(ii)

Romanticism and Consciousness

It is most unfortunate, but the point of this story has been reached where a justification of the expression "Murphy's mind" has to be attempted.

— Samuel Beckett: *Murphy*.

{...}

I have long since lost the original, but recall a notebook which commenced with the melodramatic pronouncement:
"The *central* question is the relation of the mind to Nature."

That wasn't bad, actually.

The two obvious answers are that one (pick which) contains the other.

The less obvious answer is that both are correct at once.

{...}

Bad acid (1967)

(Must I say something about the drugs? would this otherwise be *Hamlet* without the Prince of Denmark? — More like omitting Polonius, I am thinking. In any case —)

To the epistemological adventurer, the hippie always seemed unforgivably bourgeois. He took drugs and acted weird and dressed funny and thought strange things, true. But he *needed* the drugs. That was the crucial difference. — Indeed the really amazing thing was how many people hadn't realized they *had* minds, before they started fucking with them. — Thus it was always obvious that, whatever this was, it couldn't last.

{...}

I recall an expedition to the Haight in the autumn following the Summer of Love: my comrades and I, a company of geeks, wandered around the neighborhood all day, gawking at the outward signs of the New World Order; and then, needing a place to crash, followed a couple of the natives who were trying to make up the rent back to their townhouse. — Which must have been costing them more than the few bucks we ponied up, but the economic mysteries of that era then as now defied rational explanation.

There we found ourselves in the company of the usual motley crew — something like the melting pot of the Hollywood warmovie foxhole: the speed freak with fresh blood on his tracks who explained to me that the essence of Dylan was that he would take any drug, he wouldn't even ask what it was, he didn't give a shit, he'd just shoot it, snort it, smoke it, jam it up his ass in a suppository, whatever, at which I nodded sagely [thinking "What is this bullshit existential bravado?"]

and said Wow man, far out — the couple humping furiously in the back room who barely nodded on their way out to panhandle[1] among the tourists — somebody bitching about Janis and her rowdy juicer friends running around the neighborhood — somebody who gave the definitive one-line review of the latest Airplane album:[2] "It's more of a *meth* sound... ." — They were all still wearing those swinging-London faux-Victorian-dandy outfits like the Stones did but everything was dirty, because nobody bathed, or washed or even changed their clothing, because they had nothing else to change into. It was clear there was already some kind of devolution going on from love and flowers and cosmic consciousness to closed minds and harder drugs, but it was not complete — indeed one of the guys crashing there among us, mirabile dictu, was one of Owsley's own distributors,[3] and was carrying an entire gram of LSD in a vial, which we all examined with the utmost curiosity. — So we talked to these characters, cracked wise in our customary fashion, and they all kept laughing at what we were saying. — Finally someone asked "Hey, what are you guys on?" — Here followed an embarrassed pause as we looked at one another. — I wanted to say "DNA", but restrained myself because these were our hosts and it would have been rude to put them down so bluntly, even if the insult would have sailed over their heads. — Instead we performed the geek ritual of self-effacement, mumbled the usual self-deprecating apologies for being brilliant and original, and doubtless missed the opportunity to make our fortunes selling the latest three-letter wonder drug to a witless mob of scientific illiterates.

---

[1] A friend of mine — a far more gifted hustler than ever I could aspire to be — reported that, as an energetic panhandler in the Haight in the Summer of Love, he could net about seven dollars an hour from the tourists; at the time, about five times the minimum wage. (Yes, this guy later sold out, went to law school, and became a Yuppie parasite. We all know the story of the Decline of the West.)

[2] This must have been *After Bathing at Baxter's*.

[3] Really, one look at his pupils and you knew he was telling the truth.

{...}

Though I personally had never had difficulty having uncanny experiences without the aid of any other chemicals than coffee and fatigue poisons, I was nonetheless an early adopter, if then a rapid renouncer, of the abuse of psychedelics. — I dropped acid a few times without serious incident, but then had a slight disagreement with one of my best friends; this somehow became amplified and distorted, and in consequence even though I understood perfectly well that none of it was really his fault or mine, that the drugs had opened the doors of misperception, I couldn't stand to be around him for a year or more afterward.

It seemed obvious that anything that could poison my relationships so easily was not to be trusted; the argument that you could undo it all by repeating the experiment a couple of days later was obviously ridiculous, like claiming your brains would not be permanently addled so long as you kept banging yourself on the head with the same brick. — So by and large I let that wave pass me by.

{...}

But what was philosophically interesting about the experience was that the argument people like Huxley and Watts tried to make from it was completely backwards. They said the hallucinations that came from taking psychedelics proved that quotidian reality was an illusion, and the brain was a crude filter that dulled your senses and prevented you from perceiving things are they really were; from seeing the light that shone forth from the radiant center of things.

But clearly it was just the opposite, a simple matter of cause and effect. The brain was the determinant, and simply twiddling a few knobs on the back of the chassis to adjust some excitation thresholds could make the mental reality that depended on it entirely different.

In the comic books mad scientists had invented pills and potions that could turn you into a giant or make you invisible; Albert Hofman had invented one that could make you see God. — Nothing more succinctly proved the chemical basis of consciousness, the ultimate identity of the mental and the material. — That there was a speculative, a metaphysical chemistry. — A chemistry of the soul.

{...}

Drugs at any rate foregrounded the problem of consciousness: there *was* something you were altering, so what was it?

One thing was that the difference, as it were, between acting in the movie and watching it could suddenly be manifest: you seemed to step outside yourself, to realize that the point of conscious life was something like you were playing yourself, it was a role; you were a simulation of yourself.

There was some dim realization of this going around, but some inappropriate interpretation got hung on it, maybe it was the instinct to moralize — it came off as: what you think you are is phony; it's just you playing a role, it isn't the real you. — The real you, presumably, was something you had yet to discover, and a host of "spiritual guides" (aka bullshit artists) popped up like psilocybin mushrooms to make their fortunes explaining it.

But it was always obvious to me that the ultimate fact was that you *were* this simulation — this act, the game — and that the fact you could perceive this was also part of the game; was wired into it; was intrinsic to it, in fact. This didn't mean that all foundations vanished and reality was illusion — if reality was an illusion, then the distinction between reality and illusion was itself an illusion, any child should have been able to see that — it was just that reality was much more complicated, and therefore interesting, than anticipated. The Self might be an illusion, but it was easy to see that meant the universe was as well, a

complex of them, and that *this made no difference*.[4] — What is, still is, and the point is still to try to understand it.

The converse (to separation) was also true: the difference between watching or fantasizing or listening to a story and living in it could suddenly vanish, so that you couldn't tell which was which. Listening to a song might trigger a series of hallucinations consistent with your being somehow *in* it (clichéd though it must seem, my first such experience came listening to "White Rabbit", and the chessmen really did get up and tell me where to go), taking part rather than merely witnessing. — When Dylan sitting in his loft abruptly begins a verse "The peddler now speaks/To the countess who's pretending to care for him," it's because there's a movie on the television in the background (or he thinks there is), and he's suddenly not watching it from without but from within, he's entered into it. — Among other classics of Stoner Literature, in Pynchon's *Gravity's Rainbow* there is Slothrop's extended fantasy about disappearing down the toilet bowl — brought rather too vividly to the screen by Danny Boyle in *Trainspotting* [1995].

{...}

Here pro forma I should insert the Uncanny Experience anecdote that will be found in all such memoirs — the occasion when drugs did do — *something* — though I never figured out exactly what……[5]

{...}

What would it be like if telepathy were really possible? if you could enter another person's head? — Like this, you had to think: completely different though topologically equivalent, a distorted map,

---

[4] Bloom: "Shakespeare invents in *Hamlet* the destruction of any boundary between being oneself and playing oneself." Or recognizes that there never was any such distinction.

[5] Can't bring myself to do it. Too stupid.

the Self viewed in a funhouse mirror. It would be just like taking drugs.

{...}

Spinoza somewhere defines consciousness as the idea of ideas, which is close to what I eventually decided, but this wasn't the proximate cause for the inspiration. — Here I must digress: when I was a freshman I went through one of those phases in which the world seemed ripe to be remade anew — myself along with it — and briefly contemplated correcting my posture, which then as now was atrocious. I made a few vain stabs at trying to stand up straight, posing upright in front of the mirror — predictably this got me nowhere — but then more sensibly thought to compare myself to my natural peer group, the residents of Fleming House. — And then after a brief survey realized, to my considerable amusement, that the three worst slouches in the place belonged to myself and a couple of upperclassmen — call them Chris and Mitch[6] — whom I knew to be the smartest guys in the school. — "This slouch is aspirational!" I declared to myself — laughed —and never gave a thought to my posture again.

(It was here, incidentally, that I first began to realize why therapy so often turns people into assholes: because they don't really change; they only bullshit themselves into thinking they have. — It is like running a foreign emulation program: pretending to be something you are not. — I *might* have changed my posture, but not my nature.)

The two gentlemen in question were both notorious acidheads, and like all the rest of us at the time expended much effort theorizing about the drug experience.

---

[6] Both subsequently had distinguished academic careers, and why pretend I can cloud their reputations by naming them here.

Most of the commonplaces in circulation were predictably worthless, but I recall Mitch telling me[7] that Chris had said any mathematical model of the mind would have to be isomorphic to the group of its own automorphisms.

He had based this on some particular type of acid experience with which I wasn't familiar, but the idea was so obviously beautiful that it didn't matter. Out of the countless hours of stoned silly bullshit I had endured it stood out with an electric clarity. Everything else I'd heard I had ignored. But this I stole at once.

{...}

Later it occurred to me that a formal equivalent closer to the mark was the apparently paradoxical property of the lambda calculus, which Scott[8] was able to explain, that any model would have to have the property (*prima facie* impossible in set theory) that it was isomorphic to the ensemble of functions from the model to itself.

{...}

This was not the best drug-inspired contribution to the philosophy of mind, however. That was due to my friend [BA],[9] who one evening

---

[7] Chris I barely knew, but Mitch and I were, considering the disparity in our social positions, fairly tight. I must have made a favorable impression upon him, because I was told he said about me "That guy is the weirdest kid at Caltech," which I still figure to have been the finest compliment I ever received. Of course I tried to live up to his expectations.

[8] Dana Scott, "Continuous Lattices." Oxford University Computing Laboratory Technical Monograph PRG-7, August, 1971. — Quoting from the abstract: "The main result of the paper is a proof that every topological space can be embedded in a continuous lattice which is homeomorphic (and isomorphic) to its own function space. The function algebra of such spaces provides mathematical models for the Church-Curry lambda-calculus."

[9] Now deceased, alas, and never shy about crediting his drug abuse for making him a computer millionaire. But let us not name names.

came up the stairs into the hallway we shared in Fleming talking animatedly to another party, both of them stoned out of their gourds on acid and apparently arguing about the Nature of Television; as he passed my doorway he posed the question "Does the picture make the sound, or the sound make the picture?" — Of course this solved the mind-body problem at a single stroke.

{...}

And this, I suppose, has been a Flashback.

{…}

Stoned at King Soopers (1967)

An epiphany: entering the grocery, the door opening automatically for me — remembering Emerson, "All the thoughts of a turtle are turtle"; realizing that I had seen into the mental state of the photocell: pure anticipation.

The old principle that automata must have souls. — Really, this was there already in Leibniz.

{...}

Ghost in the shell (1971)[10]

There is supposed to be an argument from Gödel's theorem to show that the mind can't be a machine, but I've never understood it. Of course I have never thought that was my fault.[11]

Penrose, for one, made a book out of it; and though I didn't believe him either it was amusing that whole issues of the journals[12] were repurposed to try to refute him.

At any rate both sides of the argument are bullshit. It doesn't matter whether minds are machines or not. Even machines aren't machines. This can be seen in at least two ways:

— First, "machine" in the sense of artificial intelligence never really means "Turing machine" anyway; rather one augmented by a (true)[13]

---

[10] The basic argument (here reconstructed) has not changed a great deal since it first occurred to me, though it has obviously been revised and amended to reflect the march of mathematical progress.

[11] There are (at least) a couple of good reasons to be skeptical. — First, Gödel himself always thought this was a consequence of his incompleteness theorem, and was said to have been working on a formal proof of the proposition; which, however, he never finished, and didn't publish. That seems suspicious. — Second, the idea that a human agent could find itself trapped in repetitive cycles of mechanical behavior is supposed to be prima facie absurd; nonetheless it is the fundamental thesis of psychoanalysis; and indeed the Freudian method looks a lot like teaching a Turing machine its Gödel sentence. (Compare Thomas Mann: "No man remains what he was once he has recognized himself.")

[12] The book [published in 1989] was *The Emperor's New Mind*. For expressions of outrage cf., e.g., *Behavioral and Brain Sciences* Vol. 13 #4 (1990), pp. 643-705, and Vol. 16 #3 (1993), pp. 611-622.

[13] I.e., one referring to what theory terms "an Oracle", some external source of input like a Geiger counter recording radioactive decays. — Purely computational (pseudo) random number generators fake it, by producing sequences which are determined by such complicated rules that they "look" random (a literature has been expended trying to define the implicit oxymoron), i.e. take a long time to repeat, but on the other hand can be rapidly

random number generator — a source of randomness for nondeterministic algorithms; neural networks, for instance, fall under this description, as do Metropolis and genetic algorithms.

— Second, even deterministic machines aren't deterministic. — That is, though you have the picture (sharpened by formal models) of a system with a delimited[14] set of states, whose behavior is determined by a function which computes the next state (in a discrete series) from the current state, and assume that knowledge of the next-step function entails knowledge of all its iterates — that "deterministic" means "completely predictable", in other words — this is a kind of optical illusion. It doesn't really work that way.

{...}

It's more amusing to explain this anecdotally.

I saw Skinner lecture once, in Berkeley in the early Seventies. This was shortly after the publication of one of his numerous paeans to Mind Control:[15] he spoke in a large lecture hall, to a full house packed with an extremely hostile crowd, and though he couldn't win them over, he did at least earn their respect. — There is a certain naive pigheaded charm some nerds possess, and he had it in great measure. If nothing else, I admired his balls.

---

computed. It is truly amazing how often the naive employment of these mechanisms leads to mortifying blunders. Nearly every serious programming project I have undertaken has been almost immediately been sidetracked by an attempt to write a better random number generator than the one that has just fucked me in the nose.

[14] This is tricky: machine theory allows not simply for the case of a finite state set, but also for a finite "internal" state set augmented by potentially-infinite auxiliary storage, the tape of a Turing machine or the stack of a PDSA, e.g., which can only be accessed finitely, e.g. one item at a time. — In practice, of course, all machines are really finite, and immense ingenuity is expended to overcome limitations of time and space.

[15] Probably *Beyond Freedom and Dignity* [1971].

We were all jammed in like sardines, and I was sitting in the aisle a few feet downhill from my girlfriend, so as it turned out I couldn't talk to her until afterward and it wasn't obvious we were attached. Instead I found myself embedded among a covey of attractive female undergraduates. One of them was lecturing her friends on the nature and context of the debate we were participating in, and every time she hesitated because she didn't quite know how to continue, I finished her sentence for her. — This provided me with the standard anecdote I used in later years to describe what Berkeley was like, in the Golden Age: this was the first, last, and only time a girl wanted to go home with me because I knew Beckett wrote *Endgame*.

At any rate I was fascinated by Skinner's insistence on the predictability of human behavior; there was an echo of that Freudian certitude that had always seemed so maddening, but his explanatory apparatus was cleaner, much more austere. So what was wrong with it?

Part of it, obviously, when I read over theoretical behaviorism later[16] to find the basis for his claims, was that the most consistent version of his approach made it a point of dogma not simply that one *should* not but that one *could* not assign internal states to the organism; since simple thought experiments showed that removing the brain from the skull would produce a noticeable difference in behavior, at least among people who hadn't voted for Nixon, that was obviously wrong. — Part of it was that the kinds of laboratory experiments to which behaviorists confined themselves made essentially meaningless measurements of a kind which could not, for instance, tell you anything about the functioning of even the simplest digital computer.[17]

---

[16] Not that I wasn't familiar with it already from, e.g., Russell's synopses in *The Analysis of Mind* [1921], but it was instructive to read the modern literature and observe how little theory had progressed since Watson and Pavlov.

[17] A technical refinement of the point, which Chomsky used to great polemical effect, was that though for the simplest class of finite-state automata internal states can in principle be defined

But the main thing — what was instantly suspect — was his claim that behaviorist methods would suffice to explain even the "behavior" of mathematicians. For this seemed, after all, to be a bizarre assertion: were we seriously to think that from considerations of elementary physics — presumably by solving some system of differential equations — not that Skinner ever wrote any down, of course, but an explanatory framework based on the measurement of quantities expressed in real numbers — i.e. founded on physics envy — would inevitably (as any real physicist could instantly see) lead to such a theory — that we could tell whether a mathematician was going to be able to prove a theorem? How was one *complicated* mathematical problem — indeed all of them at once — supposed to reduce to another which seemed so much simpler?[18] — And *why* it seemed bizarre wasn't difficult to figure out. For though if we asked the mathematician to prove, say, some statement in the predicate calculus it might seem unlikely on intuitive/romantic grounds that we'd be able to describe the necessary "creative leap", really it isn't necessary to appeal to this at all: one could simply ask the mathematician to attempt mechanically to construct a proof using some method like semantic

---

away as equivalence classes of mappings from inputs to outputs, this [a] relies on the examination of infinite sets, and [b] the conditioned-reflex prescription applied to a finite training set of stimuli and responses only works for this simplest class, and cannot determine the behavior of machines that recognize more complex grammars. Since such machines already existed and even then were generating our utility bills, this seemed a fairly crushing objection.

[18] Actually it isn't impossible that a relatively simple differential equation, or system of them, could be universal in the sense of Turing; the solution of Hilbert's tenth problem showed something analogous for Diophantine problems, i.e. that there is an equation of the fourth degree in 14 variables that is universal: see Martin Davis, "Hilbert's Tenth Problem Is Unsolvable," *American Mathematical Monthly*, March 1973, 233-269.— One could conjecture, in other words, the existence of a universal *analog* computer. — But a simulation that modeled a universal Turing machine with a differential equation wouldn't be any *simpler*. The inherent difficulty of the problem is irreducible. So the picture you have of having found a solution is a kind of optical illusion. — "All I have to do is solve this equation, and..." — but how? In practice you have only replaced one intractable computation by another of equivalent difficulty.

tableaux; and then observe that whether this procedure terminates on arbitrary input is, in general, undecidable. — I.e. you needn't appeal to a magical black-box mechanism at all; even if you know the mechanism, even if the box is transparent, it makes no difference. — So the grand reductive gesture of pretending the box *has* no internal degrees of freedom is doubly pointless.

{...}

Put another way, one need not challenge Skinner with the problem of predicting whether, say, Gauss sitting at his worktable will be able to come up with a proof of, say, Goldbach's conjecture;[19] one can simply ask Skinner to tell us whether Gauss in performing the arithmetical check will find a counterexample to Goldbach's conjecture in finite time; and if so, *when*. Because this means that the behaviorist must then in effect be able to tell us in advance whether Goldbach's conjecture is true. (And decide this by solving some magic differential equation, or system of them.)[20] — True, we can, if we are faster, stay ahead of Gauss in the computation. But this is not an *effective procedure*; we can't guarantee an answer to the question exists in advance.[21] — We can't say how the computation will come out. — And therefore, in

---

[19] Communicated in a letter to Euler in the 18th century, the statement (based then on very flimsy empirical evidence, based now on dismayingly extensive tests) that every even number greater than 2 is the sum of two primes. A proof now does appear to be closer, but the feeling has generally been that if there really are "natural" elementary statements about the integers that are true but not provable, they would look like this. (Gödel himself referred to this possibility explicitly; see the notes to *Gödel 1972a* in his *Collected Works*, Volume 2.)

[20] Given the hypothetical universal system one might solve the equations "by computer", i.e. numerically, but then we simply have one machine emulating another of equivalent complexity; nothing is *reduced*, in other words.

[21] I.e. though I may not know before I perform the computation that 16117667 times 16283543 is 262452723654181, I do know that there is an answer, and if I follow the rules for multiplication I will find it within a certain number of steps which can be bounded in advance. Not all computations come with such guarantees.

the most significant sense, we cannot *predict* what Gauss is going to do, *even if he is emulating a machine*.[22]

{...}

There are various equivalents[23] that illustrate the case equally well, but the canonical question is the halting problem for Turing machines: suppose we give Gauss the description of a Turing machine, and an input tape — all this is finite — and then ask Skinner to tell us his prescription for deciding, in the general case, when/whether Gauss will finish computing the answer, and what the result will be. — To explain his behavior, i.e. — But he can't, because this is known to be impossible. — Conceivably Skinner might object that the proof of unsolvability assumes the validity of Church's thesis, an essentially metaphysical hypothesis[24] which he rejects — another myth which will dissolve in the acid bath of his scientific rationality; but then he's saying that he has some method of computation (an oracle, e.g.) that is more powerful than a Turing machine. — At which point we tell him to put up or shut up. And the rest is silence.

---

[22] I take it for granted that a human (like Gauss) can emulate any Turing machine; since after all the idea of the Turing machine is that it formalizes the abilities of a human calculator. — It is assumed, in other words, that the objection that Gauss might not have enough time or scratch paper is frivolous and irrelevant to the principle at issue. (This has nothing to do with his *behavior*.)

[23] The word problem for semigroups, e.g., which asks whether there's a general method for deciding whether two strings of symbols are equivalent under a given finite set of equational transformations, or the general Diophantine problem (Hilbert's Tenth), whether an mechanical procedure exists to determine whether a polynomial equation in a finite number of variables with integer coefficients has integer solutions.

[24] Fred Thompson was the first guy I heard call it that. He was certainly right.

{...}

You can summarize the lesson of this gedankenexperiment as follows: since prediction is simply computation,[25] machines in general are not predictable; since people can emulate arbitrary machines,[26] the behavior of people is not predictable.

So behaviorism isn't completely useless; its refutation teaches us something valuable.

{...}

This doesn't explain why a mob of hippies showed up to howl for Skinner's head, of course. That had to do with the supposed conflict between the freedom of the will and determinism. But I think the real issue there is related, essentially psychological, the anxiety that you feel about the possibility not that your actions are "determined" in some complex and unknowable fashion, but that they can be *predicted*.

We all remember Dostoevsky's lengthy rant[27] in *Notes from the Underground*, the famous Crystal Palace passage about the conflict between the freedom of the will and mathematical certainty:

> ... then, you say, science itself will teach man ... that he never has really had any caprice or will of his own, and that he himself is something of the nature of a piano-key or the stop of an organ, and that there are, besides, things called the laws of nature; so that everything he does is not done by his willing it, but is done

---

[25] This seems self-evident, but in the same way all propositions do that insinuate metaphysical hypotheses. (Here again Church's thesis.)

[26] By definition: when Turing refers to a "computer" in his original paper, he means a human following rules with pencil and paper; electronic computers did not yet exist.

[27] It would be anachronism to call it that, but this is a classic example of what is now called a flame.

> of itself, by the laws of nature. Consequently we have only to discover these laws of nature, and man will no longer have to answer for his actions and life will become exceedingly easy for him. All human actions will then, of course, be tabulated according to these laws, mathematically, like tables of logarithms up to 108,000, and entered in an index; or, better still, there would be published certain edifying works of the nature of encyclopaedic lexicons, in which everything will be so clearly calculated and explained that there will be no more incidents or adventures in the world.[28]

Or more succinctly:

> Good heavens, gentlemen, what sort of free will is left when we come to tabulation and arithmetic, when it will all be a case of twice two make four? Twice two makes four without my will. As if free will meant that!

But though the existentialist antihero of the *Notes* thus insists perversely on behaving irrationally to express his defiance of soulless rationalism, he needn't have bothered. Arithmetic itself is perverse enough.

That is, though it is already difficult enough to understand the traditional problem — your *will* is still free even if what you *want* is determined, [29]and so what — the point is really that determinism appears to entail predictability, and prediction allows control: if people are machines, then seemingly they can be *used* as machines; *that* is the terror of mechanism.

---

[28] This is the Constance Garnett translation.

[29] I cheerfully admit that emotional responses are often predictable, at least for most people much of the time; else they would be more difficult to manipulate. But here again the claims of psychology are exaggerated.

You have the oppressive sense that some puppet master like Skinner can look over your shoulder (with his "table of logarithms") and nod smugly at everything you do, because he has foreseen it all in advance; and since he knows what you will do when he pushes your buttons, *he can make you do whatever he likes*. — And Skinner of course endorses this interpretation at every turn, this is the plan for his Utopia. — That it might be determined[30] in advance but not known or even knowable — well, there is something that never occurred to the determinists; omniscient though they were supposed to be. In fact it doesn't seem to have occurred to anybody.

{...}

The anxiety is not unknown among physicists. There is a strong resemblance, e.g., between Eddington's argument (made nearly as soon as the uncertainty principle was invented)[31] that the indeterminacy of quantum mechanics permitted the freedom of the will, and Penrose's rather weird assertion (1989) that "microtubules" within the cell could turn the brain into some kind of quantum computer beyond the reach of Turing.[32] In both cases it is clearly the predictability of the mechanical that disturbs them. — Your will cannot be free if someone can know what you will do. — More than that, an artist or a musician or a mathematician cannot be truly creative, since whatever they produce is simply the result of a mechanical process. One could simply write a Shakespeare emulation program and output *Hamlet*, without the intervention of the fifty million monkeys with typewriters. — This is a slightly more

---

[30] To return to the model of the system of differential equations, there are in general existence theorems that tell you they *have* solutions which are determined uniquely by their initial conditions. This doesn't mean you can say what the solutions are. (Or — the butterfly effect — that they are stable under infinitesimal perturbations, which is a necessary condition for computer simulation.)

[31] Cf. *The Nature of the Physical World* [1927].

[32] Pure science fiction, so far as anyone can tell.

interesting problem, but the predictability issue is again key: one might in principle be able to program a (pseudo)machine to write something like *Hamlet*, but it would never turn out the same way twice, and given time and sufficiently many rewrites would turn into something else entirely. — Whether that would satisfy Penrose I don't know. But my credentials as an unreconstructed Romantic are unquestioned, and it satisfies me.

{...}

There is also an amusing functional equivalence between Skinner's implicit[33] assertion that he could predict the answer to any mathematical question from the laws governing the organism (the differential equations, or whatever) and Plato's insistence that all mathematical knowledge is something the soul obtained in a previous life/is engraved upon the Forms; it is accordingly suggestive that they envisioned similar Utopias. (And that they bore a suspicious resemblance to the Crystal Palace.) — Who were our behaviorist overlords going to be, but the new Guardians? — Moreover there are parallels with the apparent aims of the classical school of artificial intelligence, as exemplified by Minsky: if the brain was just a machine running a determined program, then those select few who could read the source code could make mere humans (aka "the lusers")[34] do whatever they wanted; traditional hacker culture was also based on fantasies of control, the domination of the programmers over the programmed.

---

[33] You have to say "implicit" because it is obvious he did not understood what was coming out of his mouth. Certainly he never understood Chomsky's critique.

[34] Traditionally the MIT school divided people who interacted with computers into two classes, programmers and users; the former were the master race, the latter, serfs and peons. — It is not an accident that, as Big Tech continues to conquer the world, more and more of it reverts to feudalism.

{...}

Another fantasy of determinism, indulged by the imaginative, is that one ought to be able to predict the course of history in advance. — This is not, precisely, the usual motivation of the self-styled grand theoreticians of history, who seem not to have advanced beyond Linnaean notions of classification — Spengler, e.g., goes on at great length in his philosophical preamble about Goethe, morphology, the incapacity of trivial concepts of causality to grasp the architecture of Destiny, etc.[35] — the game of hypothesis and prediction never caught on among the German idealists, obviously — but it is a fairly common speculation in science fiction. Isaac Asimov's *Foundation* novels are probably the most famous examples, and have been quite influential[36] in that respect: he imagines the decline and fall of a galactic empire on the pattern of Gibbon's Rome, and a dedicated cabal of monks, privy to detailed advance knowledge of the pattern history must follow, working to preserve civilization through the ensuing Dark Age, whose duration they will thus be able to minimize.[37]

The superficially convincing argument for the possibility of such prescience is the analogy with statistical mechanics: you don't need to

---

[35] In the translation of Charles Francis Atkinson: "The means whereby to identify dead forms is Mathematical Law. The means whereby to understand living forms is Analogy." — "... there can be no question of taking spiritual-political events ... at their face value, and arranging them on a scheme of 'causes' or 'effects' ... ." — "That there is, besides a necessity of cause and effect — which I may call the logic of space — another necessity, an organic necessity in life, that of Destiny — the logic of time — is a fact of the deepest inward certainty... ." — "Mathematics and the principle of Causality lead to a naturalistic Chronology and the idea of Destiny to a historical ordering of the phenomenal world." — And so on. Of course all this is nonsense.

[36] Sometimes in unobvious ways: the Nobel laureate Paul Krugman, for example, is a science fiction fan, and has often remarked that Asimov's vision of a social science that could make rigorous predictions inspired him to study economics.

[37] This idea of a monastic order preserving knowledge through a Dark Age is another favorite theme of science fiction; see for instance Walter Miller's *A Canticle for Leibowitz*.

know how each individual gas molecule is moving to calculate the pressure on a cylinder. — The argument probably fails on appeal to the butterfly effect, since there are many examples e.g. of critical battles won or lost by accidents of timing, and (pace Tolstoy) great men (and women) do seem to appear fortuitously and decisively alter the course of events — this is a more complex dynamical problem than that posed by a gas, after all — still, though one can't predict the weather exactly, one *can* predict climate change; so one might guess that on a longer time scale the rolls of the human dice may even out.

Nonetheless something similar to Skinner/Gauss does apply: the future of industrial civilization as we have it right now, for example, depends at bottom on facts of physics and astronomy as yet unknown — whether room temperature superconductors exist, whether fusion reactors can ever be practical, what results may come from mining the asteroids, whether irreversible ecological collapse is really at hand — whether an undetected asteroid is going to run into the Earth and reprise the extinction of the dinosaurs — and you can't tell how human history will turn out without knowing the answers to these external questions. — As was the past so determined: the history of the modern world follows in large part from the contingent fact that when Columbus sailed west, there was an extra continent to bump into. — So the one kind of omniscience presupposes the other. Even economics, which involves measurable quantities and superficially seems more easily predictable, depends at bottom on the ways that we can extract free energy from our environment, and thus on unpredictable boundary conditions and undiscovered facts of mathematics and physics (and chemistry and biology and geography and ...) which cannot be known without — well, without being known.[38] How could an economist in 1950 have predicted that nuclear power based on fission reactors would turn out to be more trouble

[38] Here I'm sure Heidegger would insert some rhapsody on the knowable knowingness of being-known, but — thank the gods who have not yet fled — I lack his gift for tautological obfuscation.

than it was worth, or foreseen the laser, the transistor, the photovoltaic cell, the microchip, or Moore's Law? — Von Neumann saw none of that coming, and he was as omniscient as anyone could have been at that time — for instance, he famously stated that four computers like his[39] primitive MANIAC[40] would suffice for all the computational needs of the world.[41]

{...}

A slightly weaker statement, whose relationship to undecidability is still not completely understood, is that a computation may not be impossible but nonetheless may be prohibitively difficult. This might seem like a frivolous objection were it not the case that relatively simple problems can be shown to be unsolvable within existing space and time.[42]

---

[39] Actually constructed by Nicholas Metropolis at Los Alamos following Von Neumann's IAS design, but why quibble. — Authorship of the acronym, which was meant to stamp out this reprehensible practice in its infancy and failed miserably, has been ascribed to both.

[40] Less powerful than a pocket calculator of the Seventies, and many orders of magnitude less powerful than the contemporary iPhone; which exceeds in computational power the fastest supercomputers of even the Eighties.

[41] As a final note, Lockheed Martin is supposed to be pitching a tool called the World-Wide Integrated Crisis Early Warning System (Google at your own risk), originally a project funded by DARPA, which is supposed to have had some success anticipating national and international crises. Apparently among other things it predicts the collapse of the Russian government within a couple of years; surely a consummation devoutly to be wished. — The historian Peter Turchin, on the other hand, on the basis of mathematical analysis of a large data set measuring a variety of historical trends, finds parallels between previous periods of crisis and the current situation of the United States, and predicts the disintegration of civil society within the decade. — And he does indeed begin *War and Peace and War* [2006] by invoking the example of Asimov's hero Hari Seldon.

[42] David Ruelle ("Is Our Mathematics Natural?" *Bulletin of the AMS*, Vol. 19, Number 1, July 1988) mentions a suggestion of Pierre Cartier that the axioms of set theory might be inconsistent but a proof of this would be so long that it couldn't be performed in the physical universe.

One class of examples would include the travelling salesman problem, which scales exponentially in the number of cities;[43] a greater degree of difficulty may be found in problems like computing Ramsey numbers, or evaluating the Ackermann function, which is defined as follows:

$$A(x,0) = 0$$
$$A(0,y) = 2y$$
$$A(x,1) = 2$$

else

$$A(x,y) = A(x-1, A(x,y-1))$$

Then

$$A(1,n) = 2^n$$
$$A(n,1) = 2$$
$$A(n,2) = 4$$
$$A(2,3) = 16$$
$$A(2,4) = 65536$$

in general

$$A(2,n) = A(1,\big(A(2,n-1)\big) = 2^{A(2,n-1)}$$
$$A(3,1) = 2$$
$$A(3,2) = 4$$
$$A(3,3) = 65536$$

and

$$A(3,4) = \cdots$$

[43] Given a planar map and the positions of n cities upon it, to construct a route of minimum length that visits each city exactly once; for n around 120 the number of possibilities that must be examined exceeds the number of cells of dimension the Planck length in the visible universe.

i.e., this is a recursion that will not terminate before the stars go out, and the answer couldn't be written down[44] if you used all the volumes in Borges' Library of Babel.

{...}

Regarding the Goldbach conjecture, subsequent developments have only confirmed Gödel's intuition. Certainly it is possible that there is some simple and elegant proof of this proposition, but it seems more likely there is not; and then there are curious questions about how complicated a proof, even if one does exist, might have to be. The proof of the celebrated four-color theorem,[45] for example, another result with an extremely simple statement[46] which defied demonstration for several generations, turned out not to involve (at least has not thus far) the elegant manipulation of powerful abstractions developed from mathematical theories of great scope and formal beauty — as did, for instance, the proof of the famous Weil conjectures (Deligne 1973), the proof of Mordell's conjecture (Faltings 1983), and the celebrated proof of the Taniyama conjecture (Wiles 1993/4), which entailed the last theorem of Fermat, for three

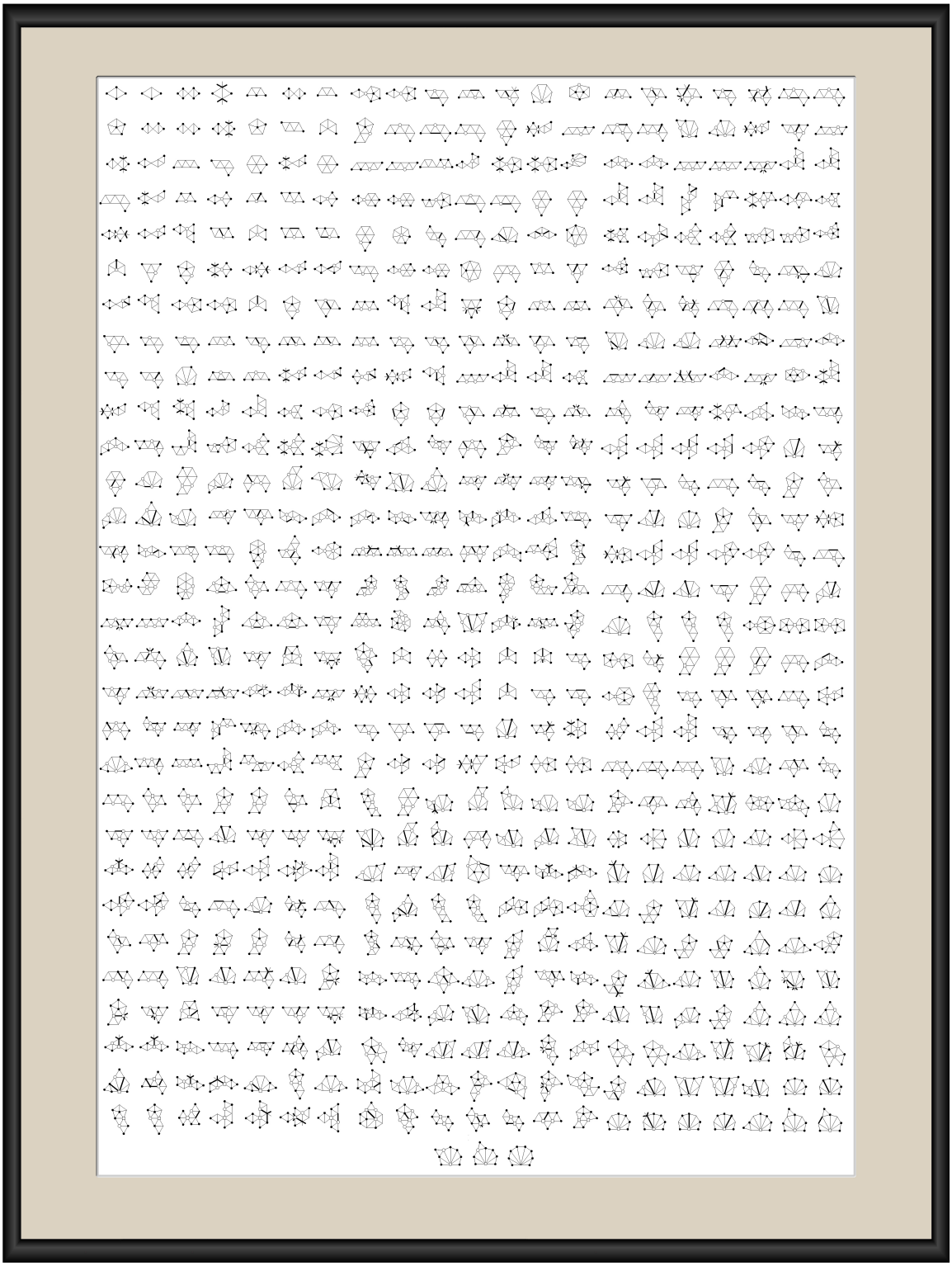
---

[44] In decimal notation, at least. Of course in effect we have already specified the number with a small finite number of symbols.

[45] Appel and Haken (K. Appel and W. Haken, "Every planar map is four colorable, Part I: discharging," *Illinois Journal of Mathematics*, **21** (1977) 429-490; K. Appel, W. Haken, and J. Koch, "Every planar map is four colorable, Part II: reducibility," *Illinois Journal of Mathematics*, **21** (1977) 491-567) considered more than 1900 configurations and more than 300 so-called discharging rules; the proof was so complicated that no one could simplify or even check it for twenty years. Finally Robertson, Sanders, Seymour, and Thomas (N. Robertson, D. Sanders, P.D. Seymour, and R. Thomas, "The four-colour theorem," *Journal of Combinatorial Theory*, Series B **70** (1997), 2-44) reduced its complexity to 633 configurations and 32 discharging rules — a simplification which allowed a complete proof to be written out and verified by computer. — An executive summary is provided by B. Bollobás, *Modern Graph Theory*. Berlin: Springer-Verlag, 1998; pp. 159-161.

[46] Specifically: that any map in the plane can be colored with no more than four colors in such a way that no two contiguous regions have the same color.

centuries the most famous unsolved problem in the subject — but rather the enumeration and systematic elimination of over a thousand separate cases, handled mechanically by a computer program and not, at least not immediately, understood directly by any human mathematician. This engendered a rather painful debate, and raised ugly questions: is there any guarantee a cleaner proof exists? are many unsolved propositions with simple statements destined to have similar resolutions? and so on. — Mathematics is supposed to be an elegant duel with light-sabers, not some kind of rude barbarian combat in which the victor clubs his opponent to death.



The 633 configurations of Robertson, Sanders, Seymour, and Thomas.

{...}

One might contrast the solution of the game of checkers, obtained by researchers at the University of Alberta; they examined 500,000,000,000,000,000,000 different configurations to show that there is a strategy for the game that does no worse than draw.[47] But there is nothing particularly shocking about this, because the rules of checkers are the product of a kind of caprice, and games of strategy in general have unbounded logical complexity;[48] in fact it's almost surprising it was this easy. — One would expect chess and Go to be solvable in similar fashion, though it is difficult to imagine that a computer could finish enumerating the cases before the heat death of the universe.

(It is instructive, incidentally, to consider the case of a human playing a machine at chess; the moves of the latter are completely determined by a set of algorithms; the moves of the former are not, and it is obvious no behaviorist ever considered the question of how they could be reduced to a finite set of conditioned reflexes[49] — this despite the fact that programming a computer to play chess was one of the first problems that occurred to the pioneers of artificial intelligence.)[50]

---

[47] Jonathan Schaeffer, Neil Burch, Yngvi Björnsson, Akihiro Kishimoto, Martin Müller, Robert Lake, Paul Lu, Steve Sutphen. "Checkers is solved." *Science* 14 September 2007: Vol. 317, Issue 5844, pp. 1518-1522. The program (Chinook) can be played online.

[48] As Ulam was fond of pointing out, questions about games of strategy nest quantifiers to arbitrary depth — the problem of chess, e.g., can be stated as whether for all opening moves by white there exists a move by black such that for all moves by white there exists a move by black such that, etc. — whereas in normal mathematics few definitions (Ulam's pet example was that of an almost periodic function) nest them more than four or five deep.

[49] Could operant conditioning be employed to teach a rat to play chess? — No? (Why not?) — What about tic-tac-toe? — If one rat can't be conditioned to play chess, can a roomful of them? Enquiring minds want to know.

[50] Turing himself wrote one of the first such programs; apparently to revenge himself upon his colleagues at Bletchley Park, who pissed him off by beating him so consistently.

{...}

But mathematics is supposed to be *necessary* truth. When a simple question has an enormously complicated answer[51] it *looks* like truth by accident.

In the case of Goldbach's conjecture results have been obtained which show that a related proposition holds for all numbers greater than an enormous lower bound; though thus far this lies far beyond the range of possible computation, it is conceivable that some combination of faster computers and improved lower bounds could make it possible to construct a complete proof by pasting together an analytic result (true for even numbers greater than some enormous N) and brute force enumeration of the rest of the cases (verified by explicit computation for even numbers less than or equal to N). If this were the case, it would present us with an example of a number-theoretic theorem about the integers, what we would like to think of as quintessential necessary truth, which would nonetheless have the appearance of being true only by accident. —Wittgenstein would have loved this, but no one else.[52]

(Obviously it is also disturbing that a proof based on a computer program depends on a proof that the program is correct; these in practice are practically impossible to provide, and, handwaving arguments about the probability of error being vanishingly small

---

[51] The usual situation goes the other way around — a complicated problem has a simple solution: the problem of the thirteen pennies, for example. [Not sure I can explain that without a diagram, and how are diagrams included in footnotes? Hmmmm.....]

[52] *Remarks on the Foundations of Mathematics*, III.42: "It might perhaps be said that the synthetic character of the propositions of mathematics appears most obviously in the unpredictable occurrence of the prime numbers. ... The distribution of primes would be an ideal example of what could be called synthetic a priori, for one can say that it is at any rate not discoverable by an analysis of the concept of a prime number." (Translated by G.E.M. Anscombe. Cambridge, M.I.T. Press, 1967.) — This sounds surprisingly Kantian, but there is something intuitively correct about it.

notwithstanding, it isn't immediately obvious that we haven't been presented with an infinite regress.)

{...}

Appended note:

The march of mathematical progress has now brought this scenario to fruition: the weak Goldbach conjecture, which states that that every odd number greater than 5 is the sum of three odd primes, had been proven by the refinement of analytical techniques due to Hardy, Littlewood, and Vinogradov, among others, to be true for all numbers greater than a bound C; a series of attempts to lower C had [2002] reduced it to about 10^1350, still far beyond the reach of computer verification. Recently, however, Helfgott[53] has lowered C to 10^27 and since computational efforts[54] have extended numerical verification nearly to 10^31, the proof-theoretic chimera has now been stitched together. — The question remains whether further refinements of these techniques can gradually reduce C to some value more satisfying to intuition: 10 certainly would work, but 100? 1000? 1000000? — Where to draw the line? — In the meantime, though the weak Goldbach conjecture is now known to be true, it falls into a kind of uncanny valley[55] between the analytic and the synthetic.

{...}

---

[53] H.A. Helfgott, "The ternary Goldbach conjecture is true"; arXiv:1312.7748v2, 17 January. 2014.

[54] These too rely on (partial) empirical verification of another open question, the Riemann hypothesis, for which enough zeroes have been computed to bound the gap between successive primes sufficiently well up to 10^27 that an odd prime can be subtracted from the triple to yield an even number less than the limit to which the even Goldbach conjecture has been verified, of the order of 10^18. Not to take anything away from Helfgott's remarkable achievement, this argument is a ridiculous kludge.

[55] A term used in computer graphics to designate the disturbing gap between the obviously phony and the photorealistic. Thus synthesized faces possess an unsettling quality.

A similar simple proposition about the primes no one has any idea how to prove is the twin prime conjecture: that there are an infinite number of pairs (p, p+2) which are both primes.[56] — About this Cohen after expressing skepticism regarding the ability of axiomatic frameworks to capture the properties of the mathematical objects they describe asks "Is it not very likely that, simply as a random set of numbers, the primes do satisfy the hypothesis, but there is no logical law that implies this?"[57]

In fact the natural metamathematical conjecture is that almost all[58] conjectures that appear to be true on probabilistic grounds are true but unprovable; i.e. that these two senses of "true, probably" and "probably true" are equivalent.[59]

The Goldbach example suggests that there may be many conjectures with relatively simple statements whose probability of truth is unity (since they can be shown to be true on the complement of a finite set) but which then are true or false globally by a kind of contingency, in that the proof can only be filled in by case by case enumeration. Indeed this situation may be typical.

{...}

To state one moral, then: philosophical intuition is not completely worthless, but like any other kind of intuition it is based on a kind of

---

[56] See (xix).2003.7.8, "Minor triumphs".

[57] Paul J. Cohen, "Skolem and pessimism about proof in mathematics", *Phil. Trans. R. Soc. A* (2005) **363**, 2407-2418. (12 September 2005.)

[58] "Almost all" has the technical definition "except on a set of measure zero" and doesn't really mean anything unless such a measure can be defined. Here it can.

[59] For reasons that may be obvious this occurred to me while meditating gloomily on a lecture about the abc conjecture.

experience; and it should not, therefore, be a surprise that its conclusions evolve when that experience broadens.

The analytic/synthetic distinction was introduced by Kant; was almost immediately questioned by Gauss, who had already understood the possibility of a non-Euclidean geometry; and then revised after radical extensions of the idea of entailment to include inferences like "7 + 5 = 12" and "a straight line is the shortest distance between two points", even though (as Kant pointed out) neither falls under the traditional definition of a conclusion being included in the premises.

Now, it becomes clear, it may be less a black and white distinction than a grayscale continuum, resolving under closer examination into an arbitrarily ramified hierarchies of the kind with which we have lately become familiar in complexity theory. — The more extensive our experience of what constitutes proof, the more baroque may our intuition of necessity become.

{...}

Fundamental misconceptions about mathematics and the nature of prediction notwithstanding, there was a larger fallacy involved in behaviorism: it was based upon an artificially limited, indeed an essentially invalid idea of what constituted science.

You could see it in the polemics Skinner's partisans wrote against Chomsky — here was a real theory of language at last, or at least a piece of one, and it was attacked as unscientific because it was (in Eddington's phrase) physics and not stamp collecting; because they not only did not recognize theory when they saw it, they did not understand its necessity — because they had trapped themselves in the most limited possible conception of empiricism, almost a throwback to Bacon, one in which scientific endeavor consisted entirely in the blind accumulation of disconnected "facts"; the reduction of the philosophy of nature to making statements in an

observation language — which, of course, they didn't even see was ill-defined.

I suppose this was natural. Psychology had spun its wheels from Hume to William James trying to found itself in introspection. A radical break seemed to be called for, what more comprehensive revolt against subjectivism than to deny the existence of the subjective entirely, and in so doing why not banish all "metaphysical" statements altogether? this was the spirit of the age, after all.

{...}

There was, in other words, a desperate anxiety among psychologists that what they were doing was not "science". And quite understandably they sought to make what they were doing "scientific" by imitating what they saw their intellectual elders doing: performing experiments in laboratories and making measurements that produced copious amounts of numerical "data" — publishing "results" in "papers" in "journals", filling them with graphs, tables, charts, and statistical analyses — *going through the motions* — hoping that, by performing the same ritual abasements as (real) biologists, chemists, and experimental physicists, psychologists could acquire their mojo. — There is a name for this, and it is not "scientific thinking".

{...}

Freud gives as examples of what Frazer called imitative or homeopathic magic the following:

> Rain is produced magically by imitating it or the clouds and storms which give rise to it, by 'playing at rain', one might almost say. In Japan, for instance, 'a party of Ainos will scatter water by means of sieves, while others will take a porringer, fit it up with sails and oars as if it were a boat, and then push or draw it about the village and gardens'. In the same way, the fertility of

> the earth is magically promoted by a dramatic representation of human intercourse...” and summarizes the principle as follows: “If I wish it to rain, I have only to do something that looks like rain or is reminiscent of rain.[60]

Later Feynman described the practice as follows:

> In the South Seas there is a cargo cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they’ve arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas — he’s the controller — and they wait for the airplanes to land. They’re doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn’t work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they’re missing something essential, because the planes don’t land.[61]

He did not, however, recognize that the cargo-cult phenomenon extends beyond pseudoscience into what is supposed to be “science” itself. — Behaviorism had a theatrical run of a couple of generations. But the planes never landed.

{...}

So that is one way of putting it: the cargo cult imitated the “behavior” of the operators perfectly, but didn’t look inside the radios to see what made them work. There must be a moral there.

---

[60] Sigmund Freud, *Totem and Taboo*, transl. James Strachey, London: Routledge Classics, 2001. Chapter 3, “Animism, Magic, and the Omnipotence of Thoughts”.

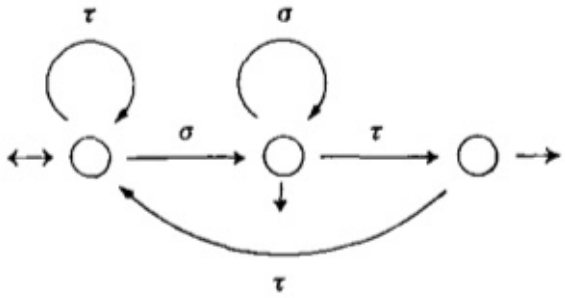
[61] *Surely You’re Joking, Mr. Feynman!* New York: W.W. Norton, 1985.

Another way of putting it is that no one ever said what "behavior" was. There was some vague appeal to observable physical states of the organism, but "observable" and "state" and "organism" weren't defined, and the "state" per se wasn't what was referenced in any case, rather some notion of "action", presumably definable in terms of a (short, finite) temporal sequence of states — though this wasn't defined either, of course. So for all anyone could tell "behavior" might include the response of the subject to cold in the form of goosebumps, or to ultraviolet light in the form of sunburn — note that the response in this case varies dramatically from one subject to another, and that no kind of input-output table relating insolation to degree of burning will say anything about the chemistry of melanin, the real causative factor — or the humidity of expelled breath, or height and weight, or for that matter what the subject said in response to the question "What are you thinking about?" — Behavior could have been *anything*, until it was defined. In fact simply by declaring it to have meant the microstructure of brain activity, to which real scientists more sensibly have turned their attention, the program could now be pronounced a success.

Again: a rigorous definition of "behavior" would entail definitions of "stimulus" and "response", and those in turn would require an enumeration of possible inputs and outputs. — Implicitly, as Chomsky pointed out, the mathematical model behind the smoke and mirrors here is that of a finite-state automaton, which takes as inputs strings of symbols selected from a finite alphabet; each one inducing a state transition for which, in turn, a string of symbols from a finite output alphabet is produced; this may be pictured, e.g., as a state transition graph with edges labelled by inputs, for instance:[62]

---

[62] Example 2.6 of Samuel Eilenberg, *Automata, Languages, and Machines*, Volume A. New York: Academic Press, 1974. — I'm cheating here, this is a slightly different type of automaton, but the principle is the same.



It would then be easy to define away the internal states of the machine as equivalence classes of maps from inputs to outputs, and a kind of behaviorist program can be said to have succeeded.

But where do the input and output alphabets *come from*? Some kind of language is presupposed to specify just what "observable behavior" is, and in practice there is that familiar philosophical bait-and-switch, the appeal to self-evidence, and a host of unexamined assumptions are insinuated by inclusion and omission. And so we have ring bell/salivate, shout/wince, electric shocks and bits of cheese, and not, say, observations of the form "I rebuked him, and observed that he took offense at the harshness of my manner of expression" — though why not, no one will ever bother to tell you. — "Measure something" — but why measure *this* and not *that*? — And the answer, of course, is that what is and is not relevant has been decided by an implicit appeal to an unstated theory, something beyond the reach of scrutiny. — Elsewhere this is styled "metaphysics".

Moreover in practice you have only a small subset of the input-output mapping, and have to guess the rest — another version of the problem of induction — and the most natural theoretical device employed to model it is, guess what, an internal state space whose transformations are induced by inputs — in the language of the electrical engineering lab, the wiring of the black box; for a given set of pairs {(input,

output)} there will be an ensemble of possible wirings, some kind of maximum-entropy probability distribution imposed upon it, and (ideally) an optimal set of yes/no experiments that will, in the limit, identify the correct internal configuration and thus determine the mapping.

But obviously this is too complicated for a psychologist to appreciate. It may be better to let them keep playing with the knobs on their empty boxes.

{...}

At any rate the simplest objection is still the most powerful: if the brain is a relatively trivial mechanism programmed by the conditioned reflex, then it shouldn't be difficult to reverse-engineer it, and build a model of one.[63] — So: Mechanical Turk; put up or shut up. — Of course this has turned out to be harder than it looked. And though admittedly the training procedure for neural networks bears a family resemblance to the process of conditioning, it is precisely that which to date has rendered it so incredibly inefficient.[64]

{...}

By way of general conclusion: though self-reproducing[65] living organisms are composed of cells, biochemical factories which contain a large but finite number[66] of kinds of molecular machines which function according to the laws of physics and chemistry — and one can, in principle, write down equations of motion for the dynamical

---

[63] Though of course: just because you can build it doesn't mean you can predict what it will do.

[64] Neural networks are trained on sample sets which number in millions, billions, or even trillions. Human infancy does not last a thousand years. Therefore, etc.

[65] Viruses are simpler, but must hijack the machinery of cellular organisms to reproduce.

[66] Counting genes, I would guess between ten and a hundred thousand. This is probably low.

system this ensemble represents — even in simplified form these would involve millions of variables, there is no sensible way in which one can suppose they could be solved, and in any case they are, strictly speaking, quantum-mechanical in nature and thus indeterministic; not that intrinsic thermal jiggle does not render the classical problem stochastic anyway. In consequence even when some kind of recognizably mechanical procedure is being implemented, in the operation of an enzyme, e.g., or the reproduction of a strand of DNA, nothing ever works the same way twice; not even the fabrication of the machines themselves. — Moreover this is not some kind of regrettable design flaw which would be eliminated in a more perfect world — as designed by Plato/Skinner/Minsky/... — this is precisely what made life possible in the first place. (It is also what renders biological design so robust.)

So in the sense that disturbs us — that mechanism is something which does the same thing the same way every time that it functions — biological machinery is not machinery at all. — Indeed to think that it is, or even that it ought to be, is simply insane. — Life is the product of evolution, and evolution consists precisely in making up rules in order to break them.

Perhaps we should call this the paradox of vitalism, then: that despite being wrong about everything it ends up winning most of the arguments anyway.

So even though there is a philosophical moral to be found here, as usual it looks like a joke.

{...}

Meno (1972)

> In the approximation of classical relativistic theory the creation of an electron pair (electron A, positron B) might be represented by the start of two world lines from the point of creation, 1. The world line of the positron will then continue until it annihilates another electron, C, at a world point 2. Between the times t1 and t2 there are then three world lines, before and after only one, However, the world lines of C, B, and A together form one continuous line albeit the "positron part" B of this continuous line is directed backwards in time. Following the charge rather than the particles corresponds to considering this continuous world line as a whole rather than breaking it into its parts.[67]

Plato in his *Meno* argues that knowledge is reminiscence: Socrates summons a slave boy ignorant of mathematics and starts cross-examining him about a geometrical demonstration; when the kid begins to evince understanding he claims this is evidence that what is known, or knowledge of abstractions, at least, is already resident in the soul; which must therefore be immortal.[68]

---

[67] R. P. Feynman, "The Theory of Positrons." *Physical Review* Volume 76, Number 6. September 15, 1949, pp. 749-759. — In his Nobel lecture [1965] he cheerfully admits to having stolen this idea from Wheeler.

[68] I think that insofar as Plato's argument works, here as always it actually shows that the soul lies outside of, is independent of time, that it is Being and not Becoming, and that if you pitch this in contemporary language most people would still buy it. Not that this has anything to do with playing harps in the Celestial choir; Plato is as usual prone to what appear to us to be purely verbal confusions like the conflation of "timeless" with "immortal." But I would point out, e.g., that though Windows 3.1 — no, too dreadful, I shouldn't say that, I should say the System 7 Mac OS — may not be running on anything at the moment, this doesn't mean it doesn't *exist*. — Though it hasn't *gone* anywhere either.— Compare also the Pythagorean doctrine of metempsychosis with (software) "installation": they are not precisely isomorphic, and indeed it's strange no one thought of the idea that more than one person may have been

It seems to me that this argument is more or less correct; though I don't think it says what Plato thought it did. — Certainly it has aged well; what Chomsky said about grammar and innate ideas was not very different.

Presumably Plato thought of this because learning something feels[69] more or less the same as remembering: you don't understand, and then you do. There is an abruptness to it, which he remarks; something that feels like the transition from lost to found.

Of course you wonder then about imagination — is this like remembering something that didn't happen? — but the real curiosity is invention, because this is exactly like remembering something you have yet to learn.

So you could with equal justice say this: if knowledge is reminiscence, then invention is remembering backwards in time.

{...}

Because Plato's argument doesn't have any direction to it. If you believed it, it would apply as easily to something no one has learned yet — to any possible result of mathematics, if not to any sort of contingent matter of fact: I could recover the memory of the proof of

---

who you were in a previous life; many distinct persons have the same physical ancestors, after all. (And do all bacteria have the same soul? really, we can do this all night —)

[69] It is amusing to try to come up with an explanation for this subjective feeling: learning involves an inductive computation which is much longer and more complex than it appears to the conscious mind; many processes go on in parallel to try to piece together a solution to the problem; when the result is presented to consciousness, the fragments of understanding are, by this time, things which are "already known," and the process of retrieving them to explain the whole is isomorphic to remembering; so when the conscious ego sees the answer it is, indeed, something already resident in the soul, albeit in a part not easily accessible to conscious inspection; the effect is functionally not dissimilar to retrieving memories from a past life.

the Birch/Swinnerton-Dyer conjecture, for instance; though presumably not of how I'd spend the prize money once I published it. — If the soul[70] is boundless and immortal and swims in the sea of eternal truth, time cannot apply to it.

(Leibniz remarks "every mind is of unbounded duration.")[71]

Admittedly this creates a problem with why you understand premises before conclusions: what kind of ordering is introduced by the arrow of logical inference? Is it the same as the arrow of time? — it doesn't appear to be, at least, since in a proof many premises may precede a conclusion, and the order among them is somewhat arbitrary, not necessarily linear. (Linearity is an artifact of exposition.)

This has something to do with the P/NP distinction, about which — ha! — more anon, but for the moment note that one way of stating that[72] is to say there is an inherent difference in difficulty between finding a proof and verifying one. — The latter is straightforward and leads from a leaf of the tree back to the root, which is linear; the former is a search, involves tracing a path from the root of the tree to the leaf that holds the solution, and is in general exponential.

In this sense the distinction between remembering and inventing is just which computation is harder. Time reversal is not a symmetry of the problem.

(Unlike quantum field theory: obviously this argument only occurred to me because I knew that Feynman identified positrons with

---

[70] "The soul" has several different meanings, and the one I take seriously (Aristotle's) is rather different from Plato's; let alone from what Catholic theology derived from it. But we're playing by Plato's rules here.

[71] Loemker, p. 160.

[72] See Jan Mycielski, "The meaning of the conjecture P≠NP for mathematical logic." *The American Mathematical Monthly*, **90** [1983], 129-130.

electrons running backward in time, but the situations are not isomorphic.)

I.e. in following a proof we have a series of applications of modus ponens: A, A —> B, therefore B. We write this down and it looks linear, but if we turn it upside-down the illusion evaporates: from B there are an arbitrarily large set of pairs B <— A, A to derive it from. Even when we can bound the number, as in a procedure like the construction of a semantic tableaux, the complexity of the search grows exponentially in length.

It is for some reason like this that you can remember where you came from but not usually where you are going. — Penrose has an elaborate argument about Fourier decompositions and the wave equation, but Patti Smith is more succinct: "I don't fuck much with the past but I fuck plenty with the future."

{...}

Personal immortality

The argument of the *Meno* is that the soul is independent of time. It says both that you always have existed and that you always will.

Young children seem to believe this instinctively. A Pythagorean belief in metempsychosis is as natural as primitive animism.

When I was a child my sister and I would address one another at the breakfast table: "When I was a bird, I used to go like this [making swooping motions with our forks]." — Of course isn't that a peculiar use of "when"? It seems to point not so much to past or future as to some location elsewhere in the manifold of possibility; somewhere sideways in time.

{…}

Meno postscript

The traditional conception of the immortality of the soul is one of an extended life: this world, *and then* the next; the linear continuation of personal identity by the accretion of memories, an uninterrupted thread. This seems strangely limited. One could attempt instead to imagine higher forms of consciousness — suppose, for instance, that one began with the original thread, the life-line with its beginning and end, and extended it in another dimension,[73] into a sort of ribbon; this might be a kind of extension into parallel worlds, but there could be other interpretations as well. — But more or less by definition this is beyond human comprehension.

Regarding it as Nietzsche did contemplating Goethe, a life taken as a whole might be regarded as a work of art — though if so one never completed but abandoned — as a kind of moment of apprehension in a larger consciousness, say; then one can ask, by analogy, what might *follow* that, and it would look more like a variation, an imitation, perhaps, or an annotated commentary, or an answer to the question "how else might it have been done?"

It probably isn't an accident that all this occurs to me while listening to Gould play the Eroica Variations.

---

[73] J.W. Dunne proposed a similar idea in *An Experiment With Time* [London: A & C Black, 1929], as a means of "explaining" the phenomenon, if it is one, of precognitive dreaming. For a while I took his stories seriously, but could never make sense of his explanation, which had something to do with his fascination with Minkowsi space.

{...}

The mental and the physical

It is not a *natural* isomorphism, in exactly the categorical sense: the functor that relates the mental and the physical is contravariant, and reverses all arrows.[74] Thus Kierkegaard said the tragedy of life was that it is lived forward but understood backward.[75]

{...}

This can be illustrated by the example with which Eilenberg/MacLane motivate the discussion in their original paper on category theory:[76] consider the class of vector spaces over a fixed base field F, together with the natural mappings between them, i.e. linear transformations

$$V \underset{f}{\rightarrow} W$$

which are closed under composition. For any vector space V, there is a dual space V* consisting of the linear mappings

$$V \underset{\phi}{\rightarrow} F$$

which is isomorphic to V, albeit in a way which depends on a choice of basis.

---

[74] This is not very different in spirit from what Russell said about the mental and the physical simply being different schemes for grouping relations, though it reflects changes in mathematical fashion.

[75] According to Susan Sontag, Walter Benjamin referred to memory as "reading oneself backward". Same thing. — The (already classic) cinematic expression of the thesis is Christopher Nolan's *Memento* [2000].

[76] Samuel Eilenberg and Saunders MacLane, "General Theory of Natural Equivalences", *Transactions of the American Mathematical Society*, Vol. 58, No. 2 (Sep., 1945), pp. 231- 294.

Then for any mapping

$$V \underset{f}{\rightarrow} W$$

there is an induced mapping

$$W^* \underset{f^*}{\rightarrow} V^*$$

defined by

$$f^*(\phi: W \rightarrow F) = \phi \cdot f: V \rightarrow F$$

A similar principle is at work in the relation of the mental and the physical: there is a mapping from the state of the world to the state of the mind at any moment; the natural development in the physical world is the dynamical evolution map from the state of the world at one time to the state of the world at a later time (in quantum mechanics this is, literally, a linear [unitary] operator mapping the Hilbert space of physical states onto itself); the corresponding induced mapping on mental states is the map from the perception of that state at the later time to the perception at the earlier time — i.e., memory.

{…}

Put another way: given two states of the world s, s', with s' later than s, causality provides a mapping from s to s'. If you are *examining* the states, however, the arrow goes the other way: given s', you *explain* what brought it about by reference to (remembered) s.

{…}

The problem of the relationship of mind to body is there already in the question of how a physical process can represent a computation; and *that* — really — is quite hard enough.

(Why it never ceases to amaze me that we can *build* computers. That these can be represented within physical systems.)

{...}

Awareness

Leibniz, "A Fragment on Dreams":[77]

> Sleep differs from waking in that when we are awake everything is directed, at least implicitly, toward an ultimate goal. But in dreaming there is no relation to the whole of things. Hence to wake up is nothing but to recollect oneself ... to begin to connect your present state to the rest of your life or with you yourself. Hence we have this criterion for distinguishing the experience of dreaming from that of being awake — we are certain of being awake only when we remember why we have come to our present position and condition and see the fitting connection of the things which are appearing to us, to each other, and to those that preceded. …
>
> [He remarks on visions in dreams, how they are superior to waking life in that respect, and concludes:]
>
> [Such visions] are sought by the waker; they offer themselves to the sleeper. There must therefore necessarily be some architectural and harmonious principle, I know not what, in our mind, which, when freed from separating ideas by judgment, turns to compounding them. A reason must be given why we do not remember waking experiences in a dream but do remember the dream when awake.

Here compare separating/compounding to serial/parallel.

---

[77] Loemker pp. 113-115.

When you awaken, your mind wanders in a kind of superposition. And then — abruptly — it observes its own state. The wave packet reduces. You are conscious.

I have to wonder whether Von Neumann had this in the back of his mind when he described the measurement process.[78]

[78] Cf. Chapter VI of *Mathematical Foundations of Quantum Mechanics*; Princeton: Princeton University Press, 1955.

{…}

Nietzsche says that consciousness it is a recent development of the organic and thus unfinished;[79] again, a kind of work in progress. — This however should be read as a gloss on the Socratic precept, to know yourself — to be able to clearly perceive and judge yourself — not simply to act without reflection; to be able to process feedback. — About which nonetheless Nietzsche, the champion of the development of the instincts — the prophet, as it were, of the unconscious — is extremely skeptical.[80]

(Oddly enough this reminds me of Whitehead's remark about inventions like algebraic notation, that advances in science largely consist of eliminating the necessity for thought, of reducing it to the automatic employment of technique; that this allows you to accomplish more with the expenditure of less mental energy.)

---

[79] Cf. *The Gay Science* #11.

[80] Thus for instance the extended rhapsody in *Ecce Homo* about the composition of *Zarathuʂtra*, how it had come to him in a burst of inspiration; as if dictated by a *∂aemon*. — "If one had the slightest residue of superstition left in one's system, one could hardly reject altogether the idea that one is merely incarnation, merely mouthpiece, merely a medium of overpowering forces." — Compare, of course, Rimbaud, "I is something else," but also Henri Poincaré [below].

{…}

Augustine in trying to resolve the paradox of the Trinity proposed an analogy with the tripartite nature of human identity:

> There are three things, all found in man himself, which I should like men to consider. They are far different from the Trinity, but I suggest them as a subject for mental exercise … . The three things are existence, knowledge, and will, for I can say that I am, I know, and I will. I am a being which knows and wills; I know both that I am and that I will; and I will both to be and to know. In these three — being, knowledge, and will — there is one inseparable life, one life, one mind, one essence; and therefore, although they are distinct from one another, the distinction does not separate them. This must be plain to anyone who has the ability to understand it. In fact he need not look beyond himself. Let him examine himself closely, take stock, and tell me what he finds.
>
> [*Confessions* XIII.11]

There is something to that, though in trying to relate mental phenomena to their evolutionary origins we might skip over Augustine and go straight back to Aristotle: any living thing has an identity, what he called the vegetable soul; some clearly have some kind of awareness, the animal soul; humans are self-conscious, and possess a rational soul.

This is not a continuum, exactly, nor is it a simple progression from simplicity to complexity; there are apparent phase transitions, something like the Chomsky hierarchy, but the ordering (this is so obvious that it has become cliché) is treelike, not linear. — One reason why any generalization about biology is probably wrong. —

There are universal principles which are manifested throughout, but this just means that pieces of functionality are scattered all about the animal and vegetable kingdoms. Plants, e.g., are supposed to be passive and insentient, animals capable of simple problem-solving but mainly relying on instinct, not reason. Nonetheless bread molds can solve mazes,[81] some birds really do have linguistic capabilities, and many animals not only have the ability to count but even seem to have a conception of zero as the cardinal of the empty set.

At any rate these appear to be distinct puzzles.

The question of identity applies to anything above the level of the individual cell. How is the organism defined? How do you draw a line around it? What is the predicate demanded by the comprehension principle?

On the one hand I want to say there is some kind of Brouwerian fixed-point theorem involved, in that a suitably-constrained system in which information is exchanged will be found to revolve, as it were, around some central axis; on the other I want to say, duh, the whole thing is grown from a strand of DNA, there is a code, a nucleus, defined physically, and the logical aspect follows from that, but it isn't simply genetic, there is the fact that, e.g., a plant can grow back from a

---

[81] See Merlin Sheldrake, *Entangled Life: How Fungi Make Our Worlds, Change Our Minds and Shape Our Futures*. London: Bodley Head, 2020. — The mechanisms are not completely understood, but fungi form extensive networks which can transfer information from one place to another, permitting a form of computation which is remarkably efficient: the mycologist Lynne Boddy made a model of Britain from soil, seeded fungi at the points corresponding to major cities, and grew a connecting mycelial network that looked just like the highway system; others have used slime molds to reproduce the subway system of Tokyo, suggesting novel applications of biological computation. — About this I remarked to a correspondent "Unfortunately fungi despite the advantages of massive parallelism have a slow clock and long cycle time, and probably can't be harnessed to crack RSA encryption. Too bad, I would love to see NSA having to learn how to cultivate slime molds."

root system, animals have central nervous systems to which signals are referred and from which they emanate. — A colony of cells is distributed, a multicellular organism has some kind of central processing. The line between them is fuzzy (Portuguese men-of-war) but whatever is emergent here, emerges rapidly; probably with the application of the principle of division of labor and its genetic correlate, the regulatory network. — At any rate there is a problem here.

The question of awareness is perplexing since it relates to the distinction between mental and physical. The old German-idealist way of looking at this invoked the distinction between noumenon and phenomenon, the difference between looking at things from the inside and from the outside. Bohr's analysis in terms of complementarity descends from that directly.

Russell's attempt to explain the distinction as one between different ways of grouping relations is similar, but has the added advantage that he saw clearly the need to formulate a kind of mathematical duality between the two viewpoints; noumenon and phenomenon should be something like Fourier transforms of one another. (Or, more generally, the set of all phenomena and the set of all noumena should be equivalent.)

Consciousness, however, is a distinct problem, and mainly has to do with the logic of memory. About that, Augustine again had the first and in some respects the definitive word.

{…}

Augustine

If there was a literary invention of self-consciousness, it was in the *Confessions* of Saint Augustine.

Who possessed an amazing philosophical acuity, really no one can touch him between the Greeks and Descartes; in the *Soliloquies*[82] he anticipated the Cartesian c*ogito* and the question of whether to be was to be perceived (he answered in the negative); in *De Trinitate*[83] he raises the problem of other minds, and gives the usual modern solution (analogy).

The philosophical passages in the *Confessions* are mainly in Books X (on memory) and XI (on the Creation and the problem of time). In brief they run as follows:

Animals have awareness [X.6] and process sensory input, but do not perceive meaning. Man, however, can *question* nature. — It is necessary, then [X.8] to go beyond the faculties of sense, the doorways into the soul, to consider memory,[84] the repository of what the senses have admitted. And this is rather mysterious:

> The power of the memory is prodigious.… It is a vast, immeasurable sanctuary. Who can plumb its depths? And yet it is a faculty of my soul. Although it is part of my nature, I cannot understand all that I am. This means, then, that the mind is too

---

[82] In *Augustine: Earlier Writings;* trans. J.H.S. Burleigh, Louisville: Westminster John Knox Press, 2006.

[83] *Augustine: On the Trinity,* transl. Stephen McKenna, Cambridge: Cambridge University Press, 2002.

[84] Augustine has an expansive interpretation of "memory", and later says [X.14] "the mind and the memory are one and the same."

> narrow to contain itself entirely. But where is that part of it which it does not itself contain? Is it somewhere outside itself and not within it? How, then, can it be part of it, if it is not contained in it?

Which states the problem of the unconscious. (Unsurprisingly, he does not resolve it.)

Also whenever he says anything about memory he always qualifies it by mentioning forgetfulness; the *fallibility* of memory puzzles him.

He draws the distinction between memory of events/sensations and functional memory, knowledge of how to do things; he notes [X.9] that besides the data of the senses memory receives and stores things like grammar. Trying to understand where the knowledge of conceptual principles comes from, he reinvents the theory of innate ideas [X.10]:

> How, then, did these facts get into my memory? Where did they come from? I do not know. When I learned them, I did not believe them with another man's mind. It was my own mind which recognized them and admitted that they were true. I entrusted them to my own mind as though it were a place of storage from which I could produce them at will. Therefore they must have been in my mind even before I learned them, though not present to my memory. Then whereabouts in my mind were they? How was it that I recognized them when they were mentioned and agreed that they were true? It must have been that they were already in my memory, hidden away in its deeper recesses, in so remote a part of it that I might not have been able to think of them at all, if some other person had not brought them to the fore by teaching me about them.

He notes among the properties of memory the ability to hold facts which are not present [X.11] but can nonetheless be retrieved; he

gives number [X.12] as an example of knowledge that does not originate in the senses. (No empiricist he.) — He remarks the oddity [X.13] that it is possible to retain both correct and incorrect logical arguments, and to review and compare them, and notes that one can remember a feeling, though this is not the same as experiencing it. (He seems here to have a keener appreciation here of the problematic character of the impression/idea distinction than Hume did.)

When he speaks of the sun [X.15], it is the sun's image he recalls, not the image of an image, but when he speaks of memory — ?! — how does that work? How can memory be present in itself except as an image of itself? — He is also [X.16] somewhat amazed that he can remember forgetfulness; this seems as though it should be self-negating. — He sees, i.e., the self-referential nature of the memory, he perceives this is the central problem. — More, he realizes "I have become a problem to myself" —

> I am not now investigating the tracts of the heavens, or measuring the distance of the stars, or trying to discover how the earth hangs in space. I am investigating myself, my memory, my mind. There is nothing strange in the fact that whatever is not myself is far from me. But what could be nearer to me than myself? Yet I do not understand the power of memory that is in myself, although without it I could not even speak of myself. What am I to say, when I am quite certain that I can remember forgetfulness? Am I to say that what 1 remember is not in my memory? Or am I to say that the reason why forgetfulness is in my memory is to prevent me from forgetting?

but rhapsodizes [X.17]

> The wide plains of my memory and its innumerable caverns and hollows are full beyond compute of countless things of all kinds. Material things are there by means of their images; knowledge is there of itself; emotions are there in the form of ideas or

> impressions of some kind, for the memory retains them even while the mind does not experience them, although whatever is in the memory must also be in the mind. My mind has the freedom of them all. I can glide from one to the other. I can probe deep into them and never find the end of them. This is the power of memory! This is the great force of life in living man… .

But then after wondering [X.19] how you can recover memories temporarily misplaced — what is the mechanism here? — he proceeds into a catalogue of the various ways the senses can admit temptation into the soul, displays predictable guilt over erotic dreams, etc., etc., and indulges in the absurdly amplified self-flagellation for which many generations of the heinously repressed have celebrated him.

Book XI is then concerned with the problem of creation *ex nihilo*: he dismisses the question of what God was doing *before* the Creation as a pseudoproblem, because it is only *within* the cosmos that time can be said to exist; rejects the Platonic picture of an artisan employing tools to shape the world from preexisting materials; and [XI.5] concludes (following the Gospels) that "you spoke and they were made. In your Word alone you created them." — Which is a very provocative suggestion of ontological sleight of hand.

Speaking, however, no matter whether any voice was heard, involves some kind of *motion*, which brings him back to the problem of time. — And here [XI.*7*] he makes remarks which seem to me to bear an uncanny resemblance to current cosmological speculation, though why, precisely, should be the subject of another postcard. — But settles on the question: how does the Word, which is timeless, somehow set in motion the engine of Becoming: "Yet the things which you create … do not all come into being at one and the same time, nor are they eternal."

God to Augustine has something like the absolute perspective we associate with Minkowski space [XI.13]: "Your years neither go nor

come … [they] are completely present to you all at once, because they are at a permanent standstill. They do not move on …they never pass at all. … Your today is eternity." — So [XI.14] what *ἰώ* time? and what has it got to do with the cosmos? The past is gone, the future isn't here yet, the present certainly exists but is always lapsing from existence. "In other words, we cannot rightly say that time *ἰώ*, except by reason of its impending state of *not being*."

Here [XI.15-16] he tries to figure out how time can be measured, a thankless chore before the invention of the clock.[85] Retreating from this question in confusion, he wonders again how the past and future can exist. The latter has been seen by prophets, the former by anyone who can describe it by examining the contents of his mind/memory. — Therefore, etc., QED? — No [X.20]:

> …it is abundantly clear that neither the future nor the past exist, and therefore it is not strictly correct to say that there are three times, past, present, and future. It might be correct to say that there are three times, a present of past things, a present of present things, and a present of future things. Some such different times do exist in the mind, but nowhere else that I can see. The present of past things is the memory; the present of present things is direct perception; and the present of future things is expectation.

Returning to the problem of measurement, he rejects the notion that time can be defined in terms of movement — quite the contrary, he suggests — in fact [XI.23] refutes a kind of argument from Mach's Principle, and concludes [XI.24] "Time … is not the movement of a body." He conjectures that one might attempt to measure duration by

---

[85] Compare, though it cannot be taken so seriously, the difficulty faced by a disciple of Lacan in defining self-consciousness before the invention of the mirror.

the lengths of syllables,[86] but gives up finally, and concedes [XI.26] "It seems to me, then, that time is merely an extension, though of what … I do not know. I begin to wonder whether it is an extension of the mind itself." And after further labor concludes

> It is in my own mind, then, that I measure time. I must not allow my mind to insist that time is something objective. I must not let it thwart me because of all the different notions and impressions that are lodged in it. I say that I measure time in my mind. For everything which happens leaves an impression on it, and this impression remains after the thing itself has ceased to be. It is the impression that I measure, since it is still present, not the thing itself, which makes the impression as it passes and then moves into the past. When I measure time it is this impression that I measure. Either, then, this is what time is, or else I do not measure time at all.

Finally he suggests that what moves through time (if anything does) is "the faculty of attention", and provides this vivid summary:

> Suppose that I am going to recite a psalm …. Before I begin, my faculty of expectation is engaged by the whole of it. But once I have begun, as much of the psalm as I have removed from the province of expectation and relegated to the past now engages my memory, and the scope of the action which I am performing is divided between the two faculties of memory and expectation, the one looking back to the part which I have already recited, the other looking forward to the part which I have still to recite. But my faculty of attention is present all the while, and through it passes what was the future in the process of becoming the past.

---

[86] This remained a problem well into the seventeenth century. Galileo, of course, referred to his pulse, but it is conjectured his musical training (he was a composer for the lute) gave him an enhanced sense of minute intervals, derived from the necessity of playing sixteenth and thirty-second notes in time.

> As the process continues, the province of memory is extended in proportion as that of expectation is reduced, until the whole of my expectation is absorbed. This happens when I have finished my recitation and it has all passed into the province of memory.
>
> {…}

To summarize: consciousness presupposes memory; memory is paradoxically self-referential; time involves similar paradoxes; the flow of time is an artifact of the function of memory.

All this is completely correct.

{…}

The intention here may not be obvious. This isn't a hermeneutic exercise. I'm not trying to *interpret* Augustine — learn his language, explicate his precise meaning, understand him on his own terms. What I'm trying to make clear is that if he spoke *my* language and knew what I know, he'd be saying the same things that I am. — Only better, of course. (And in Latin.)

{...}

So: Augustine identified the problem of consciousness with that of memory, and related both to the problem of the nature of time.

Obviously he was right about everything, and the appropriate place to begin is with an idea of Thomas Gold.[87]

Gold wished to understand how temporal succession arises in theoretical physics, and suggested that one might look at it as follows: suppose you have a set of index cards, and on each you write a description of the instantaneous state of the world. — Then you shuffle them. — How do you put them back in order? Where in the data on the cards do you find the arrow of time?

In physics, at least, it is more natural to think of this as a movie: you cut it into individual frames, snapshots of an action; scramble them; and then try to reassemble them in the proper order. — Of course this is the problem of montage, the Russian school addressed this with the famous Kuleshov experiments, and taken as a question of cinema there is not a single unambiguous answer: different orders have different meanings, and the arrangement is a matter of art. — Taken as a question of physics, it is problematic, but less ambiguous — in the usual example of frying an egg, there's no question what order the snapshots should assume, the difficulty is in explaining why that order is correct. Nonetheless under the usual laws of mechanics the problem is solvable.

---

[87] See *The Nature of Time*, Thomas Gold, ed. Ithaca: Cornell University Press, 1967. This volume summarizes the proceedings of a symposium organized by Gold, and contains not only papers presented at the meeting but the very animated discussions which followed them. Since Gold did his best to invite all of the world's leading theoreticians, these discussions are very interesting indeed.

Taken as a question of conscious experience, however, it is trivial. Mental states *refer* to one another, and that defines a natural temporal direction, provided by memory.[88] (Which is, physically, based on irreversible processes which have the thermodynamic arrow of time built into them; allowing us to cheat, and skip over the difficulties of the problem as it presented itself to Gold.)

{...}

I wouldn't say that consciousness is prior to temporal order, or vice versa; rather that the two ideas are inseparable.

I don't mean to say that consciousness is the missing link that determines the direction of time; or not that exactly. — What I mean, mainly, is that though we can imagine physical worlds in which time does not have a direction, these are not worlds in which consciousness is possible. Consciousness requires memory, and memory requires [provides] direction.

How it is that when you remember two things, you remember — or take it for granted you *ought* to be able to remember — which one came first. Because if memory were perfect, you would remember that when one occurred you remembered the other had. — Here as always we are talking about an ideal mathematical model of the phenomenon.

So the most elementary conception of the way that consciousness registers/sorts/imposes temporal order is exemplified by the ordinal numbers.

---

[88] Of course if memory *cannot* be relied upon, the questions of order and meaning again become problematic. This is exactly the problem faced by Guy Pearce with his Polaroids in *Memento* [Christopher Nolan, 1999]. As the authors were undoubtedly aware.

Von Neumann defined an ordinal as the set of all preceding ordinals, grounding the series in the empty set.

Literally, the sequence is generated by bracketing:[89]

$$\emptyset,\ \{\emptyset\},\ \{\emptyset,\{\emptyset\}\},\ \{\emptyset,\ \{\emptyset\},\ \{\emptyset,\{\emptyset\}\},\ \ldots$$

so that **m** < **n** if and only if **m** is contained in **n** if and only if **m** is an element of **n**, and each ordinal *contains* its predecessors. — This is it exactly.

{…}

The infinite descending epsilon-chain is a major no-no of set theory, of course, and though it's obvious how we avoid this in the case of the ordinals, in the case of conscious states it requires postulating a first memory.

The problem with that is not that it doesn't exist, but that it is usually an isolated incident which is not an element of a connected progression, as the model requires. — Memory functions erratically in early childhood, consciousness in consequence takes a while to turn itself on.

So this is the other thing we have to say about this model: it represents an ideal limit, something that is only approximated by the phenomenon as we observe it in the physical world, and thus in the mental world that we inhabit. Consciousness as we possess it is

---

[89] This may, by some coincidence, be more or less the sense in which Husserl used the notion, but of course I'll never bother to find out. Let's simply give credit where credit is due: consciousness is, indeed, consciousness-of (and thus self-consciousness is consciousness-of consciousness itself); and "bracketing" isn't a bad way to describe the elementary mental operation.

fragmentary and imperfect, because memory is fragmentary and imperfect.

{..}

Perhaps the best representation of the idea in the cinema is the famous shot in *Citizen Kane* of Kane/Welles walking down a corridor in his mansion and passing between two long mirrors, which produce an infinite series of reflections of reflections of a myriad of Kanes, stretching off into obscurity.[90] — In reality, with mirrors which are not perfectly reflective, the images will grow fainter as they recede in the apparent distance. So also memory dims — though not, really, in the same way, it is not an analog copying process, like making a copy of a tape; it involves some kind of digital representation, as if there were some sort of sentence recorded at every instant of awareness that contains a summary not only of what is transpiring at that instant but also a thumbnail description of what preceded that — or rather: of *what you thought* about what preceded that. — Else the past would fade into an undifferentiated blur. Instead we have a sort of album, with photographs and jottings and guitar licks and (Proust) odors, in my case probably of farts.

But in the moment, in the experience of awareness, what connects one moment to what preceded it is just that: the near-perfect retention of the previous thought, the present thought, and the perception that the latter *contains* the former; *is about* the former; *refers to* the former. — At bottom the relation is that of metalanguage to language; there is a process of *evaluation* going on.[91] — To represent it naively, supposing I

---

[90] It is also a metaphor for the multiple-perspective insect-eye view the film creates of Kane, an intimation of the now-ubiquitous idea of the multiverse, and much else besides. But I'm doing this one first.

[91] Another purely logical way to provide an arrow for time is the one implicit in Kripke's semantics for intuitionistic logic:, which can be interpreted as referring to a set of states of knowledge: you picture a set of elementary propositions, and a series of partial valuations

have just formed the internal sentence "The coffee cup is on the table to the left of the laptop," the present sentence is something like "I am reaching for the coffee cup with my left hand, having asserted that 'The coffee cup is on the table to the left of the laptop.'"

(With which I do so. — Cold. — Into the microwave again.)

{…}

In reality the mechanisms that create and store memories are (out of physical necessity) complex, and employ representations whose nature and functioning no one as yet understands. — This has to be the case, because, for instance, if we take the ordinal representation literally then the amount of storage required to keep your experiences straight would grow exponentially with time. This isn't possible, of course. — In fact even a linear growth is impractical; this is the point of Borges' story "Funes el memorioso". One cannot absorb the past if it is so vividly present that you are still living in it.[92] — A back-of-the-envelope calculation shows that the brain would already have popped like a balloon in infancy if it retained everything; a ruthless process of data compression which proceeds continuously is an engineering requirement. This entails the necessity of generalization — you have no room for the heap of facts, you must summarize them succinctly or

---

whose domains are ordered by inclusion. The problem of the sea-battle, for instance, is resolved by noting that the truth or falsehood of the statement belongs to an extension which has "not yet" been performed, because (the tautological formulation) "the future" is simply what you don't know about — that to which a valuation has not yet been assigned. — Or: "past" = "that to which a valuation has been assigned"; "future" = "that to which a valuation has not been assigned"; "present" = "the process of valuation. Another expression of the principle that information *by definition* always propagates forward in time.

[92] The problem would be solved if the longer you lived, the faster you thought, i.e. if at a minimum the speed of thought increased linearly as a function of time. Of course if anything the opposite is the case.

forget them entirely — thus induction — and the sketchy and imperfect character of any recollection.[93]

But that's real life, and this is more like mathematics. The question to be addressed by speculation is whether the ideal of consciousness is attainable; whether it can be physically realized. — What is the end point of mental evolution? Because it obviously isn't me.

In fact that's the most obvious thing about consciousness: that you haven't attained it. That you are always stumbling around half-awake, never really aware of who you are and what you are doing. That it is all out of control, that you have no grip on it.

{...}

The feeling of not being on top of things, of reeling out of control — what Heidegger called *Geworfenheit*, the sense of having been *thrown* into the world[94] — is the feeling of never being able to catch up; the incompleteness of consciousness is at the root of that. — If it were possible to retain everything and perceive it all at once — if we were fully conscious — complete understanding might be possible. — But it isn't, and it isn't. At best we perceive things in flashes, and perceive only fragments.— Schopenhauer said philosophy begins upon a minor chord; in truth the notes sound indistinctly, and never all at once.

---

[93] Of course this may be advantageous, for the same reasons that the imperfection of DNA transcription is a fundamentally creative principle. — Nietzsche: "The artist needs the infidelity of memory in order not to copy but to transform nature."

[94] Oddly enough — or not — this aspect of the human condition found its perfect cinematic expression in a silly scifi action movie — *Predators* [Nimród Antal, 2010] — which commences abruptly *in medias res* with an abducted mercenary warrior (Adrien Brody) regaining his senses, confused and disoriented, to find himself plummeting through the atmosphere of an unknown planet, toward a surface where he is about to be hunted by — well, you guessed it —

The imperfection of memory guarantees that you are never really yourself.

The imperfection of memory: this is why you always have the feeling that you aren't all quite there; that you are out of control, somehow, off balance and reeling forward against your will — *losing track* of yourself. You want to say, get a grip, but you never quite can: the present is *there*, but slips away and fades into the past before you can grasp it. Because there is this mathematical ideal, that is what defines you, and you are only an imperfect realization of it. — It is a kind of broken symmetry. You never attain the Platonic ideal.

A shape not seen directly; its shadow only.

Lived experience is not like a movie that you can pause, and step forward and backward. But it *ought* to be.

The principle is no different (in kind, anyway) from the limited realization of the Turing machine: a physical computer does not have an infinite supply of scratch paper. — But it *ought* to, and we reason about it as if it did. — This is one of those instances in which the difference between large finite models and the infinite limit is an annoying distraction from the real principles at issue.

(We do have to leave open the possibility, however, that there's some subtlety here that can only be addressed by taking these limitations into account; as the theory of computation develops another dimension when you consider constraints of time and space.)

{…}

The definition of an ordinal is inductive, i.e.

(i) Ø is an ordinal.

(ii) If **x** is an ordinal, then the union of **x** and {**x**} is an ordinal.

Which constructs the ordinals as Peano did the integers, from the bottom up. One can invert the procedure, and work from the top down; this is the variant which in programming is called recursion.

At first glance it looks like a flagrant violation of the vicious circle principle, since it defines a concept in terms of itself.

I.e. the factorial function[95] on a positive integer argument[96]

$$n! \equiv n \cdot (n-1)(n-2) \cdots 1$$

so that, e.g.,

$$5! = 120$$

can instead be specified (in Lisp notation) as

```
(defun factorial (n)
   (if (= n 1) 1
      (* n (factorial (- n 1)))))
```

[95] Dana Scott somewhere remarks that the factorial is the most overdefined function of all time.

[96] Euler, no disciple of Wittgenstein, extended the definition from positive integers to any real or complex argument; in this reinvention it is referred to as the gamma function.

which can be read as "the factorial is defined as the function of n such that if n equals 1, it is 1, otherwise it equals n times the value of the factorial for n minus 1."[97]

Similarly the Fibonacci sequence

$$1,1,2,3,5,8,13,21,34, \ldots$$

can be defined as the function

$$f(1) = 1$$
$$f(2) = 1$$
$$f(n) = f(n-1) + f(n-2)$$

[97] It is assumed that the arguments are positive integers.

{…}

Each conscious moment *contains* the one that preceded it.

Linguistically, each statement is *about* the last. The tower of inclusion is metalinguistic. — Viewed subjectively "now" is the assignment of a truth value.

(Why fucking with temporal order leads you straight into the Cretan liar.)

Consciousness and recursion.

(Here as everywhere we owe the first recognition of the importance of the idea to Kurt Gödel.)

Since I lacked access to computers until I could actually buy one, I came to coding as a pastime relatively late in life, and thus wrote my first recursive function call when I was trying to draw a picture of a (self-similar) tree.

For a simple binary tree, the algorithm is as follows: you define a function which takes as inputs a length, a minimum length, a multiplier less than one, a starting-point, a direction, and an angle; if the length is less than the minimum length, it returns without drawing anything (this is the cutoff that prevents the program from running forever); otherwise it draws a line segment from the starting point at an angle left by the given angle from the input direction and calls itself with new inputs the length times the multiplier, the minimum again, the multiplier again, the starting-point the endpoint of the segment just drawn, the direction that of the segment just drawn, and the given angle; when the function called to the left returns from execution it

repeats the procedure on the right. Thus the function calls itself not once but twice, and the result of its execution is to draw a tree which (modulo the asymmetry introduced by the finite cutoff) draws a tree which consists of a line segment to which are attached two isomorphic copies of itself, one angled to the left by a fixed amount, one angled to the right. — The ungainliness of this verbal description explains the usefulness of programming languages, which are simultaneously simpler and more precise than English prose.

At any rate the first time I did this it was peculiarly difficult (the fact that I was trying to do it in a mediocre implementation of a bad language[98] didn't make it any easier), and when I reached the stage in the code where it committed the sin of self-reference and *called itself* — not once but twice — I felt a peculiar vertigo, as if my eyeballs were trying to turn around and look back into my head.

And of course as a veteran of philosophical perplexity I recognized this instantly! as precisely the disorientation I had always felt when I tried to think about the problem of consciousness, to think about thinking, exacerbated enormously when later I began to worry about the foundations of mathematics — can you reason about logic? make mathematical models of mathematics?

Because *it is all the same problem*.

What is my argument here? — *It gave me the same headache*. — Oddly enough though it is more than a trifle silly it is the most powerful argument of all.

Consciousness is consciousness-of; self-consciousness is consciousness of being conscious.

[98] Microsoft Basic, on my first Mac.

Descartes based his existence not so much on the fact that he thought, but on the fact he was *aware* that he was thinking.

The sense of self is that of the subject as object.

The degree of consciousness is indexed by the efficiency of the implementation of recursion. (Depth of stack is one determinant. — How far short we fall of the "potentially infinite".))

There is (the ideal limit) some perfect mathematical model, which would at the least require perfect eidetic memory. And then the actual realization of the idea, which is erratic, and works by fits and starts.

Something in this captures the difficulty of the mental: there is a dual foundation for the world, but it has been obscured by a kind of broken symmetry. (Of course the material also has its limitations — in effect in quantization we posit the Hamiltonian picture as an unrealizable ideal as well.)

The old idea was "partial representation". Like the realization of the Platonic Idea of the circle.

Not unrelated: the denial by Hume and Nietzsche, among others, of the reality of the Ego. The argument is valid as far as it goes — there is reference to an ideal object which does not "really" exist — but does not grasp the simplification made possible for, say, a geometry, when ideal points are adjoined. — It is something like what happens when you stare too closely at a set of colored squares — they are just pixels, after all — but if you move back and take the set in as a whole, it resolves into a picture. You see a human face.

In principle you leave a physical record behind you, in the form of a series of patterns of neuronal excitation. (Say.) But all this is like the print on paper, there is something it points to; something that it means.

{…}

What is conscious is what enters into memory. — That's simple enough. That is what Augustine said.

A bit *too* simple, as the overwhelming evidence for ongoing unconscious labor shows. But something like this must be true nonetheless.

There is a kind of SIMP model that naturally imposes itself. A function is being evaluated. It depends on many variables, each the result of a functional evaluation. — And so on, recursively. — The separate subevaluations can be performed in parallel; and surely are, all the evidence suggests it.

But what defines the fixed point, the center? — The traditional objections to the Ego don't address this problem at all, why the mass of thoughts should even *appear* to be directed toward a center. — Why there is one voice in your head that is the loudest.

E.g. Nietzsche (*Late Notebooks* 34[123]): "man is a multiplicity of forces which stand in an order of rank… . All these living beings must be related in kind, otherwise they could not serve and obey one another … . The concept of the 'individual' is false. … the center of gravity is something changeable; the continual generation of cells, etc., produces a continual change in the number of these beings. And mere addition is no use at all. Our arithmetic is too crude for these relations, and is only an arithmetic of single elements." — Which actually contradicts itself: the equilibrium may not be static, but there is some kind of "order of rank" defined by the logic of the dynamics. — But he is quite acute on these questions: [24] "The logic of our

conscious thinking is only a crude and facilitated form of the thinking needed by our organism, indeed by the particular organs of our organism. For example, a thinking-at-the-same-time is needed of which we have hardly an inkling."

Later he refers to "the development of consciousness as an apparatus of government"; almost too cute —

{…}

But really the arguments against personal identity are the same as the ones that would try to persuade you that the rotation group in three dimensions does not "exist" either; for all that you need to study its representations to understand the quantum mechanics of angular momentum.

Still, why is there only one center of attraction? It seems like a fixed point theorem, something like the reason when you stir the cream in a coffee cup there's a single point at the center of the vortex that isn't turning.

A bundle of functions which are calling one another. The arrows of direction converging on a single one to which all the results are returned.

{…}

A related mystery: the evolution of arithmetic.

Where does counting come from? how did it arise?

The operation of abstraction. In the lambda calculus this is simply function abstraction, lambda: to take the functional form as an object in itself.

Something like this is made possible by the reduction of memories to data; by being able to examine these like sense data.

Perceiving the similarity of the five cows to the five trees. Perceiving the one-to-one correspondence.

But how would you *invent* counting? There must be some evolutionary pathway that leads to it. How could you discover it?

{…}

The imperfection of consciousness is the imperfection of memory. — In which one must include the imperfection of the mechanisms that retrieve and examine memories, but — I have listened to the Haydn symphonies dozens of times, so that anything I hear sounds familiar, but I have nothing like the memory of Wittgenstein, who could whistle through his favorite passages from Beethoven and analyze them as he went; and Mozart was undoubtedly better, more completely the master of everything he had ever heard. — But could either do calculations in his head, like Euler? and though my usual example is Von Neumann, I am usually thinking "if only I could keep track of everything like that" but he wasn't particularly musical, and even his mathematical intelligence was curiously limited, he had very little geometric intuition, for instance. — So it seems very obvious that there has never been a fully conscious human individual. — Save in the imagination of Borges, who could see the way that this would be an affliction.

{...}

Matters of degree: my dogs are conscious. (How do I know this? because they can be *unconscious*; I have often watched them dreaming.) But do they inspect their internal states in the same way I do? — Yes, but not as well. I have a larger brain, a better memory, and (in the use of language) employ a better shorthand for summarizing previous

experience. — If there is a phase transition here, it is in whatever makes possible the use of language — in symbolic representation. — But basically our feeling is that *Natura non facit saltus*; and even when she does, she approximates the jumps by continuous functions.

Though dogs are not as intelligent as birds in some respects — and so on — and so on. — The picture (Dawkins?) of a multiplicity of functionalities, developed and distributed among a multitude of species.

{…}

Some Cartesian I am if I don't revert to paranoia, however. — Given the dependence of consciousness on the chain of memory, then — what if it's fake? — Of course this is the entire oeuvre of Philip K. Dick.

What you have to wonder here is whether it really makes a difference. Superficially you have a kind of grue paradox applying to the past, at every step in the historical record you can insert a discontinuity, but isn't it Dick's point that fiction here has the same ontological weight as reality? that this is where the two interpenetrate? — The replicants in *Blade Runner* with their fetishistic reliance on their collections of yellowed photographs. — Rachael remembers a mama spider eaten by her babies, and this is "really" a memory of Tyrell's niece. — But is it? doesn't it define Rachael just as well as it did its originator? Doesn't the shared memory entail a shared identity? an overlap between two persons?

I don't feel like chasing Dick down this particular rabbit-hole at the moment, but my basic feeling, that the anxiety you are supposed to eel here is unfounded, remains the same.

{…}

The ∂eck of car∂s (again)

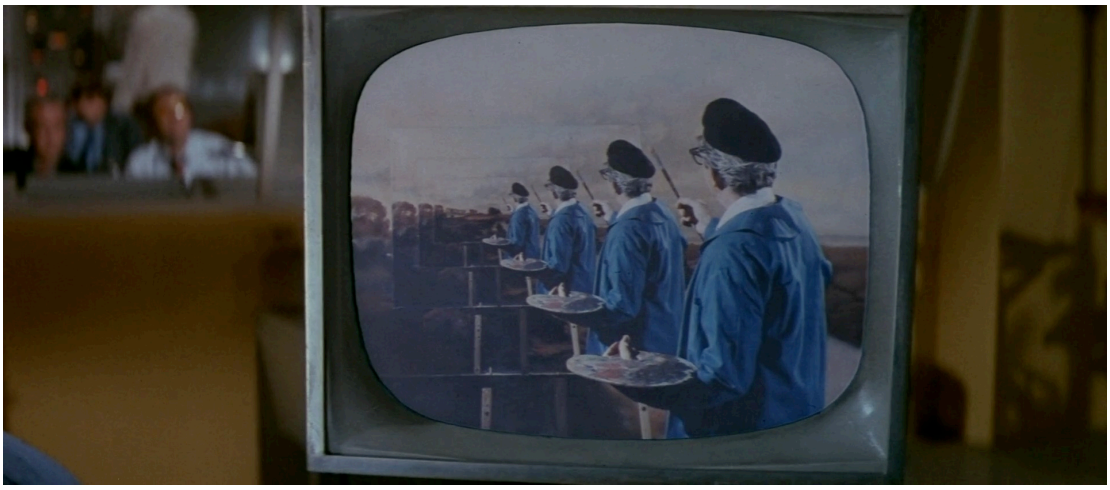
A natural way to look at it is to think of a strip of film, representing a series of snapshots taken along a four-dimensional world line. (Life is one long tracking shot.) — Consciousness (assumed to be external) is then a light illuminating the individual frame (the present) — the bulb in the projector, what throws the picture onto the screen: "Attention is *here*."

This is incorrect. The light — the act of projection — is superfluous. The idea of order emerges from the way the frames refer to one another; from the way they fit into one another. This putting-in-order is *itself* consciousness. — Which is not outside the picture, but part of it.

{…}

In the spirit of recalling all my favorite episodes of cognitive dissonance, here I must interpolate the observation that, during the period when I was first thinking about all this, I was flabbergasted to discover an exposition of the essential idea in a ridiculous scifi movie I happened across on the late show — namely *Escape from the Planet of the Apes* [Don Taylor, 1971]. — In this, after it is realized that the ape astronauts who have landed on the Earth didn't come from another planet, but rather our own future, an industry-standard Scientist With A German Accent ("Dr. Otto Hasslein") is dispatched to the television studio to explain the nature of time to the national audience. — "I think that time can only be fully understood by an observer with the godlike gift of infinite regression," he says. And offers as an illustration a painter at work upon a landscape, who realizes the picture is not complete unless he includes himself in it; painting a picture of himself painting a picture of, etc. — As an explanation of how consciousness constructs the temporal ordering, this is essentially correct.[99]

[99] This episode in the Apes franchise was written by Paul Dehn, noted mainly for screen adaptations like *Goldfinger* and *The Spy Who Came In From The Cold*. How inspiration possessed him on this occasion one can only guess. Maybe he was a neighbor of Fred Hoyle's.



{…}

This isn't quite it, of course. The fundamental paradox is one that Francis Crick used to emphasize, one exhibited in another (better) silly movie, Woody Allen's *Everything You Always Wanted to Know About Sex But Were Afraid To Ask* [1972]. In this there is a scene in which, while a guy is trying to achieve success in a sexual encounter, a crew of operators in something that looks suspiciously like NASA Mission Control frantically shout orders and throw levers and switches in the control room of his brain. — This, Crick remarks, is a very natural picture, one which is automatically assumed: there is a kind of nerve center to which all the messages are relayed, and someone is sitting there watching all the monitors — the master of the (inverse) Panopticon, see Fritz Lang in *The 1000 Eyes of Dr. Mabuse*, *The Matrix*, etc. — issuing commands. — Yes, asks Crick; but *who* is sitting there? — Well, *you* are; there *is* a regress. But in the nested chain of control rooms (the inclusion relation runs the other way in this picture) each one contains the *next* one, the one corresponding to the next instant; just as the definition of the factorial refers to its value at the preceding integer. A new one is created, as it were, at every step; (n + 1) within n.

It's possible to play games with a video camera and a monitor that illustrate the paradox directly, as Hofstadter pointed out.[100] — Still photographs don't really do it justice: when you aim the camera at the screen that is displaying its output, you get the infinite descending chain of pictures within pictures, sure, but they can be skewed with respect to one another, which is disorienting, and the dynamical effects produced by *turning* the camera are uncanny. — Most disconcerting of all, I discovered, was to employ the automatic zoom. The coordinated creep of the entire visual field is the only thing I have

[100] See Douglas Hofstadter, *Gödel, Escher, Bach*. [New York: Basic Books, 1979.]

ever seen, cold sober, that reproduces the effects of LSD: the famous Hitchcock trick[101] on steroids.

[101] The effect produced by simultaneously moving the camera toward an object while zooming out to compensate, thus optically amplifying the sense of depth in the field of view. First used in *Vertigo,* and imitated a thousand times thereafter.

{…}

The consistency of visual metaphor

The representation of time as a spiral, or a concentric set of circles, like tree rings [*Vertigo*] — is the self-similar: that which can be mapped into itself one-to-one.

(In three dimensions the onion and its layers.)

{…}

Are other models possible? Yes, certainly. You can see the hint of one in the final confrontation in *Fight Club* [David Fincher, 1999], when Ed Norton shouts "You're just a voice in my head!" and Brad Pitt replies calmly, "And you're just a voice in mine."[102] Because there is no *logical* objection to the idea of two control centers, each nested within the other.

It is easy to write something like this in Lisp, e.g.:

```
(defun even (n)
   (if (= n 0) 0
      (+ 1 (odd n))))

(defun odd (n)
   (+ 1 (even (- n 1))))
```

yielding

```
?     (odd 20)
      39
?     (even 20)
      40
```

and so on. — One or the other of the recursions must be grounded, in other words, but it is arbitrary which. Nor is there any difficulty in generalizing this to any finite number. The real question is how the separate identities could be distinguished, if each has access to the memories of the other, and in fact multiple personality disorders, insofar as they cannot be consigned to urban legend, appear to involve

[102] This exchange occurs in the movie, but not the novel. Presumably it should be attributed to Fincher or the screenwriter, Jim Uhls, and not to Mr. Palahniuk.

a partitioning of memory and consciousness, which mathematically would correspond to the partitioning of a graph into disconnected components. The idea that alternating personalities might, e.g., see the world differently, and (adopting Russell's simile) switch back and forth from red- to blue-tinted spectacles, is a trifle more baroque, and a little more difficult to enter into, if not necessarily to understand.

{...}

Hegel in the *Phenomenology of Spirit* #80 characterizes consciousness as "explicitly the *Notion* of itself";[103] which I have to admit is the right idea, even though there is very little else in this work that makes any sense. — Perhaps #177: "A self-consciousness exists *for a self-consciousness*" — ? — At any rate he appears to say that consciousness of self presupposes acknowledgement by another consciousness, something like the mirror principle, perhaps something like the principle I intuited on my tricycle.

But this is Hegel, of course, so who the fuck knows.

[103] Translated by A.V. Miller. Oxford: Oxford University Press, 1977.

{…}

In Shakespeare, the play that is *about* self-consciousness, *Hamlet*,[104] has as the midpoint and summit of its action a play *within* the play, designed by the protagonist to model the action of the play itself. This says everything.

[104] Bloom refers to Hamlet as "the hero of self-consciousness."

{…}

Hegel was right that consciousness has a history, that it is a work in progress, something in development. I doubt, however, that he realized it has been evolving for a couple of billion years, and that nothing about it is unique to *Homo sapiens*.

Because though on the one hand you suspect some kind of phase transition to consciousness in the higher mammals (perhaps also invertebrates, see the curious case of the octopus), on the other it's obvious that this is a biological invariant, coded in at the lowest level:

> A bacterium is so small that its sensors alone can give it no indication of the direction that a good or bad chemical is coming from. To overcome this problem, the bacterium uses time to help it deal with space. The cell is not interested in how much of a chemical is present at any given moment, but rather in whether that concentration is increasing or decreasing. After all, if the cell swam in a straight line simply because the concentration of a desirable chemical was high, it might travel away from chemical nirvana, not toward it, depending on the direction it's pointing. The bacterium solves this problem in an ingenious manner: as it senses its world, one mechanism registers what conditions are like right now, and another records how things were a few moments ago. The bacterium will swim in a straight line as long as the chemicals it senses seem *better* now than those it sensed a moment ago.[105]

This must be nearly the simplest possible implementation of the sense of time: two states, present and past, and execution of a single command ("change direction") conditional on the computation of a

---

[105] Peter Godfrey-Smith, *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. New York: Farrar, Strauss, and Giroux, 2016.

gradient. But it shows that the fundamental operation, the comparison of snapshots, is possible even in bacteria. And everything follows from this.

{...}

The Interpretation of Dreams (1973)

There's a famous story about a horse named Clever Hans, who was supposed to be able to do arithmetic: some Kantian would ask him what was seven plus five, and he would paw the ground twelve times. This baffled and amazed everyone who witnessed it until someone noticed that, if he couldn't see his trainer, Hans didn't know where to stop: he read the correct answer from subliminal cues. He didn't really know how to add, he just knew what his trainer wanted to hear.[106] — People are (usually) more clever than horses, and can interpret subtler signals.

When I first read Wittgenstein I spent about a month going through the *Philosophical Investigations*, taking copious notes and arguing with the author line by line. I don't think I ever finished the book, at least not on that occasion; the project devolved finally into making notes upon my notes upon my notes. When I had written thirty or forty thousand words, I figured I had done enough.

The experience taught me I could turn the reading of anything into a melodramatic struggle between myself and the author over the real meaning of the text, however, and for a while I read everything that way, because I had developed a narcissistic fascination with watching myself think. — Or at least pretending to think. — "Wrestling is not a sport, it is a spectacle," says Barthes.[107] No shit.

Thus when I decided to make a study of Freud — not that I thought for a moment that psychology (as opposed to, say, anthropology) was

---

[106] The phenomenon is general, in that, e.g., sniffer dogs also respond to cues from their supervisors: Lisa Lit, Julie B. Schweitzer, Anita M. Oberbauer, "Handler beliefs affect scent detection dog outcomes." *Animal Cognition*, 2011 May; 14(3); 387-394.

[107] Roland Barthes, "The World of Wrestling." *Mythologies*. New York: Hill and Wang, 1972.

a real science — let alone psychoanalysis, which was more like a form of literary criticism, an attempt to read the history of the Ego as if it were a work of literature — though there seemed to be a strange sort of depth in it — maybe this was what Kant saw in Swedenborg, it had that element of spirits speaking to us from other dimensions, accessing the unconscious mind in the same way mediums accessed the spirit world in séances — I got a copy of *The Interpretation of Dreams* and began reading it the same way, annotating nearly every line and writing comments continuously as I went. Progress was not rapid, but it was cheap entertainment, and I pressed forward, the intrepid European explorer hacking his way through the hermeneutic jungle with machete and rifle in search of a lost city of gold (hopefully ruled over by Ursula Andress). After a few days I had read several chapters and trashed the best part of a notebook in this heroic quest to fathom the nature of the psyche.

Then I had a dream about fucking my mother. I tossed the notebook into the fireplace, took my paperback Freud to the used book store to trade it in for a couple of space operas, and decided to let someone else refute psychoanalysis.[108]

And went back to reading fast and sloppy.

---

[108] I gather Grünbaum also figured out why the patients of Freudian analysts — subject to far more efficient imprinting than mere readers — had Freudian dreams, and, in his own inimitable fashion, expanded the argument into an enormous book. But of course I never read it. It could only have provoked a relapse.

{…}

Compare Harold Bloom:

> Though a kind of Freudian for a few years……a proposed study of him called *Transference and Authority* was the one book I have never been able to finish. And I had to abandon an annual graduate course on Freud, because as the term neared its end, my verbal slips, the parapraxes of Freud's *Psychopathology of Everyday Life*, augmented, until the final class became involuntarily hilarious, since I suffered a kind of misspeaking in tongues.[109]

[109] *Genius*, p. 181. New York: Warner, 2002.

{...}

Da(i)sein

Parenthetically: though you take it for granted that an artificial intelligence would have a mind and a soul,[110] it seems obvious it would not have a psyche; at least nothing a human would recognize as such. In that respect it would indeed be alien and uncanny.

(Ironically this doesn't mean a robot couldn't be an effective psychotherapist, as Weizenbaum[111] demonstrated long since with ELIZA, a program based on a few simple tricks which required neither mind nor soul. — What this says about psychotherapy I'm not sure. But it can't be good.)

---

[110] In the sense of Aristotle, *De Anima*. We'll get to that.

[111] See Joseph Weizenbaum, *Computer Power and Human Reason* [1976]. — What really amazed him, he said, was that when he had his secretary try talking to ELIZA, even though she had witnessed him writing it and knew it was just a sort of magician's trick, after a couple of minutes interacting with the program she asked him to leave the room to respect her privacy.

{...}

Mimicry

— is actually problematic. — As well it ought to be, this is almost the problem of induction.[112]

What did the neoclassicists mean when they said Art put the mirror up to Nature? (What gets reflected? How? — The question of what constitutes *representation*.)

The protagonist of William Gaddis's *The Recognitions* doesn't copy artist's paintings, but artist's styles. How can that be possible?

Bloom quotes a discussion[113] of how Boswell "impregnated himself with the Johnsonian ether"; a paradigmatic example of learning a style in the sense of Gaddis. — For the distinction here between remembering precisely what Johnson had said and reconstructing it from scribbled notes and an uncanny ability to

We learn words and phrases by rote but this isn't all of learning language; we can absorb grammars and generate new ones, internalize theories, fulfill much more elaborately articulated expectations — can interpret much subtler signals, and can respond in more sophisticated

---

[112] In the sense of what it is that allows you to continue a sequence, almost exactly.

[113] By Frederick Pottle: "Does Boswell, then, report Johnson's conversation verbatim? In particular sentences and in some brief passages of an epigrammatic cast, yes. In general, no. The crucial words, the words that impart the peculiar Johnsonian quality, are indeed *ipsissima verba*. Impregnated with the Johnsonian ether, Boswell was able confidently to recall a considerable body of characteristic diction. Words entail sense; and when elements of the remembered diction were in balance or antithesis, recollection of words and sense would almost automatically give 'authentic' sentence structure. But in the main Boswell counted on … an understanding, grown intuitive, of Johnson's habits of composition … to construct epitomizing sentences in which the *ipsissima verba* would be at home." [*Genius,* p. 170.]

fashion. The process is so opaque that Chomsky resurrected the theory of innate ideas to explain it. (His version of the theory of *Meno*.)

We don't simply answer by picking multiple choice.

The brothel scene in *Amadeus*, where Mozart improvises in the style of a series of composers — Salieri last and most ludicrously.

The reductio ad absurdum, Tony Hendra on Douglas Kenney: "The very first time I met Doug, he was talking like William Makepeace Thackeray. ...I mean, he was actually improvising his prose...and it was Thackeray, it wasn't Dickens...and about 30, 45 seconds after that he was demonstrating that he could put his entire fist into his mouth... ."[114]

Not so funny, David Halberstam on Robert McNamara:

> One particular visit seemed to sum it up: McNamara looking for the war to fit his criteria, his definitions. He went to Danang in 1965 to check on the Marine progress there. A Marine colonel in I Corps had a sand table showing the terrain and patiently gave the briefing: friendly situation, enemy situation, main problem. McNamara watched it, not really taking it in, his hands folded, frowning a little, finally interrupting. "Now, let me see," McNamara said, "if I have it right, this is your situation," and then he spouted his own version, all in numbers and statistics. The colonel, who was very bright, read him immediately like a man breaking a code, and without changing stride, went on with the briefing, simply switching his terms, quantifying everything, giving everything in numbers and percentages, percentages up, percentages down, so blatant a performance that it was like a satire. Jack Raymond of the *New York Times* began to laugh and

[114] *Drunk, Stoned, Brilliant, Dead.* [Douglas Tirola, 2015.]

> had to leave the tent. Later that day Raymond went up to McNamara and commented on how tough the situation was up in Danang, but McNamara wasn't interested in the Vietcong, he wanted to talk about that colonel, he liked him, that colonel had caught his eye. "That colonel is one of the finest officers I've ever met," he said.[115]

— Almost like — dare we say it? — the kind of patient one might write a paper about — ? —

{…}

Peirce distinguished induction from abduction; a more provocative name for it, though he seems to have meant something more like Popperian hypothesis. — What you wonder about is more like black magic, however, almost a species of Promethean theft: it is as if the idea has been stolen from the mind of God. — Almost like a form of telepathy, you might speak of "channeling" e.g., the Zeitgeist, perhaps, or speaking as the voice of the people.

Can you talk seriously of reading the mind of nature? — Here again though usually you seem to read it from the face, it sometimes seems that some deeper connection has been established.

---

[115] *The Best and the Brightest*, Chapter Thirteen. Here, alas, we see the origin of the infamous body counts.

{...}

I did keep track of my dreams for a while in high school, but this had nothing to do with Freud; rather I was intrigued with the speculations of J.W. Dunne, who thought dreams incorporated experiences not only from the past but also the future.[116] (I was not yet enough of a wiseass as to say this meant they employed the advanced as well as the retarded Green's function,[117] but that was the idea he was straining toward.)

So I wrote them all down in a diary for a month or two, to see if I could find evidence of precognition. Apart from a premonition about the outcome of the Rose Bowl,[118] I could find none. And the dreams turned out to be very dull, at least the ones I could remember. — Moreover contrary to accepted wisdom it did not get easier to recall them with practice, and I suspect those who claim that it does have learned instead not to notice they've just started making shit up.

{...}

Nonetheless occasionally I have had good ones. I recall one in particular, when I was seven or eight, in which I flew up and down through history in a time machine that looked like the *Spirit of St. Louis*;[119] it featured stampeding dinosaurs and a supernova

---

[116] J.W. Dunne, *An Experiment With Time*. London: A. & C. Black, Ltd., 1929.
[117] Wheeler, John Archibald, and Richard Feynman, "Interaction with the Absorber as the Mechanism of Radiation." *Reviews of Modern Physics*, Vol. 17 Number 2, April-July 1945.
[118] As has been pointed out by every skeptic since the world began, premonitions recalled after the fact are meaningless. — Conversation with a friend after the Challenger disaster: "I realized that I had had a premonition before the launch that the thing would blow up. And then I realized I had had such premonitions before every single manned launch going back to Alan Shepard." — He: "The exact same sequence of thoughts went through my head."
[119] Lindbergh's memoir was one of my favorite books as a kid. He and Saint-Exupéry epitomized the romance of aviation that NASA later did its best to destroy.

accompanied by a choral ode, and somehow affected me profoundly; at least I've never forgotten it.

{...}

In general, however, the most interesting part of dreaming is waking up and trying to continue the dream into conscious life. — There is a strange character to these meditations. They are like dreams themselves; not conscious analyses, but variations on a theme. — In this cross-conscious state the mind moves sideways at extraordinary speed, like motion on a frictionless surface, a sort of associational superfluidity, with the result that even though the dreams may be pedestrian the interpretations can be remarkable. I have thought of things this way that would never have occurred to me otherwise.

In fact I half suspect that remembered dreams are basically constructed in this state, partial rationalizations of mental states that don't translate into the representations of conscious life.[120] — I might go on at length here about an analogy with the relationship between language and music, but Nietzsche did it all better in *The Birth of Tragedy*.

{...}

On one occasion I was (I guess) engaged in some kind of flashback to my school days, which involved a chase through the steam tunnels in the style of the finale of *The Third Man*. But then my companion and I

---

[120] All this would be different if there were some way of monitoring the brain of a sleeping subject and turning that into words and images; as, e.g., in the scene in *Prometheus* [Ridley Scott, 2012] in which the android David/Michael Fassbender watches the dreams of the sleeping Noomi Rapace with a sort of hologram projector. If brain states could be recorded and played back, obviously, in some sense they could be remembered, and thus be accessible to consciousness. — There are recent experiments which seem to indicate this may indeed be possible; if so the terms of the discussion will be transformed completely.

tried climbing out to escape through some kind of narrow grating: he got away, but I got my head caught in a tight place. — Panicked. — And woke immediately, in full analytical mode: was this really birth trauma? It seemed too cute to be believable. — Then began to drift again, and imagined a cartoon: the moment of birth; doctors, taken aback; no head has yet emerged, but, first, a hat on the end of a stick. — I told this to an artist friend later; he was enormously amused, and at once began to sketch the idea.

{…}

Identity

Hume[121] extended skepticism to the existence of the soul or Ego: there is no direct perception of self, he maintained, only of its components, impressions and ideas.

I would say on the one hand yes, this is true, in that self-consciousness in real and limited human life is fragmentary; but on the other hand no, the perception of self is direct and unambiguous, in the same way that you don't perceive the *whole* brick when you look at it, you turn it over and look at different sides, assess its texture, its color, heft it to get a sense of its weight, etc., before (in imitation of Dr. Johnson)[122] you throw it through Hume's window to get his attention.

One can analyze the glimpses down to their individual fragments and there isn't much brick left either; nor window, nor Hume, nor *Treatise*, nor sentences nor words nor sense to be found in them — nor impressions nor ideas either. — It is altogether too easy, with analysis, to reduce the world to undifferentiated dust.

But — granting his hypotheses — ideas are impressions of impressions; still there are ideas of ideas; of ideas; of ideas; of — Hume only carries this to first order, in other words, and there is a hierarchy here of indefinite extent.

(Ulam quotes Banach: "Good mathematicians see analogies between theorems or theories, the very best ones see analogies between analogies.")

---

[121] See his *Treatise*, Book I, Part IV, Section VI, Of Personal Identity.

[122] It is probably superfluous to quote the relevant passage here;. see Boswell's *Life of Johnson* 6 August 1763.

{...}

Is it just intension and extension again? — The set versus the predicate that defines it.

(Something about the visible tree, and the root system that is not perceived. — Well, we can multiply metaphors at will….)

{…}

Factorization

Adjoint to the problem of identity is a related question which has always bothered me, what might be called the problem of factorization.

That is, we have on the one hand the totalitarian fantasy[123] of the master computer, the monster intellect which like Alpha-60[124] sits at a central location, receives all information, processes it, and issues orders; thus controlling a city, a nation, an empire — whatever. But on the other, the democratic hand, we may imagine a collective intelligence[125] that via some kind of telepathic mechanism shares the thoughts of many individuals, perhaps an entire race — brains as cells in a meta-organism — and multiplies their capacities to godlike proportion.

But there are reasons that the human body does not consist of one big cell, or one undifferentiated colony of cells. The logic of functional organization entails a kind of factorization, a division of labor.

If one tried to build a computer[126] the size of the solar system, it would take several hours at the speed of light to get a signal from a neuron[127] on one side to a neuron on the other; you also know that neurons can't be scaled down indefinitely, that they must have some finite minimal

---

[123] One which for obvious reasons has always been particularly dear to the military mind.

[124] It is, of course, ironic that Godard should have invented this parable of rebellion against totalitarian control, and then embraced Mao, who personified it.

[125] Cf., e.g., Arthur Clarke's *Childhood's End*, or its original, Olaf Stapledon.

[126] We phrase this in the terms of computer architecture because we understand that a little better. Obviously.

[127] There have to be fundamental elements of some kind, why not call them this.

size. So physical limitations suggest that partitioning computations into pieces which can be performed locally must be necessary. Also, the inherent limitations[128] on the efficiency of a single bandwidth-limited central processing unit that performs all computations serially vis-a-vis a parallel architecture that distributes them are also well known, and this argument leads to the same conclusion. So the picture of a distributed network with many centers of computation emerges naturally by more than one line of reasoning.

Godfrey-Smith notes that, in distinction to the design of the chordate nervous system, which channels signals along the spinal cord to a central brain, "… much of a cephalopod's nervous system is not found within the brain at all, but spread throughout the body. In an octopus, the majority of neurons are in the arms themselves—nearly twice as many as in the central brain. The arms have their own sensors and controllers. They have not only the sense of touch, but also the capacity to sense chemicals—to smell, or taste. Each sucker on an octopus's arm may have 10,000 neurons to handle taste and touch. Even an arm that has been surgically removed can perform various basic motions, like reaching and grasping."[129] — Exactly.

So there is a tricky optimization problem here, one whose solution is problem-specific, involving the best balance between the number of cells and their individual capacities. In the original design of his Connection Machine,[130] Danny Hillis imagined the individual elements would be minimal, with very little local memory; in subsequent realizations of his ideas economics dictated wiring together existing computers with rather complex architectures instead, though

---

[128] See John Backus….

[129] Peter Godfrey-Smith, *Other Minds,* Chapter 3.[New York: Farrar, Straus and Giroux, 2016.]

[130] A radically elegant departure in computer architecture, based on the idea of wiring 2^n processors together in parallel as nodes in an n-dimensional Boolean hypercube; described in his thesis: W. Daniel Hillis, *The Connection Machine.* [Cambridge: MIT Press, 1986.]

the trend has been to reduce their size. The design of the neuron, in other words, is still evolving.

So if one really did attempt to build some kind of maximal intelligence, it would look more like a society than an individual mastermind, and it isn't obvious where the tipping-point lies. What you have to guess is that there is some kind of theoretical upper limit to (individual) intelligence, that manifests itself as a threshold of stability, and beyond this if you try to add more processors the whole will begin to behave like a group of individuals communicating — indeed probably arguing — with one another; not like an undivided unity. — That there is a logical necessity that turns Alpha-60 into the City Council, in other words.

And clearly to some extent though there is a unifying identity this has already happened in the brain as we see it, though we don't understand the interaction of the component parts.

Here also I should mention something that I think of as the principle of the cell wall: that parallel computation can only be efficient if possibilities can be evaluated independently, without interference; you must be able to do one thing and work it out to a conclusion. The evolution of life required the separation of genetic material into separate individuals to ensure that experiments would be independent.

— However. — Identity and factorization *are* adjoint, the dialectic of analysis and synthesis is universal (or as universal as anything gets), and though treating an arbitrarily large collection of individual neurons as a single intelligence may not be valid, this doesn't rule out the possibility of hierarchical organization, with collectives at one level forming individuals at the next;[131] this is the idea of the monadology, after all, and this too corresponds to some fundamental principle in

[131] It's also possible that "levels" may parse differently depending on point of view. But one conceptual morass at a time.

nature. — So we haven't really eliminated the possibility of Gaia. We're still articulating the reasons that we don't understand how it could work.

{...}

Of course given any mathematical model of consciousness the first question is how to generalize it. We can easily imagine more primitive modes of consciousness — I would certainly say that dogs have minds and souls, for instance, but doubt their memories are quite so acute as ours, and suspect therefore that their ability to think *about* what they are thinking is considerably less[132] — there is a continuum, clearly, but there may be some kind of abrupt transition when complexity attains some critical value — but can we imagine higher forms? by which we don't refer to some form of cosmic insight (the drugs trivialized all that, I'm afraid) but some extension of the idea —

— Perhaps to higher dimension? suppose we picture a two-dimensional time, and try the string-theoretical trick of extending world lines to world sheets: one might imagine a manifold of sequences of lived experience, each taken as a state, and a — continuous? — progression through them.

But it isn't obvious this amounts to a real extension: taking an entire world line as a state is simply adopting a more complicated definition of "state", Turing machines with multiple tapes can be linearized, functions of more than one variable can be reduced to functions of a single variable by the trick of "currying", etc.

— A better idea is to relax the requirement of linear memory. (To split the causal chain of consciousness.) Nothing in principle prevents the extension of the reference/containment idea to more general partial orders, and though the notion of a tree of possibilities which branches at every moment is both overfamiliar and rather nebulous, one might, for instance, suppose the possibility of cloning a personality completely, say in a science-fiction scenario in which (I will get to this

---

[132] They are however obviously conscious, since they can be unconscious, sleep, and dream.

later) you transmit a copy of yourself to Alpha Centauri, the two versions of yourself have parallel series of adventures, and then the copy is transmitted back and reintegrated with the original.[133] — It is difficult to believe that the human brain as presently constituted could absorb this shock, but a form of consciousness that could is a very natural idea of the superhuman.

(Perhaps worth noting that though there's no logical difficulty in reconciling two separate threads of past experience there wouldn't be any subjective sense of whether an event on one timeline occurred before or after an event on another; no matter that some external clock might be able o decide the proposition objectively. — Threads that are not recombined, on the other hand, are simply separate personalities; rather odd to contemplate but then in some sense every living creature represents a branch on the tree of life, and, etc.)

Strange but true there's even the hint of such speculation in the classical literature: there is an old tradition, referred to in the Second Part of Goethe's *Faust*, that the real Helen went to Egypt and it was a phantom double who went to Troy. When asked about this she says [8880]

> This is a superstition of dark-tangled sense!
> Which of them am I? Even now I do not know.[134]

Though in principle nothing prevents her from having been both.

— The extent to which multiple simultaneous *conscious* threads of execution may be sustained by real existing humans, on the other hand, is not obvious. It can be verified with simple psychological tests that the vast majority of people (in excess of ninety percent) are

---

[133] It would be more natural, of course, to imagine a swap, in which both copies continue to exist after integrating their separate threads of experience.

[134] David Luke translation.

incapable of true multitasking, for instance; a small minority, on the other hand, seem to be much more capable. — It is also said that some gifted individuals can attend to several conversations at once, or read and converse at the same time — Caesar, for instance, is supposed to have had this ability — and if I ask myself whether when Bach played the organ he was doing four intricate things at once, or just one really complicated thing that involved both hands and both feet, I don't know what to answer. — But the *possibility* certainly exists.

{…}

Memory in childhood

I had noticed by the time that I was seven or eight[135] that my earliest memories were already fossilized; no longer the living originals, but copies, the imprints those had left in stone.[136]

I asked my mother about a few of these, because I had no context for them other than they seemed to come before any others. — One was of very bright sunlight on white sand, with a blue expanse of water beyond it. She said that was Lake Huron, at the tip of the thumb, when I was two years old. — Another was of lunch at a kitchen table, and a radio playing *The Romance of Helen Trent*. She said that was the first apartment they'd had after they got married, and I may have been less than a year old. — I think this was an exaggeration, but I also retained a subliminal impression of a brick building, which must have preceded the first of many suburban houses. — I also remembered my playpen, though not what she claimed were my effortless escapes from it. — In any case I have seen home movies of my third birthday, and I remember nothing of it.

I also remembered standing in our attic bedroom, looking at my sister in her crib, and forming the perfectly subvocalized sentences: "I am three. That is my sister. She is two." — There is a charming simplicity in this. It is almost worthy of a British philosopher.

But the continuous thread of conscious life and memory only began around the time that I was four. My impression, which can only be slightly exaggerated, is that I understood what had happened almost

---

[135] It had to be about that time, since the analogy would not have occurred to me before I had read George Gamov's *Biography of the Earth*, which explained the formation of fossils.

[136] As for how this metaphor would have occurred to me — well: I had already read George Gamov's *Biography of the Earth*.

immediately. With this commenced a process of self-examination which continues to this day.

Before that self-awareness had been patchy, and interrupted by blackouts and lapses. — I recall vividly, for instance — or did, it is hard to get the tenses right — one day when I was left alone to take a bath. When my father came back to check he found me playing contentedly with my toys at one end of the tub; but pointed out, as sternly as he could manage, that there was something at the other. I perceived a discoloration in the water, and realized this was the signature of the presence of a large turd, which had somehow materialized on the bottom. Clearly I had to be responsible, but I remembered nothing, and said, truthfully, that I hadn't been aware of what I was doing. He didn't understand me — in fact this was the reason I remembered the incident so clearly, because it puzzled me that he didn't understand — and asked me, in apparent seriousness, whether I thought the Lone Ranger would have done something like that. — I stared at him blankly, because it seemed like such a strange question. — As indeed it still does.

— Well. — What can I tell you. — Marcel Proust I am not.

{…}

The unconscious

The inherent necessity of shutting the mechanism down for regularly scheduled maintenance became apparent to me during a (thankfully brief) period when I abused the use of amphetamines, and became habituated to staying up three and four days at a time. What I discovered was that dream interludes began inserting themselves into waking life whether you wanted them or not. — A canonical instance was an occasion when I was (so I thought) trying to follow an argument in Whittaker/Watson, and realized when I turned one page that the cavalry had just arrived in the nick of time on the one preceding. — Wait a minute, said I….

At any rate dreams, though fascinating, are not some kind of unique window into "the unconscious", which is essentially a kind of triviality.

This was less obvious before the invention of digital clocks. Now it happens constantly that I glance at the readout in the upper right hand corner of the screen and notice — what a coincidence — that the time is something like 3:14, or 2:34 (you look up a moment later and it is 2:46), or 3:43, or 12:16, or (this only began after 2001) 9:11.

My laptop has a sexed-up version of a Unix utility which in the MacOS is called the Activity Monitor; it list all the processes resident on the CPU at any moment, and numbers the subprocesses subordinate to each. This fluctuates continuously, but typically there are around 180-190 processes and 900-1000 threads of execution dependent on them.[137]

---

[137] For the new generation of laptops, double the numbers.

Clearly something similar holds for the human brain. Some thread is constantly checking the clock for me, and promotes the time to attention when there's something significant about it. And whatever it is recognizes arithmetic sequences, and knows Beethoven's birthday, and the cube of 7, and a great deal besides. — Other illustrations may be multiplied at will, but consider, e.g., the common experience[138] of glancing at a page without seeming to see it, having the feeling there was something funny about it, and then realizing on closer inspection that it contains the name of someone you knew in elementary school.

So this answers the question "Where do thoughts come from?" — It is just as with Santa Claus and his toys: an unseen army of little elves labor to produce them.[139]

{…}

The world's longest "Duh"

The premise of *Blink* [Michael Apted, 1993] is that a woman (Madeleine Stowe) who has been blind since childhood has had her sight restored by an operation; her capacity for processing visual input, however — the software, it is intimated, not the hardware — has atrophied from disuse, and until it is reconstituted she is unable to interpret what she is seeing; images pop into her head at unpredictable intervals after the stimulus that prompted them, often days later. — Of course this is immediately complicated by her being witness to a murder, but — never mind — this really does happen, though I have usually observed it with auditory rather than visual inputs.

In particular it is a frequent occurrence, for someone surrounded by speakers of an unfamiliar language, to hear some word or phrase

---

[138] At least, this sort of thing happens to me all the time.

[139] Henri Poincaré had a lot to say about these little elves; see below.

spoken and be unable to recognize it; until some pattern-recognition engine which has been chugging away in the interim announces its result, often minutes later.

The record, which is indicative either of the unusual strength of my unconscious muscles or of just how stupid I really am, is held by the answer to a question I put to my hosts on the first New Year's Eve I spent in Argentina: we were eating something, I asked what it was, they made some incomprehensible noise[140] in reply, I pawed about aimlessly in the dictionary, I gave up. — "Looks like breaded veal," I thought. — Six months later while browsing in the carnecería I saw a sign saying "Milanesa" and knew instantly what I had heard. — And glancing downward, saw a familiar set of cutlets on display in the cooler below it.

[140] There is no harm in belaboring this elementary point of epistemology: you have no idea what you have heard if you don't know what to listen for; pattern recognition is a matter of matching pre-existing templates — not to concede Chomsky's point, but he was certainly onto something — and when they don't exist, the result is chaos. — The same is true of vision, of course, and even — for reasons inherent to the logic of perception — machine vision; which goes a long way toward explaining the UFO phenomenon, even when it appears to be confirmed by electronic sensors and computer software.

{…}

Of course there are less obvious cases — waking in the middle of the night, checking the digital clock and noting that it is 4:20 on the morning of 4/20, for instance — is there a clock in my head? maybe, it sometimes seems that way — weirdly inappropriate to apply such mathematical precision to a stoner holiday though; is this the work of the Inner Prankster? — and there is disturbing evidence of a kind of slippery slope: I could laugh it off, for instance, when knowing from previous experience that I would be spending the entire day waiting in office antechambers for someone to hand me the next form to be filled out in the paper chase I carried *Ulysses* with me to my first day of work on a government summer job, read half of it with a curiously obsessive focus, and only realized later that the day in question was, indeed, June 16th — Bloomsday — but it was more perplexing when I was seized one evening by the impulse to read a book about Boethius, and discovered halfway through it that this was his feast day (October 23) — how could I have known that? I couldn't recall ever having read anything about him; Russell mentions him with admiration but never refers to the Church calendar. So where did that come from? How did that fact enter my head? — Jung did have something right. Sometimes it looks as though external reality is also wired into the unconscious mind. That dreams connect us by some subterranean channel to *anima mundi*.