevertheless have such a fine track record of success in giving the details, and such a stable and fecund background theory to use in generating and confirming new details, that they simply dismiss the rhetorical implication: "You'll never succeed!" They calmly acknowledge that they may need to develop a few new wrinkles before they can declare victory. Believers in strong AI are similarly content to concede that all AI models to date have been deficient in many respects, orders of magnitude too simple, many of them pursuing particular visions of AI that are simply mistaken. They go on to note that Searle isn't challenging particular details of the attempts to date; he purports to be offering an argument for the *in principle* impossibility of strong AI, a conclusion that he insists is meant to cover all *imaginable* complications of the underlying theoretical framework. They know that their underlying theoretical framework is nothing other than the straightforward extension, into the human brain and all its peripheral devices and interfaces, of the Darwinian program of mindless mechanism doing, in the end, all the work. If Darwinian mechanisms can explain the existence of a skylark, in all its glory, they can surely explain the existence of an ode to a nightingale, too.[6] A poem is a wonderful thing, but not clearly more wonderful than a living, singing skylark.

Unsupportable antipathies often survive thanks to protective coloration: they blend into the background of legitimate objections to overstatements of the view under attack. Since the reach of Darwinian enthusiasm has always exceeded its grasp, there are always good criticisms of Darwinian excesses to hide amongst. Likewise, of course, for the excesses of the ideologues of AI. And so the battle rages, generating as much suspicion as insight. Darwinians who are sure that a properly nuanced, sophisticated Darwinism is proof against all the objections and misgivings–I am one such–should nevertheless recall the fate of the Freudian nags of the 50s and 60s, who insisted on seeing everything through the perspective of their hero's categories, only to discover that by the time you've attenuated your Freudianism to accommodate *everything*, it is Pickwickian Freudianism most of the way. Sometimes a cigar is just a cigar, and sometimes an idea is just an idea–not a meme–and sometimes a bit of mental machinery is not usefully interpreted as an adaptation dating back to our ancestral hunter-gatherer days or long before, even though it is, obviously, descended (with modifications) from some combination or other of such adaptations. We Darwinians will try to remind ourselves of this, hoping our doughty opponents will come to recognize that a Darwinian theory of creativity is not just a promising solution but the only solution in sight to a problem that is everybody's problem: *how can an arrangement of a hundred billion mindless neurons compose a creative mind, an I*?

William Poundstone has put the inescapable challenge succinctly in terms of "the old fantasy of a monkey typing *Hamlet* by accident." He calculates that the chances of this happening are "1 in 50 multiplied by itself 150,000 times."

> In view of this, it may seem remarkable that anything as complex as a text of *Hamlet* exists. The observation that *Hamlet* was written by Shakespeare and not some random agency only transfers the problem. Shakespeare, like everything else in the world, must have arisen (ultimately) from a homogeneous early universe. Any way you look at it *Hamlet* is a product of that primeval chaos.[7]

Where does all that design come from? What processes could conceivably yield such improbable "achievements of creative skill"? What Darwin saw is that design is always both valuable and costly. It does not fall like manna from heaven, but must be accumulated the hard way, by time-consuming, energy-consuming processes of mindless search through "primeval chaos", automatically preserving happy accidents when they occur. This broadband process of Research and Development is breathtakingly inefficient, but–this is Darwin's great insight–if the costly fruits of R and D can be thriftily conserved, copied, and re-used, they can be accumulated over time to yield "the achievements of creative skill." "This principle of preservation," Darwin says, "I have called, for the sake of brevity, Natural Selection."[8]

There is no requirement in Darwin's vision that these R and D processes run everywhere and always at the same tempo, with the same (in-)efficiency. Consider the unimaginably huge multi-dimensional space of all *possible* designed things–both natural and artificial. Every imaginable whale and unicorn, every automobile and spaceship and robot, every poem and mathematical proof and symphony finds its place somewhere in this Design Space. If we think of design work or R and D as a sort of *lifting* in Design Space[9] then we can see that the gradualistic, frequently back-sliding, maximally inefficient basic search process can on important occasions yield new conditions that speed up the process, permitting faster, more effective local lifting.[10] Call any such product of earlier R and D a *crane*, and distinguish it from what Darwinism says does not happen: *skyhooks***.**[11] Skyhooks, like manna from heaven, would be miracles, and if we posit a skyhook anywhere in our "explanation" of creativity, we have in fact conceded defeat–'Then a miracle occurs."[12]

What, then, is a mind? The Darwinian answer is straightforward. A mind is a crane, made of cranes, made of cranes, a mechanism of not quite unimaginable complexity that can clamber through Design Space at a giddy–but not miraculously giddy–pace, thanks to all the earlier R and D, from all sources, that it exploits. What is the anti-Darwinian answer? It is perfectly expressed by one of the 20th century's great creative geniuses (though, like MacKenzie, he probably didn't mean by his words what I intend to mean by them).

Je ne cherche pas; je trouve.

–Pablo Picasso

Picasso purports to be a genius indeed, someone who does not need to engage in the menial work of trial and error, generate-and-test, R and D; he claims to be able to *leap* to the summits of the peaks–the excellent designs–in the vast reaches of Design Space without having to guide his trajectory (he searches not) by sidelong testing at any waystations. As an inspired bit of bragging, this is *non pareil*, but I don't believe it for a minute. And anyone who has strolled through an exhibit of Picasso drawings (as I recently did in Valencia) looking at literally dozens of variations on a single theme, all signed—and sold–by the artist, will appreciate that whatever Picasso may have meant by his *bon mot,* he could not truly claim that he didn't engage in a time-consuming, energy-consuming exploration of neighborhoods in Design Space. At best he could claim that his own searches were so advanced, so efficient, that it didn't seem–to himself–to be design *work* at all. But then what did he have within him that made him such a great designer? A skyhook, or a superb collection of cranes?[13]

We can now characterize a mutual suspicion between Darwinians and anti-Darwinians that distorts the empirical investigation of creativity. Darwinians suspect their opponents of hankering after a skyhook, a miraculous gift of genius whose powers have no decomposition into mechanical operations, however complex and informed by earlier processes of R and D. Anti-Darwinians suspect their opponents of hankering after an account of creative processes that so diminishes the Finder, the Author, the Creator, that it disappears, at best a mere temporary locus of mindless differential replication. We can make a little progress, I think, by building on Poundstone's example of the *creation of the creator* of *Hamlet.* Consider, then, a little thought experiment.

Suppose Dr. Frankenstein designs and constructs a monster, Spakesheare, that thereupon sits up and writes out a play, *Spamlet.* My question is not about the author of *Waverley* but about the author of *Spamlet.*

Who is the author of *Spamlet*?

First, let's take note of what I claim to be irrelevant in this thought experiment. I haven't said whether Spakesheare is a robot, constructed out of metal and silicon chips, or, like the original

Frankenstein's monster, constructed out of human tissues–or cells, or proteins, or amino acids, or carbon atoms. As long as the design work and the construction were carried out by Dr. Frankenstein, it makes no difference to the example what the materials are. It might well turn out that the only way to build a robot small enough and fast enough and energy-efficient enough to sit on a stool and type out a play is to construct it from artificial cells filled with beautifully crafted motor proteins and other carbon-based nanorobots. That is an interesting technical and scientific question, but not of concern here. For exactly the same reason, if Spakesheare is a metal-and-silicon robot, it may be allowed to be larger than a galaxy, if that's what it takes to get the requisite complication into its program–and we'll just have to repeal the speed limit for light for the sake of our thought experiment. These technical constraints are commonly declared to be off-limits in these thought experiments, so so be it. If Dr. Frankenstein chooses to make his AI robot out of proteins and the like, that's his business. If his robot is cross-fertile with normal human beings and hence capable of creating what is arguably a new species by giving birth to a child, that is fascinating, but what we will be concerned with is Spakesheare's purported brainchild, *Spamlet*. Back to our question:

Who is the author of *Spamlet*?

In order to get a grip on this question, we have to look inside and see what happens in Spakesheare.[14] At one extreme, we find inside a file (if Spakesheare is a robot with a computer memory) or a basically *memorized* version of *Spamlet*, all loaded and ready to run. In such an extreme case, Dr. Frankenstein is surely the author of *Spamlet*[15], using his intermediate creation, Spakesheare, as a mere storage-and-delivery device, a particularly fancy word processor. *All* the R and D work was done earlier, and copied to Spakesheare by one means or another.

We can visualize this more clearly by imagining a sub-space of Design Space, which I call the Library of Babel, after Jorge Luis Borges' classic short story by that name.[16] Borges invites us to imagine a warehouse filled with books that appears to its inhabitants to be infinite; they eventually decide that it is not, but it might as well be, for it seems that on its shelves—in no order, alas—lie all the *possible* books.

Suppose that each book is 500 pages long, and each page consists of 40 lines of 50 spaces, so there are two thousand character-spaces per page. Each space is either blank, or has a character printed on it, chosen from a set of 100 (the upper and lower case letters of English and other European languages, plus the blank and punctuation marks).[17] Somewhere in the Library of Babel is a volume consisting entirely of blank pages, and another volume is all question marks, but the vast majority consist of typographical gibberish; no rules of spelling or grammar, to say nothing of sense, prohibit the inclusion of a volume. Five hundred pages times two thousand characters per page gives a million character-spaces per book, so there are $100^{1,000,000}$ books in the Library of Babel. Since it is estimated[18] that there are only 100^{40} (give or take a few) *particles* (protons, neutrons and electrons) in the region of the universe we can observe, the Library of Babel is not remotely a physically possible object, but thanks to the strict rules with which Borges constructed it in his imagination, we can think about it clearly.

We need some terms for the quantities involved. The Library of Babel is not infinite, so the chance of finding anything interesting in it is not literally infinitesimal.[19] These words exaggerate in a familiar way, but we should avoid them. Unfortunately, all the standard metaphors—astronomically large, a needle in a haystack, a drop in the ocean—fall comically short. No *actual* astronomical quantity (such as the number of elementary particles in the universe, or the time since the Big Bang measured in nanoseconds) is even visible against the backdrop of these huge-but-finite numbers. If a readable volume in the Library were as easy to find as a particular drop in the ocean, we'd be in business! Dropped at random into the Library, your chance of ever encountering a volume with so much as a grammatical sentence in it is so vanishingly small that we might do well to capitalize the term—*Vanishingly* small—and give it a mate, *Vastly*, short for Very-much-more-than-astronomically.[20]

It is amusing to reflect on just how large this finite set of possible books is, compared with any actual library. Most of the books are pure gibberish, as noted, so consider the Vanishing subset of books composed entirely of English words, without a single misspelling. It is itself a Vast set, of course, and contained within it, but Vanishingly hard to find, is the Vast subset whose English words are lined up in grammatical sentences. A Vast but Vanishing subset of this subset in turn is the subset of books composed of English sentences that actually make sense. A Vast but Vanishing subset of these are about somebody named John, and a Vast but Vanishing subset of these are about the death of John F. Kennedy. A Vast but Vanishing subset of these are true . . . and a Vast but Vanishing subset of the possible true books about the death of JFK are written entirely in limericks. There are many orders of magnitude more possible true books in limerick form about the death of JFK than there are books in the Library of Congress.

Now we are ready to return to that needle-in-a-haystack, *Spamlet*, and consider how the trajectory to this particular place in the Library of Babel was traversed in actual history. If we find that the whole journey was already completed by the time Spakesheare's memory was constructed and filled with information, we know that Spakesheare played no role at all in the search. Working backwards, if we find that Spakesheare's only role was running the stored text through a spell-checker before using it to guide its typing motions, we will be unimpressed by claims of Spakeshearian authorship. This is a measurable, but Vanishing, part of the total R and D. There is a sizable galaxy of near-twin texts of *Spamlet*—roughly a hundred million different minor mutants have but a single uncorrected typo in them, and if we expand our horizon to include one typo per page, we have begun to enter the land of Vast numbers of variations on the theme. Working back a little further, once we graduate from typos to *thinkos*,[21] those arguably mistaken, or sup-optimally chosen, words, we have begun to enter the land of serious authorship, as contrasted with mere copy-editing. The relative triviality of copy-editing, and yet its unignorable importance in shaping the final product gets well represented in terms of our metaphor of Design Space, where every little bit of lifting counts for something, and sometimes a little bit of lifting moves you onto a whole new trajectory. As usual, we may quote Ludwig Mies van der Rohe at this juncture: "God is in the details."

Now let's turn the knobs on our thought experiment, as Douglas Hofstadter has recommended[22] and look at the other extreme, in which Dr. Frankenstein leaves most of the work to Spakesheare. The most realistic scenario would surely be that Spakesheare has been equipped by Dr. Frankenstein with a virtual past, a lifetime stock of pseudo-memories of experiences on which to draw while responding to its Frankenstein-installed obsessive desire to write a play. Among those pseudo-memories, we may suppose, are many evenings at the theater, or reading books, but also some unrequited loves, some shocking close calls, some shameful betrayals and the like. Now what happens? Perhaps some scrap of a "human interest" story on the network news will be the catalyst that spurs Spakesheare into a frenzy of generate-and-test, ransacking its memory for useful tidbits and themes, transforming–transposing, morphing–what it finds, jiggling the pieces into temporary, hopeful structures that compete for completion, most of them dismantled by the corrosive processes of criticism that nevertheless expose useful bits now and then, and so forth, and all of this multi-leveled search would be somewhat guided by multi-level, internally generated evaluations, including evaluation of the evaluation… of the evaluation functions as a response to evaluation of… the products of the ongoing searches.[23]

Now if the amazing Dr. Frankenstein had actually anticipated all this activity down to its finest grain at the most turbulent and chaotic level, and had hand-designed Spakesheare's virtual past, and all its search machinery, to yield just this product, *Spamlet*, then Dr. Frankenstein would be, once again, the author of *Spamlet*, but also, in a word, God. Such Vast foreknowledge would be simply miraculous. Restoring a smidgen of realism to our fantasy, we can set the knobs at a rather less extreme position and assume that Dr. Frankenstein was unable to foresee all this in detail, but rather delegated to Spakesheare most of the hard work of completing the trajectory in Design Space to *one literary work or another*, something to be determined by later R and D occurring within Spakesheare itself. We have now arrived, by this simple turn of the knob, in the neighborhood of reality itself, for we already have actual examples of impressive artificial Authors that Vastly outstrip the foresight of their own creators. Nobody has yet created an artificial playwright worth serious attention, but an artificial chess player–IBM's Deep Blue–and an artificial composer–David Cope's EMI–have both achieved results that are, *in some respects*, equal to the best that human creative genius can muster.

Who beat Garry Kasparov, the reigning World Chess Champion? Not Murray Campbell or any of his IBM team. Deep Blue beat Kasparov. Deep Blue designs better chess games than any of them can design. None of them can author a winning game against Kasparov. Deep Blue can. Yes, but. Yes, but. I am sure many of you are tempted to insist at this point that when Deep Blue beats Kasparov at chess, its brute force search methods are *entirely* unlike the exploratory processes that Kasparov uses when he conjures up his chess moves. But that is simply not so–or at least it is not so in the only way that could make a difference to the context of this debate about the universality of the Darwinian perspective on creativity. Kasparov's brain is made of organic materials, and has an architecture importantly unlike that of Deep Blue, but it is still, so far as we know, a massively parallel search engine which has built up, over time, an outstanding array of heuristic pruning techniques that keep it from wasting time on unlikely branches. There is no doubt that the investment in R and D has a different profile in the two cases; Kasparov has methods of extracting good design principles from past games, so that he can recognize, and know enough to ignore, huge portions of the game space that Deep Blue must still patiently canvass *seriatim*. Kasparov's "insight" dramatically changes the shape of the search he engages in, but it does not constitute "an *entirely* different" means of creation. Whenever Deep Blue's exhaustive searches close off a *type* of avenue that it has some means of recognizing (a difficult, but not impossible task), it can re-use that R and D whenever it is appropriate, just as Kasparov does. Much of this analytical work has been done for Deep Blue by its designers, and given as an innate endowment, but Kasparov has likewise benefitted from hundreds of thousands of person-years of chess exploration transmitted to him by players, coaches and books. It is interesting in this regard to contemplate the suggestion recently made by Bobby Fischer, who proposes to restore the game of chess to its intended rational purity by requiring that the major pieces be *randomly* placed in the back row at the start of each game (random, but mirror image for black and white). This would instantly render the mountain of memorized openings almost entirely obsolete, for humans and machines alike, since only rarely would any of this lore come into play. One would be thrown back onto a reliance on fundamental principles; one would have to do more of the hard design work in real time–with the clock running. It is far from clear whether this change in rules would benefit human beings more than computers. It all depends on which type of chess player is relying most heavily on what is, in effect, rote memory–reliance *with minimal comprehension* on the R and D of earlier explorers.

The fact is that the search space for chess is too big for even Deep Blue to explore exhaustively in real time, so like Kasparov, it prunes its search trees by taking calculated risks, and like Kasparov, it often gets these risks pre-calculated. Both presumably do massive amounts of "brute force" computation on their very different architectures. After all, what do neurons know about chess? Any work *they* do must be brute force work of one sort or another.

It may seem that I am begging the question in favor of a computational, AI approach by describing the work done by Kasparov's brain in this way, but the work has to be done somehow, and no *other* way of getting the work done has ever been articulated. It won't do to say that Kasparov uses "insight" or "intuition" since that just means that Kasparov himself has no privileged access, no insight, into how the good results come to him. So, since nobody knows how Kasparov's brain does it–least of all Kasparov–there is not yet any evidence at all to support the claim that Kasparov's means are "entirely unlike" the means exploited by Deep Blue. One should remember this when tempted to insist that "of course" Kasparov's methods are hugely different. What on earth could provoke one to go out on a limb like that? Wishful thinking? Fear?

But that's just chess, you say, not art. Chess is *trivial* compared to art (now that the world champion chess player is a computer). This is where David Cope's EMI comes into play.[24] Cope set out to create a mere efficiency-enhancer, a composer's aid to help him over the blockades of composition any creator confronts, a high-tech extension of the traditional search vehicles (the piano, staff paper, the tape recorder, etc.). As EMI grew in competence, it promoted itself into a whole composer, incorporating more and more of the generate-and-test process. When EMI is fed music by Bach, it responds by generating musical compositions in the style of Bach. When given Mozart, or Schubert, or Puccini, or Scott Joplin, it readily analyzes their styles and composes new music in their styles, better pastiches than Cope himself–or almost any human composer–can compose. When fed music by two composers, it can promptly compose pieces that eerily unite their styles, and when fed, all at once (with no clearing of the palate, you might say) all these styles at once, it proceeds to write music based on the totality of its musical experience. The compositions that result can then also be fed back into it, over and over, along with whatever other music comes along in MIDI format, and the result is EMI's own "personal" musical style, a style that candidly reveals its debts to the masters, while being an unquestionably idiosyncratic integration of all this "experience." EMI can now compose not just two-part inventions and art songs but whole symphonies–and has composed over a thousand, when last I heard. They are good enough to fool experts (composers and professors of music) and I can personally attest to the fact that an EMI-Puccini aria brought a lump to my throat–but then, I'm on a hair trigger when it comes to Puccini, and this was a good enough imitation to fool me. David Cope can no more claim to be the composer of EMI's symphonies and motets and art songs than Murray Campbell can claim to have beaten Kasparov in chess.

To a Darwinian, this new element in the cascade of cranes is simply the latest in a long history, and we should recognize that the boundary between authors and their artifacts should be just as penetrable as all the other boundaries in the cascade. When Richard Dawkins notes[25] that the beaver's dam is as much a part of the beaver phenotype–its *extended phenotype*–as its teeth and its fur, he sets the stage for the further observation that the boundaries of a human author are exactly as amenable to extension. In fact, of course, we've known this for centuries, and have carpentered various semi-stable conventions for dealing with the products of Rubens, of Rubens' *studio*, of Rubens' various students. Wherever there can be a helping hand, we can raise the question of just who is helping whom, what is creator and what is creation. How should we deal with such questions? To the extent that anti-Darwinians simply want us to preserve some tradition of authorship, to have some *rules of thumb* for determining who or what shall receive the honor (or blame) that attends authorship, their desires can be acknowledged and met, one way or another (which doesn't necessarily mean we should meet them). To the extent that this is not enough for the anti-Darwinians, to the extent that they want to hold out for authors as an objective, metaphysically grounded, "natural kind" (oh, the irony in those essentialist wolf-words in naturalist sheep's clothing!), they are looking for a skyhook.

The renunciation of skyhooks is, I think, the deepest and most important legacy of Darwin in philosophy, and it has a huge domain of influence, extending far beyond the skirmishes of evolutionary epistemology and evolutionary ethics. If we commit ourselves to Darwin's "strange inversion of reasoning," we turn our backs on compelling ideas that have been central to the philosophical tradition for centuries, not just Aristotle's essentialism and irreducible *telos*, but also Descartes's *res cogitans* as a causer outside the mechanistic world, to name the three that had been most irresistible until Darwin came along. The siren songs of these compelling traditions still move many philosophers who have not yet seen fit to execute the inversion, sad to say. Clinging to their pre-Darwinian assumptions, they create problems for themselves that will no doubt occupy many philosophers for years to come.[26] The themes all converge when the topic is creativity and authorship, where the urge is to hunt for an "essence" of creativity, an "intrinsic" source of meaning and purpose, a locus of responsibility somehow insulated from the causal fabric in which it is embedded, so that within its boundaries it can generate, from its *own* genius[27], its *irreducible* genius, the meaningful words and deeds that distinguish us so sharply from mere mechanisms.

Plato called for us to carve nature at its joints, a wonderful biological image, and Darwin showed us that the salient boundaries in the biosphere are not the crisp set-theoretic boundaries of essentialism, but the emergent effects of historical processes. As one species turns into two, the narrow isthmus of intermediates disappears as time passes, leaving islands, concentrations sharing family resemblances, surrounded by empty space. As Darwin noted (in somewhat different terms), there are feedback processes that enhance separation, actively depopulating this middle ground. We might expect the same sort of effects in the sphere of human mind and culture, cultural habits or practices that favor the isolation of the processes of artistic creation in a single mind. "Are you the author of this?" "Is this all your own work?" The mere fact that these are familiar questions shows that there are cultural pressures encouraging people to *make* the favored answers come true. A small child, crayon in hand, huddled over her drawing, slaps away the helping hand of parent or sibling, because she wants this to be *her* drawing. She already appreciates the norm of pride of authorship, a culturally imbued bias built on the palimpsest of territoriality and biological ownership. The very idea of being an artist shapes her consideration of opportunities on offer, shapes her evaluation of features she discovers in herself. And this in turn will strongly influence the way she conducts her own searches through Design Space, in her largely unconscious emulation of Picasso's ideal, or, if she is of a contrarian spirit, defying it, like Marcel Duchamp:

> Cabanne: What determined your choice of readymades?
>
> Duchamp: That depended on the object. In general, I had to beware of its "look." It's very difficult to choose an object, because, at the end of fifteen days, you begin to like it or to hate it. You have to approach something with an indifference, as if you had no aesthetic emotion. The choice of readymades is always based on visual indifference and, at the same time, on the total absence of good or bad taste…[28]

There is a persistent problem of imagination management in the debates surrounding this issue: people on both sides have a tendency to underestimate the resources of Darwinism, imagining simplistic alternatives that do not exhaust the space of possibilities. Darwinians are notoriously quick to find (or invent) differences in *genetic fitness* to go with every difference they observe, for instance. Meanwhile, anti-Darwinians, noting the huge distance between a beehive and the *St. Matthew Passion* as created objects, are apt to suppose that anybody who proposes to explain both creative processes with a single set of principles must be guilty of one reductionist fantasy or another: "Bach had a gene for writing baroque counterpoint just like the bees' gene for forming wax hexagons" or

"Bach was just a mindless trial-and-error mutator and selector of the musical memes that already flourished in his cultural environment." Both of these alternatives are nonsense, of course, but pointing out their flaws does nothing to support the idea that ("therefore") there must be irreducibly *non-Darwinian* principles at work in any account of Bach's creativity. In place of this dimly imagined chasm with "Darwinian phenomena" on one side and "non-Darwinian phenomena" on the other side, we need to learn to see the space between bee and Bach as populated with all manner of mixed cases, differing from their nearest neighbors in barely perceptible ways, replacing the chasm with a traversable gradient of non-minds, protominds, hemi-demi-semi minds, magpie minds, copycat minds, aping minds, clever-pastiche minds, "path-finding" minds, "ground-breaking" minds, and eventually, genius minds. And the individual minds, of each caliber, will themselves be composed of different sorts of parts, including, surely, some special-purpose "modules" adapted to various new tricks and tasks, as well as a cascade of higher-order reflection devices, capable of generating ever more rarefied and delimited searches through pre-selected regions of the Vast space of possible designs.

It is important to recognize that genius is itself a product of natural selection and involves generate-and-test procedures all the way down. Once you have such a product, it is often no longer particularly perspicuous to view it solely as a cascade of generate-and-test processes. It often makes good sense to leap ahead on a *narrative* course, thinking of the agent as a self, with a variety of projects, goals, presuppositions, hopes, …In short, it often makes good sense to adopt the intentional stance towards the whole complex product of evolutionary processes. This effectively brackets the largely unknown and unknowable mechanical microprocesses as well as the history that set them up, and puts them out of focus while highlighting the patterns of rational activity that those mechanical microprocesses track so closely. This tactic makes especially good sense to the creator himself or herself, who must learn not to be oppressed by the revelation that on close inspection, even on close *intro*spection, a genius dissolves into a pack rat, which dissolves in turn into a collection of trial-and-error processes over which nobody has ultimate control.

Does this realization amount to a loss–an elimination–of selfhood, of genius, of creativity? Those who are closest to the issue–the artistic and scientific geniuses who have reflected on it–often confront this discovery with equanimity. Mozart is reputed to have said of his best musical ideas: "Whence and how do they come? I don't know and I have nothing to do with it."[29] The painter Philip Guston is equally unperturbed by this evaporation of visible self when the creative juices start flowing:

> When I first come into the studio to work, there is this noisy crowd which follows me there; it includes all of the important painters in

history, all of my contemporaries, all the art critics, etc. As I become involved in the work, one by one, they all leave. If I'm lucky, every one of them will disappear. If I'm really lucky, I will too.[30]

In closing, I would like to acknowledge a few of my co-authors:

Anonymous
Jorge Luis Borges
David Cope
Charles Darwin
Richard Dawkins
Susan Dennett
René Descartes
Marcel Duchamp
Thomas Edison
Bobby Fischer
Philip Guston
Douglas Hofstadter
Nicholas Humphrey
Robert MacKenzie
Tony Marcel
Victoria McGeer
Ludwig Mies van der Rohe
Pablo Picasso
William Poundstone
John Searle
William Shakespeare
Mary Shelley
Paul Valéry

Endnotes

1. Paul Valéry, *Cahiers*, ed. Judith Robinson, 2 vols. (Paris: Edition de la Pleiade 1973-74). vol. 2 (1974), p. 1388.

2. Dennett, D. 1984, *Elbow Room*, p13.

3. Robert Beverley MacKenzie, 1868, *The Darwinian Theory of the Transmutation of Species Examined* (published anonymously "By a Graduate of the University of Cambridge"). London: Nisbet & Co. Quoted in a review, *Athenaeum*, no 2102, February 8, p217.

4. Searle, John, 1980, "Minds, Brains and Programs," *Behavioral and Brain Sciences*, 3, pp.417-58.

5. This is obviously true of all competences of information-processing or control, but not of productive or transformative processes, such as lactation, which requires the transport and assembly of particular

materials. Since Searle purports to distinguish the brain's "control powers" from its "bottom-up causal powers" that "produce intentionality," some have thought Searle imagines intentionality to be a special sort of substance secreted by the brain. Since he denies this, he owes us some other way to distinguish these mysterious causal powers from the control powers that software can implement and an explanation of why they are not implementable in a virtual machine.

6. This perspective helps to explain the visceral appeal to many onlookers of the various *apparent* alternatives to Darwinian mechanism that have flourished over the years. The most prominent recently have been the appeal to "self-organization" "on the edge of chaos" (Stuart Kauffman, Per Bak, and others), and "dynamical systems theory" in both evolution and cognition (Esther Thelen, Walter Freeman, Timothy van Gelder, and others), and, of course, Stephen Jay Gould's insistence that evolution is not, as I have claimed (building on the work of theorists from Darwin to Fisher and Haldane to Williams and Maynard Smith), fundamentally an algorithmic process. After the smoke of battle clears, these ideas can be readily seen to be, at best, interesting complications of the basic Darwinian mechanisms, just as connectionist architectures and embodied cognition models are interesting complications of the basic ideas of AI. These controveries are, at best, constructive disagreements over how to "give the details", not challenges to the basic Darwinian vision.

7. William Poundstone, 1985, *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge*, New York: Wm Morrow, p23.

8. Charles Darwin, *Origin of Species*, Ch. 4 summary (p.127 in facs. edition)

9. This tactic of mapping evolutionary processes and results onto space is a natural and oft-used metaphor, exploited in models of hill-climbing, and peaks in adaptive landscapes, to name the most obvious and popular applications. Its naturalness does not guarantee its soundness, of course, and may even mask its limitations, but since the basic mapping strategy has proven to be particularly useful in expressing *criticisms* of over-simple evolutionary ideas (e.g., Kauffman's "rugged landscape", Eigen's "quasi-species"), it is not obviously biased in favor of simplistic visions of Darwinism.

10. John Maynard Smith and Eörs Szathmary, 1995, *The Major Transitions in Evolution*, Oxford: Freeman, identify eight occasions (major transitions) when the evolutionary process became more efficient, creating cranes.

11. Daniel C. Dennett, *Darwin's Dangerous Idea*, 1995, "The Tools for R and D: Skyhooks or Cranes?" pp 73-80.

12. See the famous cartoon by Sydney Harris, in which the physicist's blackboard is covered with impressive formulae, except for this bracketed phrase in the middle, which leads the onlooker scientist to say "I think you should be more explicit here in step two." (reprinted in Daniel Dennett, *Consciousness Explained*, 1991, p38).

13. I have been unable to discover the source of Picasso's claim, which is nicely balanced by a better known remark by a more down-to-earth creative genius, Thomas Edison: "Genius is one per cent. inspiration and ninety-nine per cent. perspiration." (in a newspaper interview. *Life* [1932], ch. 24, according to the *Oxford Dictionary of Quotations*.)

14. Yes, I intend the homage to an old favorite of mine, *What Happens in Hamlet*, by J. Dover Wilson (1951, Cambridge Univ. Press).

15. unless we find there is a Ms. Shelley who is the author of Dr. Frankenstein. . . . !

16. Borges, J. L. 1962, *Labyrinths: Selected Stories and other Writings*, New York: New Directions. [La Biblioteca de Babel, 1941, in El jardin de los senderos que se bifurcan, published with another in *Ficciones*. 1956, Emece Editores, S. A., Buenos Aires.]

17. Borges chose slightly different figures: books 410 pages long, with 40 lines of 80 characters. The total number of characters per book is close enough to mine (1,l312,000 versus 1,000,000) to make no difference. I chose my rounder numbers for ease of handling. Borges chose a character set with only 25 members, which is enough for upper-case Spanish (with a blank, a comma and a period as the only punctuation), but not for English. I chose the more commodious 100 to make room without any doubt for the upper and lower case letters and punctuation of all the Roman alphabet languages.

18. Steven Hawking insists on putting it this way: "There are something like ten million million million million million million million million million million million million million (1 with eighty zeroes after it) particles in the region of the universe that we can observe." *A Brief History of Time*, New York: Bantam, p.129. Michael Denton (*Evolution: A Theory in Crisis*, London: Burnett Books. 1985) provides the estimate of 1070 atoms in the observable universe. Manfred Eigen (*Steps Towards Life*, Oxford University Press 1992, p.10) calculates the volume of the universe as 1084 cubic centimeters.

19. The Library of Babel is finite, but curiously enough, it contains all the grammatical sentences of English within its walls. But that's an infinite set, and the library is finite! Still, any sentence of English, of whatever length, can be broken down into 500-page chunks, each of which is somewhere in the library! How is this possible? Some books may get used more than once. The most profligate case is the easiest to understand: since there are volumes which each contain a single character and are otherwise blank, repeated use of these one hundred volumes will create any text of any length. As Quine *(Quiddities: An Intermittently Philosophical Dictionary*, Cambridge, MA: Harvard Univ. Press. 1987) points out, in his informative and amusing essay, "Universal Library," if you avail yourself of this strategy of re-using volumes, and translate everything into the ASCII code your wordprocessor uses, you can store the whole Library of Babel in two extremely slender volumes, in one of which is printed a 0 and in the other of which appears a 1! (Quine also points out that Theodor Fechner, the psychologist, propounded the fantasy of the universal library long before Borges.)

[20] Quine, *loc.cit.*, coins the term “hyperastronomic” for the same purpose. The previous two paragraphs are drawn, with minor changes, from *Darwin’s Dangerous Idea*, pp108-9.

[21] For more on this concept, see my “From Typo to Thinko: When Evolution Graduated to Semantic Norms,” forthcoming in the Fyssen conference volume on cultural evolution.

[22] Douglas R. Hofstadter, 1981, “Reflections,” (on Searle) in Hofstadter and Dennett, eds., *The Mind’s I: Fantasies and Reflections on Self and Soul*, 1981, New York: Basic Books and Hassocks, Sussex: Harvester

[23] Shakespeare himself was, of course, a tireless exploiter of the design work of others, and may well have been poking fun at his own reputation, quoting a critic, when he had Autolycus describe himself as “a snapper-up of unconsidered trifles” in *A Winter’s Tale* (Act IV, scene iii). Thanks to Tony Marcel for drawing this passage to my attention.

[24] For the details, see David Cope, ed., *Virtual Music* (forthcoming from MIT Press), including my commentary, “Collision Detection, Muselot, and Scribble: Some Reflections on Creativity.”

[25] Richard Dawkins, 1982, *The Extended Phenotype*, Oxford and San Francisco: Freeman.

[26] Three examples: Jerry Fodor’s series of flawed theories of psychosemantics; John Searle’s inability to account for how “intrinsic intentionality” could evolve when it has no “control power” consequences visible to selective pressure; John McDowell’s quest for a non-Darwinian alternative to what he calls “bald naturalism,” a struggle to secure a variety of normativity that is not the mere *as-if* normativity he finds discernible in evolution. See Dennett, *Darwin’s Dangerous Idea*, 1995, and “Granny versus Mother Nature — No Contest,” *Mind & Language*,11 no.3, 1996, pp 263-269, and “Review of John Searle, *The Rediscovery of the Mind*” *Journal of Philosophy*, 60, (4), 193-205, Apr. 1993, for my analyses of Fodor’s and Searle’s difficulties. My discussion of McDowell must be deferred to another occasion.

[27] See “Do-It-Yourself Understanding,” in *Brainchildren*, Cambridge MA: MIT Press, 1998, pp59-80, for my analysis of this theme in Fred Dretske’s search for a privileged place where the understanding happens.

[28] *Dialogues with Marcel Duchamp*, Pierre Cabanne (transl Ron Padgett), New York, Viking Press, 1971, p.48. Thanks to Nicholas Humphrey and Victoria McGeer for ideas expressed in the previous paragraph.

[29] In an oft-quoted but possibly spurious passage–see *Darwin’s Dangerous Idea*, p346-7.

[30] I have been unable to locate the source of Guston’s quote, but I have found much the same remark attributed to the composer, John Cage, a close friend and contemporary of Guston’s, who [is said to have]said this about painting:

> When you are working, everybody is in your studio-the past, your friends, the art world, and above all, your own ideas-all are there. But as you continue painting, they start leaving, one by one, and you are left completely alone. Then, if you are lucky, even you leave.

Like all other creators, Guston and I like to re-use what we find, adding a few touches from time to time.