Indiscrete Affairs

Volume II — Engaging the World

Brian Cantwell Smith

Last edited: 17 January 2010

Draft only (version 0.80

Faculty of Information
45 Willcocks St W, Toronto
Ontario M5S 1C7 Canada
brian.cantwell.smith@utoronto.ca

— Were this page blank, that would have been unintentional —

Table of Contents

— Were this page blank, that would have been unintentional —

Prior Publication Details (3 out of 11)

1. **From E&M to M&E:** in Luciano Floridi, ed., *Philosophy of Computing and Information: 5 Questions*, Automatic Press/VIP, 2008?
2. **Requiem for the Computational Theory of Mind:** Never published; presented as Presidential Address, 25th Anniversary Meeting, Society for Philosophy & Psychology, June 21, 1999, Stanford.
3. **How Can I Miss You If You Won't Go Away?:** Never published
4. **Subjectivity and Objectivity:** Never published. Presented as the Thomas Langford Inaugural Lecture at Duke University on Dec. 4, 2001, inaugurating the Kimberly Jenkins University Professorship in Philosophy and New Technology.
5. **The Nonconceptual World**: Never published
6. **Dennett on Smith, and Reply:** in Hugh Clapin (ed.), *Philosophy of Mental Representation*, Oxford University Press, 2002
7. **Rehabilitating Representation:** Never published
8. **Devil in the Digital Details:** Not really published. Appeared (I think) in la *Calcografía Nacional Simposio internacional arte gráfico y nuevas tecnologías* (an exhibit catalog), 2002
9. **Deconstructing Digitality:** appeared in Idea&s, a publication of the Faculty of Arts & Science, University of Toronto.
10. **Indiscrete Affairs:** Never published
11. **Interview:** Never published in its entirety; an abbreviated version appeared in W. Mark Richardson and Philip Clayton, eds., *Science and the Spiritual Quest: New Essays by Leading Scientists*, Routledge, 2002.

— Were this page blank, that would have been unintentional —

A • Introduction

— Were this page blank, that would have been unintentional —

1 — From E&M to M&E

A journey through the landscape of computing[†]

Not only are the philosophy of computing journey through the landscape of computing and the philosophy of information new fields, still theoretically unstable, but the subject matters they span are exceptionally broad. "Information" covers so many phenomena as to be threatened by vacuity—though that has not deterred people from using it as an explanatory concept in fields as diverse as biology, computer science, medicine, journalism, electrical engineering, literature, the arts. Computation is narrower, and seems better understood, in part because of half a century's work on mathematical theories of computability. But here I believe appearances are misleading. Not only do we not understand computing as well as is generally thought, I will argue, but making progress will require upending all sorts of fundamental assumptions in ontology, epistemology, and even metaphysics.

This combination of newness and breadth means that no contributor to this volume can assure the reader that the path they have traveled through the landscape may not be due as much to their own philosophical predilections as to any intrinsic geography. So there is merit to Floridi's suggestion that we start with biographic details. However it also means that all writing in these

[†]Originally published in Luciano Floridi, ed., *Philosophy of Computing and Information: 5 Questions*, Automatic Press/VIP, pp. ■■–■■.
"E&M" is physics' moniker for *electricity and magnetism*, the field from which I entered computing. "M&E" is philosophy's parallel epithet for *metaphysics and epistemology*, the landscape to which my travels through computation have led.

areas (my own included) is liable to fall prey to Isaiah Berlin's challenge that "writing is amateur when you learn about the author, not about the subject matter."[1]

Forewarned is forearmed.

1 Origins

My own interest stems from my first semester at university, when an IBM 360/44 was delivered into the basement of the Oberlin College physics department. Riven by a naïve version of C P Snow's two-culture dilemma, I wrestled with whether to drop physics and major in religion, debated politics with anyone who was awake, and spent the remainder of my nights ferrying stuffed boxes of punched cards back and forth to the operator's window at the Computing Center. Crazed, yes; but it made a kind of manic sense. Knowing nothing of hermetic methods or intellectual precursor, I was possessed by a conviction that the power and elegance of science, the gravity and richness of politics and religion, and the intensity of intimate human communion were ultimately more similar than they were different.

Within two months I had made two life-altering decisions. First, I vowed to dig deep enough to get to the place where these superficially different perspectives could be understood, if not as "one," then at least as integral—as part of a single encompassing reality. Second, at a more pedestrian level, I asked my physics professor for six weeks off from doing problem sets, to figure out whether the school's new computer might help with this quest. What I wanted to know, I told him, was whether computing could be understood with all the power and insight and elegance that I loved in the sciences, but nevertheless do justice, in a way no prior scientific account ever had, to the richness and complexity of the human condition.

It was a classic sophomoric venture: wisdom shot through with foolishness. All told, it was not a bad question. But six weeks turned into forty years.

The first results stemmed from those long nights of debugging. Inchoately at first, but more articulately as the years went by, I

[1]«Ref»

came to believe that the understanding of computing I was deriving from concrete engagement—not just at Oberlin, but later writing operating systems, implementing data bases, designing programming languages—was not accounted for by what was being taught about computing in the nascent field of computer science. The problem wasn't just that theories idealized, or ignored practical realities with which one had to come to grips in real-world settings. As much is true of any engineering practice. Rather, I could never shake the feeling that the accounts were profoundly wrong, misguided at their core—"missing" what mattered most about the territory we were tacitly and somewhat blindly exploring.

In parallel, motivated by an interest in people and mind, I was drawn into artificial intelligence (AI) and cognitive science, initiatives whose fortunes were on the rise, as society grappled with the monumental idea that computing was not just a technology of disruptive impact, but also a powerful idea-perhaps even one that applied to us. Maybe we, too, were computers. Debates raged, with endorsements rung from the MIT, Carnegie Mellon, and Stanford AI laboratories,[2] critiques lobbed back by Weizenbaum, Dreyfus, and Searle,[3] and more speculative analyses taken up across the philosophy of mind.[4]

Naturally, I wanted to formulate my own position. But I was blocked by my underlying sense of discrepancy between how we thought computers worked and my blood-and-bones intuitions as to how they actually worked. The situation is depicted in figure 1. Debate on what came to be known as the computational theory of mind (CTOM) was presumed to have the structure la-

[2]Particularly Marvin Minsky & Seymour Papert at MIT, Allen Newell and Herbert Simon at Carnegie Mellon (CMU), and John McCarthy at Stanford.

[3]Especially Joseph Weizenbaum's ELIZA program (1966), Hubert Dreyfus' *What Computers Can't Do: A Critique of Artificial Intelligence* (1972), and John Searle's "Chinese Room" thought experiment (1980)

[4]E.g., Haugeland's *Mind Design* (1981), and *Artificial Intelligence: The Very Idea* (1985). Mid 19th-century philosophical discussions of computing were primarily conducted in the philosophy of mathematics, pursuant to Gödel's proof and Turing's computability results; by the end of the century, the debate had moved to the philosophy of mind. It is only now that an authentic philosophy of computing is coming into its own.

beled (. What it is to be a computer was assumed to be uncontentiously formulated in the "received" theory of computation, labeled q_c in the diagram. At stake was how to understand the mind: q_m. The substance of the CTOM was taken to be the thesis that $q_m \approx q_c$.[5]

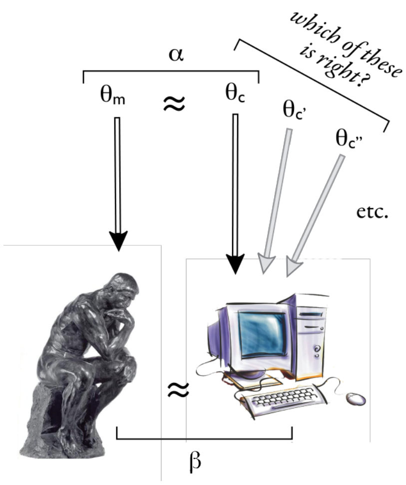


Figure 1 — The computational theory of mind

My problem was straightforward. Fundamentally, I took the CTOM not to be a *theory-laden* proposition, in the sense of framing or resting on a specific hypothesis q_c about what computers are, independent of whether q_c held of real-world computers. Rather, I took it to have an ostensive or "transparent" character: that people (i.e., us) are computers in whatever way that computers (i.e., those things over there) are computers—or at least in whatever way some of those things are or might be computers.[6] It wouldn't be interesting, I felt (this was no attempt to vindicate Weizenbaum, Dreyfus or Searle), if it emerged that, sure enough, q_c was not true of people, but q_c was not true of the IBM 360, ABS brake systems, or my word processor, either. Suppose, in particular, as I suspected, that q_c was not the right theory of computing, but instead that q_c', or q_c'', or some other account, were "correct" or anyway to

[5]Not, of course, that anyone thought that all computers were minds (M=C). Even if all minds are computational, the class of computers is larger, and so clearly M , C. This raises the question of how what was specific to mind, and how that would be articulated. The complexity of the questions was rarely explicitly addressed, but it was presumed that the restriction of C to M would also be expressed *in computational terms*—as opposed, say, to minds being those computers that "weigh more than 1 lb but less than 10," a restriction that leaves minds as computers, but where the restriction itself is not, as it were, a "computational" restriction, not being framed in terms of a property (weight) that is itself a "computational property."

[6]Cf. footnote ■■.

be preferred?[7] Then the only interesting question, I believed, was whether q_c' or q_c'' (or whatever) held of people.

A myriad challenges can be raised against this approach, including: that an "empirical" stance cannot be right, because q_c (i.e., the accepted mathematical theory of computation and computability) is how computing is *defined*; that because we build computers, we must understand them; etc. While I disagree with all of this, this is not the place to address it. The point is simply that I believed (i) that the only interesting version of the CTOM needed a theory that did justice to computing, and (ii) that q_c was not it. And so, around the mid 1970s, I took up the project that occupied me for the next twenty-five years: to figure out what computing is (i.e., which variant of q_c is right), at a level of depth strong enough to found an adequate theory of computing, and richly enough articulated to support substantive debate about a relevant computational theory of mind.

Before I could address the question of whether the CTOM was true, that is, I needed to know what it said.

By 1972 I had moved to MIT, an epicenter of AI and cognitive science. Instead of entering those programs directly, I first enrolled as a "social inquiry" major-reflecting my interest in assessing, rather than embracing, the CTOM. But the same problem of inadequacy in reigning conceptions of computing impeded my participation in that fledgling STS program—for example, the assumption that computing is a technology, as opposed, say, to a form of art or sculpture. Recognizing that insight could only come from substantial engagement, I transferred to the Artificial Intelligence Laboratory for the remainder of my education.

2 Preliminaries

Philosophy of computing is in its infancy. Whether history will even notice us I do not know, but we have certainly just scratched its surface. Take a dozen terms of the computational art: *program*, *process*, *algorithm*, *symbol*, *data structure*, *implementation*, *architecture*, *complexity*, *object-oriented*, *user-friendly*, *nondeterminism*, and

[7] Whatever "correct" comes to, whether that is even the right term, etc. For simplicity, I have phrased the issue conservatively.

procedural. Every one remains unreconstructed—some more so than others, but all to an extent that five minutes in an undergraduate class is enough to raise questions that outstrip contemporary comprehension. Even such fundamental notions as *being computational, carrying information, being algorithmic* or *effective,* etc., remain open. No one knows whether they are: *intrinsic,* like mass and momentum; *abstract,* like numbers or types; *relational,* like being well-loved; or *require external ascription or interpretation,* like the meanings of books and text.[8]

I do not say this to be negative. On the contrary, the inchoate state of our understanding greatly energizes the field. It is like participating in the early days of physics. Graduate students can still read everything that has been written, and set out to explore largely uncharted intellectual realms. Critical issues are at stake—not just fundamental ones of meaning, mechanism, and reality-but also such notions as credibility, authenticity, engagement, and the like.

Nevertheless, the modest state of the art does suggest that students enter the field with some humility, lest they be misled into taking more for granted than is warranted. Four cautions strike me as especially important:

C1 Computers, computing, computation: It is essential not to assume pretheoretically any particular conception of—or distinction among—such familiar notions as *computer, computing, computation, computable,* etc. One such view has become something of a commonplace in computer science: that *computations,* viewed as abstract objects, are the entities of theoretical interest; and that *computers,* merely physical devices that realize or implement computations, are of no theoretical significance (no matter how economically and pragmatically consequential). This is the stance immortalized in Dijkstra's famous claim that "computer science is no more about computers than astronomy is about telescopes."[9] But any substance to that blunt pro-

---

[8]Everyone, including I, can raise objections to every one of these examples, and to the four-way typology. That is the point. The intellectual structure of the inquiry is still up for grabs.

[9]«Ref»

nouncement,[10] including the distinctions it is framed in terms of, depend on a theoretical framework the adequacy of which should be in question in any foundational analysis.[11]

C2 Equivalence: It is critical not to give undue weight, especially conceptual weight, to the famous equivalence proofs underlying mathematical computability-proofs according to which various different "models" of computing (Turing machines, the l-calculus, Kleene's recursion equations, etc.[12]) are shown to "compute" the same class of mathematical functions. Not only is the legitimacy of these proofs rarely questioned; it is also common to assume-falsely, in my view—that they show that the different models are "equivalent" for other purposes as well. Some illustrative problems:

a. As in C1, the proofs rely on a conception of what it is to "compute"—a notion that should be questioned, not assumed, in a foundational account. To assume such "post-theoretic" equivalences in advance will inevitably prejudice, and at worst render circular, any account of computing based on them.

b. The notion of "compute" on which the equivalence proofs rely is *extremely restricted*. Issues of input and output, and any other form of interaction or engagement with the world, are not so much ignored as banished—kept outside the framework. No mention is made of how the tape is initialized, how the results are "read out" (or interpreted; see C4), or anything of the sort. Time and timing are similarly dismissed. While complexity analyses pay some attention to resources, the claim that a universal machine can do "anything"—at least, "anything that can be done by machine"—is ex-

---

[10]With which, as it happens but perhaps not surprisingly, I disagree.

[11]Note the contrast with cognitive science and philosophy of mind, which (especially in the analytic tradition) used to view "mind" as essentially independent of body-a dualism that has come crashing down in recent years.

[12]Paradigmatically, devices of minimal structure given access to indefinite storage.

cruciatingly narrow. Tap out the differences among rhumba, reggae, and bebop? Make a cup of coffee? "Out of bounds!" is the standard reply. But who said so—and why? These are things a philosophical analysis should explain, not presume.

c. Conversely, the equivalence metric used in the equivalence proofs is extraordinarily broad—so broad as to sweep under the rug virtually every distinction that might be relevant to a theory of mind: how the device works; whether the resulting computation is intrinsic, ascribed, relative, relational, etc.; how long it would take to run; and so on. The standard way one shows that one machine can "do the same thing" as another, in fact, is to have the first machine model or simulate the second-the very distinction on which Searle based his critical distinction between "weak" and "strong" AI.[13] Distinctions on which competing theories of mind are distinguished-behaviourism, representationalism, type or token reductionism, materialism, etc.—are similarly obliterated in the quest for isomorphism.

d. All semantic issues, about meaning and interpretation, are again ignored or banned. In his original work on information theory Shannon was particularly articulate about this setting aside of issues of meaning and content; for reasons described below, the situation in computing is more complex. But independent of the use of words, fundamental issues of how systems signify, represent, carry information about, are interpreted as, or otherwise relate to the world around them are not addressed by any received theory of computing.

C3 Semantic Soup: In days of Ptolemaic and pre-Copernican astronomy, it was easy to distinguish among the various accoutrements of inquiry: *theory*, *experiment*, *equipment*, *model*, *representation*, *subject matter*, etc. *Theories* were viewed as abstract; *representations* were written down,

[13]Just because *x* simulates *y*, that doesn't mean that *x* is *y*. In fact it may imply that statement's denial.

probably on parchment; *models*, such as brass orreries, likely sat on tables; celestial *subject matters* were a long way away. In computational times, however, one encounters claims that instances of all these categories are of the same kind: computational processes of one sort or other.[14] Even in Turing's original paper, distinctions among numbers, representations of numbers, and numeric models are conflated after just a few pages. The mathematical proofs mentioned above, along with such kin as Gödel's incompleteness theorems, category theory, and the like, identify (i.e., conflate) all manner of isomorphic things. Current writers sometimes muse about the overlap,[15] but by and large it receives little attention. The caution, here, is not so much an injunction not to do this or that, but to keep a strict eye on the soup of semantic relationships in which computational systems simmer, lest the intentional character of the phenomenon dissolve from view.

C4 Mathematics: In part because of the prior three cautions, I enjoin students never to use mathematical examples as paradigmatic illustrations of computing, or as case studies on top of which to develop a general account. Numbers, numerals, mathematical models, and the like are simply too easy to confuse or conflate for it to be possible to "extract" the true nature of what is going on. Not only that; people's philosophies of mathematics differ by more than the issues at stake in philosophy of computing and/or philosophy of mind. Some people take numbers to be concepts; others, to be Platonic abstractions; still others, to be numerals or expressions; etc. How can one forge a cogent philosophy of computing in the face of such ontological profusion? Better to pervert Gertrude Stein to our purposes: "Forget numbers; think about potatoes."

By way of preparation, especially for those new to the field, two

[14] I am not saying a theory can be a computation (as opposed to something more abstract); merely, that some people claim so.

[15] Cf. Edelman's ironic comment that he had validated his emphatically non-computational model by "implementing it on a computer." «ref»

additional observations need to be added to these four cautions:[16]

P1 Terminological Archeology: Much of the theoretical vocabulary we use to study computing was not invented de novo. A great many terms of art were borrowed from logic and metamathematics—the areas in which Turing, Kleene, and other computational progenitors worked. Thus such notions as *syntax*, *semantics*, *symbol*, *identifier*, *variable*, *reference*, *interpretation*, *model*, etc., were used technically in logic long before they were pressed into computational service. This overlap has generated more confusion, I believe, than has been adequately recognized.

Searle's two arguments against the possibility of artificial intelligence are striking examples: (i) his "Chinese Room" argument, that semantics does not inhere in syntax;[17] and (ii) his parallel argument that syntax does not inhere in physics.[18] Searle was trained as a philosopher, and would have learned the words 'syntax,' 'semantics,' 'formal,' etc., from logic. To a person, as far as I know, computer scientists, on reading his arguments, feel that Searle "just doesn't get it." What I have told students for more than twenty years, however, is that *Searle would have been right*, if the words meant what he was taught that they mean—if, that is, by 'syntax' and 'semantics,' computer scientists meant what the people they took those words from (i.e., logicians) meant by those terms. This is not to excuse Searle, whose conclusions I am not endorsing;[19] but it does throw down a gauntlet that we say, in language that non-computer scientists can understand, what computing is. Yet another reason why the philosophy of computing is so important.

P2 Interdisciplinary Theory: Finally, it pays to attend to the

[16] It will be obvious later why these two deserve mention here.

[17] «Ref BBS»

[18] Chapter 9 of *Rediscovering the Mind*, MIT Press, 1992.

[19] I agree with him, as it happens, both that syntax does not inhere in physics, and that semantics does not inhere in syntax—at least on local interpretations of all those words. Where I disagree with him is on the underlying assumption that computation is syntactic.

> relations between substantive issues that arise in computing and allied questions addressed in other fields-especially as the place of computing in the overall intellectual landscape is not yet well understood or agreed. Just one example: computer science has extensive vocabulary to talk about the relation between one system understood as "an a" and that very same system understood as "a b"—i.e., as we say, one and the same system analysed at different levels of abstraction. As well as using such basic terms as 'implementation' and 'realization,' computational discussions involve such notions as *abstraction*, *modularity*, "black-box" and "grey-box" *implementation boundaries*, *importation*, *exportation*, *interoperability protocols*, *interfaces* (including APIs[20]), etc. Philosophy of mind and philosophy of science have developed their own theoretical apparatus to deal with what looks at first blush to be the same subject-under such terms as type- and token-reduction, (local and global) supervenience, multiple realisability, etc.
>
> For years I have offered to supervise a doctoral student to conduct a theoretical analysis of trans-disciplinary vocabulary in this or various allied areas, since I am not aware of any other systematic investigation of how the two analytic frameworks relate. No takers so far, but the offer remains open.

Enough preliminaries. Once the land is cleared, the project of developing an adequate philosophy of computing opens up into something like Frege's investigation of number-except that the empirical commitment requires maintaining focus on concrete, in-the-world phenomena. In that way it is also reminiscent of questions in the foundations of physics: about meaning, interpretation, measurement, and reality.

It does rather mean starting from scratch. But such is the nature of the enterprise.

3 Project

Given these considerations, how can one proceed? My approach has been to intersect three cross-cutting "axes" along which com-

[20] "Application specific interfaces"—…

putation has historically been analysed—generating something like an informal coordinate system in terms of which to map the computational territory. In no way do I endorse the resulting cartography as theoretically sound, or even as particularly coherent. By the time I am done, in fact, I discard every one of these distinctions, or reconfigure them beyond recognition. Still, the system pays its way as an initial guide.

3a Construals

The first axis enumerates seven "construals" of computing, as I put it, that have variously held sway in our intellectual discourse:

1. **Formal Symbol Manipulation (FSM):** the idea, derivative from a century's work in formal logic and metamathematics, of a machine manipulating symbolic or (at least potentially) meaningful expressions independent of their interpretation or semantic content;

2. **Effective Computability (EC):** what can be done, and how hard it is to do it, "mechanically," as it were, by (an abstract analogue of?) a "mere machine";

3. **Execution of an Algorithm (ALG)** or **Rule-Following (RF):** what is involved, and what behaviour is thereby produced, in following a set of rules or instructions, such as when making dessert;

4. **Calculation of a Function (FUN):** the behaviour, when given as input an argument to a mathematical function, of producing as output the value of that function applied to that argument;

5. **Digital State Machine (DSM):** the idea of an automaton with a finite, disjoint set of internally homogeneous machine states-as parodied in the "clunk, clunk, clunk" gait of a 1950's cartoon robot;

6. **Information Processing (IP):** what is involved in storing, manipulating, displaying, and otherwise trafficking in information, whatever information might be;

7. **Physical Symbol Systems (PSS):** the idea, made famous by Newell and Simon (1976), that, somehow or other,

computers interact with, and perhaps also are made of, symbols in a way that depends on their mutual physical embodiment.

I do not claim this list is exhaustive. Several more have recently made it onto the scene: non-linear dynamics, complex adaptive systems, a view of computing in terms of interacting agents, and so forth-all of which could be used to extend the list. Contrapuntally, a host of familiar ideas must be set aside as inappropriate for foundational duty: (i) demeaning characterisations, that computing is just something or other (machine, mechanism, artefact, attributed, etc.); (ii) negative construals, such as that computing is not some way (conscious, original, alive, and so on); and (iii) "higher-order" or adverbial specifications, such as abstract, universal, formal, etc., which only gain traction against some presumed prior property. The members of all three categories implicitly rely on another conception of computing, in order to have any substance.[21] But leaving such complexifications aside, it is the seven listed above—what I call the **classic construals**—that, at least to date, have shouldered the weight of the intellectual debate.

It is critical to recognise that all seven construals are both intensionally (conceptually) and extensionally distinct. In part because of their great familiarity, and in part because "real" computers apparently exemplify more than one of them, but perhaps especially because of the pernicious influence of those pesky equivalence proofs, it is often thought that the seven are roughly synonymous. This conflationary tendency has been especially rampant in cognitive science and philosophy of mind, both of which tend to move around among the seven with abandon. But to do so is a mistake. The supposition that any two of these construals amount to the same thing, let alone the whole group, is simply false.[22]

---

[21]How do we know that computers are just machines, not conscious, etc.? Only if we have some other account of what they are like, from which such a conclusion could then be derived.

[22]Formal symbol manipulation is explicitly characterized in terms of a semantic aspect of computation, for example, if for no other reason than that without it there would be no warrant in calling it symbol manipulation-to say nothing of there being nothing for it to work independently of.

Clarifying the issues raised in these construals, bringing salient assumptions to the fore, showing where they agree and where they differ, tracing the roles they have played in computing's first century-questions like this must be part of any foundational reconstruction. But in a sense these issues are all secondary. For none has the bite of the reason we are interested in the set in the

The digital state machine construal, in contrast, makes no such reference to semantic properties. If a Lincoln-log contraption were digital but not symbolic, and a continuous symbol machine were formal but not digital, they would be differentially counted as computational by the two construals. Not only do FSM and DSM mean different things, in other words; they have overlapping but distinct extensions.

The effective computability and algorithm execution construals also differ on semantics. Whereas effective computability seems free of intentional connotation, the idea of algorithm execution seems not only to involve rules or recipes, which presumably do mean something, but also to require something like "understanding" or at least "semantic compliance" on the part of the agent producing the behavior. It is also unclear whether the notions of "machine" and "effectiveness" refer to causal powers, material realization, or other physical properties—or, as current theoretical discussions suggest, effective computability should be taken as an abstract mathematical notion. (This is no small question; if we do not yet understand the mind/body problem for machines, how can we expect computational metaphors to help us in the case of people?) The construals also differ on whether they focus on internal structure or on input/output—i.e., on whether (i) they treat computation as a way of being structured or constituted, so that surface behavior is derivative (FSM and DSM), or whether the having of a particular surface behavior is the essential locus of computationality, with questions about how that is achieved left unspecified and uncared about (EC, perhaps ALG).

Not only must the construals be distinguished, moreover; further distinctions are required within each one. Thus the notion of information processing—responsible for such slogans as *The Information Age*, and the link between philosophy of computing and philosophy of information-must be broken down into at least three sub-readings, depending on how information is understood: (i) as a lay notion, dating from perhaps the 19th century, of an abstract, publicly-accessible commodity carrying a degree of autonomous authority; (ii) so-called "information theory," the semantics-free notion originating with Shannon & Weaver (1949), which spread out through much of cybernetics and communication theory, is implicated in Kolmogorov and other complexity measures, and has been tied to notions of energy and entropy; and (iii) the semantical notion of information advocated by Dretske (1981), Barwise & Perry (1983), Halpern (1987), and others.

first place: whether any of the enumerated accounts is *right*.

That question, too, must be addressed: to what jury a proposed theory of computing should be held accountable. But for now let me cut straight to the chase: not one is correct. Forty years after that freshman year in college, I am prepared to argue that, when subjected to the empirical demands of practice and the conceptual demands of theory, all seven construals fail—for deep, overlapping, but distinct, reasons. No one of them, nor any group in combination, is adequate to meet the requirements of a foundational account.

3b Dialectics

To understand the reason for this failure, and grasp the picture of computing that comes out of it, it helps to identify the other two "axes" I use as an initial guide to the territory—both of which cross-cut the first division into construals.

The second involves a set of four "dialectics"—fundamental metaphysical distinctions particularly applicable to "things computational," and necessary to understand if we are to claim to have an intellectual grasp on computing.

1. **Meaning and mechanism:** The first dialectic involves the only *substantial* thesis about the nature of computing I adopt as an investigative guide (again, not as necessarily true of the subject matter, but indicative of issues to be investigated): that, in one way or other, computation involves an interaction or interplay of *meaning* and *mechanism*. That computation is somehow mechanical is reflected in the fundamental effectiveness limits that permeate computational theory and experience. As already suggested, there is disagreement about the nature or origin of this "efficacy"-whether it is (i) an abstract notion, as gestured towards in the notion of an "effectively computable function," taken by logicians and mathematicians to be an entirely abstract notion, unrelated to physical constraint; or (ii) a physical notion, tied to underlying physical law. But as so powerfully demonstrated by Turing in his original paper, that computation is in one way or another limited both in principle and in practice is as deep a fact about

the topic as any that exists.[23]

That computation has anything to do with meaning, interpretation, semantics, etc., is much less widely agreed—in spite of the use of logical language discussed above. I take the semantic nature of computing to be compelling, however, both from the nature of existing theoretical debate and from the character of the phenomenon.

2. **Abstract and concrete:** A second distinction that permeates computing, which has arisen several times already, is that between the concrete and the abstract. What "degree of concreteness" computation manifests, if I can put the question that way, is deucedly difficult to figure out—to say nothing of what, under scrutiny, the terms even mean. Are arrangements of physical things themselves physical, abstract, or somewhere in between? ("I like what you've done with your living room; that's a great arrangement of chairs.") What about abstractly specified concrete properties, or concretely specified abstract properties? Or do the words signify neither a binary distinction, nor two ends of a continuum, but some third possibility entirely? Perhaps they aren't even the right contrast pair. Only the philosophy of computing knows for sure.

3. **Static and dynamic:** Less philosophically vexed, but as crucial to computing, is the distinction between static and dynamic. Programs, it would seem, are static entities, or anyway passive;[24] compilers translate them into other static entities (programs in a lower-level language); interpreters "run them," generating dynamic processes, etc. Or rather: interpreters are programs too; it is when interpreters run that they take programs and generate further process or behaviour—behaviour somehow different from, and yet in other ways coincident with, the behaviour of the interpreter's own running.

[23]Students think Turing is famous because he introduced the notion of a computer, and demonstrated its power. It is important to remind them that he demonstrated both its power and its limitation.

[24]This is not to say that people don't update them—i.e., make better versions of the same program. The identity conditions are complex, but programs certainly exist over time.

terpreter's own running.

Some immediate facts aren't hard to delineate, in other words—even if the saying gets pedantic. Still, the distinction is important, as is the question of whether the way we currently arrange things is necessary or mere historical contingency. Is it just habit, or lack of imagination, that makes us think process specification should take static form? Could a dynamic process itself describe, represent, or specify?[25] If so, surely there could be computational analogues, suggesting that we shouldn't build static specification into our framework. In this and other cases, it is clearly important, in so far as it is possible, to avoid shackling our philosophy of computing to the tiny fraction of possible computational architectures that have so far been explored.

4. **One and many:** Finally, any account of computing worth its salt must deal with a bewildering plethora of distinctions between "things that are one" and "things that are many," such as a single program, web page, file, etc., and multiple distributed "copies" or "versions" of it (a distinction that bedevils software projects and replicated data bases), or the issues that arise when you call a procedure on a matrix: do you pass a distinct copy or, as it is said, "the address," so that there is only one—or is that rather a new copy of "the same address," i.e., two pointers (copies?) that point to the same location, or... Attempts at ultimate clarity can lead to madness.[26]

We speak of many/one relations in many different ways: (i) in terms of types, classes, categories, templates, patterns, schemata, etc., where a single (abstract?) entity is taken to have multiple instances; (ii) as a (concrete?) unit thing with different copies, editions, or versions; (iii) as a set with distinct members; (iv) as a role played by different individuals; and so on. It is far from clear that we under-

[25]Ask a friend to describe a spiral staircase, and watch their hands; you will see a dynamic representation.

[26]It is uncanny how sophisticated expert programmers are at navigating these singular/plural shoals.

stand the distinctions between and among these ways of speaking, and why, exactly, we use one or other in any given case. More seriously, the profusion of possibilities, and the diabolical fact that on reflection whether something is "one" or "many" can seem to be a matter of perspective or even degree, rather than being an intrinsic property (of it? them?), can send metaphysical tremors through the foundations.

Another issue on which to keep an eagle eye.

3c Formality

The third axis has to do with **formality**—one of the most recalcitrant properties underlying the entire field. Somehow or other, it is thought, computation is a *formal* phenomenon, or amenable to *formal* analysis, or *works formally*, or something like that. Just which of these is true, what they mean, and how they relate, are additional issues that any philosophical analysis of computing must investigate.

The near-universal allegiance to formality is both curious and fertile. It is not as if "formal" is a technical or theory-internal predicate, after all-no one writes FORMAL(x) in their equations. Moreover, informal usage seems to range across as many as a dozen meanings of the term: *precise*, *abstract*, *syntactic*, *mathematical*, *explicit*, *digital*, *a-contextual*, *non-semantic*, etc. Far from engendering debate, this profusion or outright ambiguity has probably helped to cement consensus. Because it remains tacit, cuts deep, has important historical roots, and permeates practice, formality is an ideal foil with which to investigate computation.

Once again I will cut straight to the bottom line. The moral for computer and cognitive science here is similar to the claim made earlier about the seven construals: *no plausible reading of 'formal,'* in my view, *applies to the computational case*. Needless to say, negative claims are tricky to prove. To make such a conclusion watertight, one would need both an agreed theory of computing and a definitive analysis of 'formality.' But certainly my investigations have led me to conclude that there is no substantive reading of 'formal' under which concrete, in-the-world computing—**computation in the wild**, as I sometimes call it—is, in fact, necessarily formal. As I put it in another context, "one cannot avoid the ultimately ironic conclusion: that the computer, darling child of the formal tradition, outstrips the bounds of the very tradition that gave rise to it."

4 Results

The issues discussed above are given slightly more treatment in Smith (19■■), and will be explored in depth in Smith (forthcoming). Here, though, it is time to assemble these piecemeal results into the ultimate conclusion, and sketch some of the issues it opens up in front of us.

The bottom line again is simple. Not only do none of the seven construals, understood formally or informally, serve as an adequate account of computing. More seriously, no other construal, of my own or anyone else's making, will serve either. The reason is stark: there is no theory of computing to be had. This, too, is a result that after this long journey I am prepared to claim: *the term 'computational' does not name a property of theoretical significance.*[27] A philosopher who believed in such things might say that computation is not a natural kind, though not being such a philosopher, that is not how I would put it. I would rather just say this: that there is ultimately nothing *special* about computing or computers-nothing to give substance to a theoretical notion of computing-beyond the thesis of the first dialectic: computers are systems or devices that involve an interplay of meaning and mechanism, the best we know how to build. Period.[28]

There is nothing more to say.

The first comment to make—and it should be made straight away—is that this is wildly optimistic claim. Far from being negative, the fact that there is no theoretical substance to something's being computational (i) not only opens up the realm of computing to possibilities not heretofore imagined, but (ii) from an intellectual point of view, makes the development of computers vastly more significant than it would otherwise have been. Sure enough, a number of popular hypotheses end up on the cutting room

[27]Or computing, or computer; it does not matter.

[28]I heard this saying, growing up, from my late father; whether he had heard or created it I do not know.

floor—including the vaunted computational theory of mind.[29] On the other hand, it follows that all of the specific details and understandings and intricacies and mechanisms and architectures developed in computer science are "unrestricted": rather than applying to just a subset of the world's systems, they apply to all systems-at least all that involve the fundamental meaning/mechanism dialectic, which is a lot. So to take just one example: the relation discussed above between philosophy's notions of reduction, supervenience, etc., and computer science's understanding of architecture, implementation, abstraction boundaries, etc., are not just parallel developments. First blush was right: they are theoretical perspectives on the same subject matter. That is why that doctoral dissertation would be important—a synthesis of the two perspectives is mandated by the simple but compelling fact that the subject matters do in fact coincide.

I am not saying that the development of computing is not a theoretical (as well as practical) accomplishment of the utmost magnitude. The discovery of how to arrange physical matter in such a way as to implement digital processes, for example, is a staggering achievement—easily worth a passel of Nobel prizes. Rather, the point is that, instead of being viewed as a restricted species, as is implicit in the idea that 'computational' is a property of theoretical significance, computers are better understood as a site—a "**laboratory of middling complexity**," where we can work out the best understandings we can muster about how meaning and mechanism interact.

5 "Internal" Prospects

What lies ahead?

For discussion purposes, I will address this question using a distinction that is at the very least not black-and-white, and ultimately not one I believe in at all, but which will nevertheless bring some order to the discussion. I will divide my remarks into two categories: (i) "internal" prospects for working out the theoretical and scientific consequences of the views to which I argue, and (ii) more profound "external" implications, regarding our fundamen-

[29] *Of course* we are computers (unless substance dualism is true). We are physical beings, and we mean, or deal with meanings.

tal approach to metaphysics, ontology, and epistemology.

I said that computing involved a mixture of meaning and mechanism.[30] I also said that computer science uses a spate of terms borrowed from logic—including some (*identifier*, *symbol*, *reference*, *interpretation*, etc.) that, in logic, have squarely to do with semantics. It would be natural to suppose that these terms are used to describe the semantic or "meaningful" aspect of computing, rather than the "mechanism" side. Perversely, however, the converse turns out to be true. This is one reason why both Searle and his interpreters get confused. After recruiting semantical concepts (or at least terminology) from logic—terms or concepts that logic uses to analyse meaning—computer science deployed them to study additional aspects of mechanism.

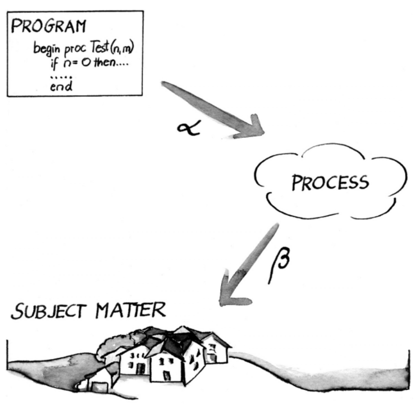


Figure 2 — Program, Process, Task Domain

How this came to pass is a complex story, but the result can be roughly caricatured. Computer science needed to understand the relation between a *program*, taken as a static or anyway passive entity that, plus or minus, both describes and prescribes a "computation," which for present purposes we can take to be the dynamic process that takes place when the program, as we put it, "runs" (see figure 2).

Because of the descriptive element, it was easy to parlay logic's notion of semantics to this purpose, since it took the form of mapping between one thing, "syntactic" or "grammatical" in form, and another, which logic had analysed in terms of a mathematical model. Logic's "semantic interpretation function" could thus be used, in computer science, for the relation dubbed a in the figure: to map programs onto the resulting processes, mathematically modeled or abstractly described. Some work needed to be done.

[30]Admittedly, I haven't defended this statement—merely assumed it as the first dialectic. Given what is said in this and the next two paragraphs, it is not as simple a thesis to defend as one might expect, and so I will continue to do so here. See Smith (forthcoming).

In order to capture the prescriptive part, for example, the interpretation relation needed to be constrained to be effective—in a way that would be perverse, if not outright unimaginable, in classical logic. From computer science's point of view, however, the move made sense. It is this restriction of interpretation to effectiveness that has spawned computer science's obsession with constructive mathematics, intuitionistic type theory, and eventually the development of linear logic, all in service of a kind of ultimately concretized meaning.

The problem with all this, however, is that the genuine semantical question, from a philosophical point of view, is not about the relation between program and process, but between process and world—between NASA's system to calculate trajectories and the orbits of distant planets, for example, or between the proximal results of NATO's early warning systems and the distal fact of whether an intercontinental missile strike has in fact been launched. Schematically, that is, as the figure indicates, we are faced with two relations connecting three realms: (i) a in the figure, between a program P and the process R that results from running it; and (ii) that labeled b, between that process R and the task domain D that the process is about, or that is the subject matter of the information that the process manipulates, or whatever. Computer science has used logic's semantical vocabulary to study the first relation, a, from P⇒R, whereas the tough semantical question is about the second relation, P⇒D.[31] That remains to be theorized. (It is also a tough point to make to computer scientists, since you cannot use any of logic's classic semantical concepts to describe it, as those terms have already all been "used up.")

One more technical result needs to be brought out. Earlier I mentioned debates about the "origin" of the computability limits first demonstrated by Turing, which form the foundations of computability and complexity theory. One of the results that emerges from the analysis of the effective computability (EC) construal is that recursion theory, the notion of computability, etc., turn out to be *mathematical models of physical constraint*. So that

[31]As I once put it to Gordon Plotkin, a programming language semanticist, "I am interested in the semantics of the semantics of programs."

theory, too, is about the "mechanism" side of the substantive dialectic. It is framed as if it were a theory about the computation of numbers, but in fact it is a mathematical theory about reconfigurations of marks (i.e., of physically distinguishable states). This is obvious to programmers, increasingly recognized by computer scientists, and anathema to logicians and recursion theorists. But no matter; it is again a tremendously positive result. The development of the theory of effective computability, once reconstructed as a mathematical theory of causation (below), is another intellectual achievement worthy of several trips to Oslo.

Given these understandings, I would identify the following four projects as the first half of my answer to Floridi's last question, about the most important issues facing the philosophy of computing. As I say, I see these four as "internal" to the subject matter. Even they are huge, and barely begun-but that just underlines what I said above: this is a new field, with most of the work remaining in front of us.

5a Concretisation

If computability and complexity theory is about mechanism and process, not numbers (cf. C4, above), then the entire theory must be recast in concrete terms. To take just one example, consider the infamous equivalence proofs discussed earlier. As currently cast, they claim that, if set up with appropriate inputs, one machine m_1 can "do the same thing"—that is, can "compute the same function"—as another m_2, if given the same input. As I said, that statement relies on a notion of "computing," which for two reasons must now be rejected. First, since 'compute' can no longer figure as a substantial property, we need to cleanse all theoretical statements of its use. Second, to the extent that there was any meaning to the phrasing "machine m computes function f," it is this: given "input" marks j denoting x, machine m can produce "output" marks k denoting $y=f(x)$. "Computing a function" has to do with mathematical entities—that is, with f, x, and y. On the recommended concrete overhaul, the theory would have to *eschew all mention of such entities*, and speak instead about machines and marks: *m*, *j*, and *k*.[32]

This transformation will be massive. Just one example: It is widely understood, and used to encrypt credit card numbers on the internet, that "factoring the products of two large primes" is a difficult task. But factoring is a phenomenon in the realm of numbers, not in the realm of physical arrangements. Moreover, factoring numbers is trivial if numbers are represented in non-standard ways—for example, as lists of their prime factors. So whatever is going on, "what is hard" must have to do with the nature of *numerals*. As such, it deserves framing in terms of marks, directly.[33]

I dub the recommended reconfiguration of the equivalence proof a **motor theorem**, with roughly the following content: Given a motor m_1, and an adequate stock of other passive but perfect parts,[34] one can assemble a configuration *p* of those parts, such that the resulting device, consisting of m_1 appropriately connected up to *p*—a device of potentially Rube-Goldberg complexity-will produce "isomorphic" behaviour, under a hugely broad metric of "equivalence," to that of any other machine m_2 that can be built.

Is the motor theorem impressive? Should we be impressed that such a theorem can be proved? Who knows? Personally, I should not have thought so.[35] Most concrete devices consist of some

---

[32]The theory might (though I do not know whether it will) use functions and numbers to model or measure the concrete phenomena, in the way that physics uses numbers as a basis of measurement; but the resulting theory will no more be about numbers than to say that earth's escape velocity is 11,200 meters per second is a statement about the number 11,200.

[33]What kind of fact is that? It must have something to do with the composition of prime factorization with the inverse of the interpretation function for radix numerals—or rather that composition (or something like it) may play a role in its mathematical characterisation. But what that comes to, concretely—and why such an operation should be hard for a mechanism subject to our physical laws to perform—is going to take some work to figure out.

[34]Friction-free, totally discrete, and idealized in various other ways—these are the consequences or strictures of digitality, perhaps the most significant notion in the entire computational pantheon.

[35]I could never understand, when I first learned about the proofs of universal computability, why, in spite of their surface brilliance, I found them, au fond, to be so fundamentally boring. It took almost twenty years to figure

number of motors, gears, pulleys, containers, pipes, etc., or other forms of motive force. It does not seem to me especially odd that with one motor, of sufficient (that is: indeterminate) size, plus an indefinitely large supply of other perfect, friction-free parts (switches, ropes, pulleys, etc.) one could construct a device functionally isomorphic to any "perfect" device—especially if the metric of equivalence that one is mandated to meet, as in this case, is sufficiently broad. But normative assessment may be personal, and anyway should wait until much more of the reconstruction worked out in detail-something not yet done. The present point is simply that something like the motor theorem, plus appropriately detailed variants for all of the complexity variants, is mandated by the concrete understanding of the notion of "universality" that comes out of this analysis.

In passing, one salutary effect of this "concretization" of our understanding of computation may be to help rid popular culture of various myths about computing, including the ubiquitous belief (false, in my view) that there is a fundamental distinction between the "virtual" world, on the one hand, and the "physical" or "real" one, on the other. Not only are computational processes (and worlds) real; they enjoy a materiality that, while different in tenor than that of our direct experience, is undergirded by the same physical laws and participates in the same temporality.[36]

5b Physical states

Another issue brought onto centre stage by the concrete reformulation of the mechanism side of computing is that of individuating physical states (or perhaps more correctly: physical state types). As Putnam points out,[37] one can claim that a rock implements any computation one wishes so long as one divvies up the physical states "appropriately"—which is to say, in completely

out the answer.

[36]Curiously, given contemporary processor speeds, the material worlds of computing manifest its relativistic character quite directly. It matters, if you are a thread running on a contemporary processor, that a nanosecond is approximately "equal" to a foot—in a way that does not enter our everyday phenomenology.

[37]«reference: the appendix to *Representation and Reality*»

unnatural ways. Deviant physical typing can produce lots of strange results: solving the halting problem, decrypting the most challenging encodings, solving traveling salesman problems in unit time. Of course, such Goodmanesque predicates[38] violate both intuition and utility. But what is an appropriate physical state? No one knows. All we can say is that an adequate theory of meaning and mechanism depends critically on the answer.

5c Process

"Computation," it used to be said, "is mathematics plus time." I disagree with the mathematics part, but the inclusion of temporality as computationally fundamental is unarguable. Process, doing things, what can happen, how long it takes-these are constitutive of the computational realm. Yet I think it is fair to say, almost a century after Husserl, Whitehead, and Heidegger, that we still do not have a theory of process and temporality worthy of the name. The temporally dependent variables of physics, enshrined in the calculus, are one spectacular success story, but they are extremely specific. There is no reason to believe that the "soul of a meaningful machine" will be disclosed through numerically valued measure properties.[39] By and large, computer science does not use them, instead analysing dynamic systems in terms of static structures—programs, inputs, outputs, requirements to be met, conditions to be honoured, contexts viewed as static abstractions. Why do we not deal with time more directly? As I continually ask graduate students, where is the programming language that is as natural for expressing jazz rhythms as Lisp is natural for expressing recursive functions? Should such a language itself be dynamic? Even in cases where we do employ mathematics, process and dynamics are modeled in terms of abstract a-temporal structures. Is the atemporality of mathematics a metaphysical, epistemological, or cognitive necessity? Is a dynamic mathematics[40] an unthinkable possibility?

---

[38]Such as Goodman's famous grue and bleen: green before some time t, and blue thereafter; and its converse.

[39]Experience with computing to date suggests that architecture is a deeper analytic concept than measurement.

[40]Not just dynamic notation, nor a dynamic system about mathematics, nor a dynamic system mathematically modeled, but dynamic mathematics

History, I suspect, will laugh at us three times over: once for our reliance on objects, twice for the skittishness with which we approach relations, and three times for our naïveté about time.

5d Semantics

Finally, and most obviously, we need a theory of semantics—of the "meaning" side of that first substantive dialectic. In spite of the fact that essentially all of contemporary computer science's theoretical apparatus deals only with the mechanism side, it is not pure physicality that we are up against,[41] but, as I keep saying, meaningful mechanisms. Cognitive science recognizes the problem, e.g., in the so-called "symbol grounding" problem. But for an adequate intellectual understanding of semantics, intentionality, meaning, interpretation, reference, modelling, analysis, simulation, etc., we remain woefully in the dark.

It is on the semantic side of the equation that some look to the notion of information. The question is whether the "counterfactual correlation" analysis of information content inaugurated by Dretske in 1981,[42] or perhaps the teleo-semantic variants developed since then, could be pressed into service for a genuinely semantic account of what information is, on which, in turn, a semantically grounded information-processing (IP) construal of computing could then rest. There are huge challenges to these accounts, involving such issues as how to deal with "misinformation" (if that is a species of information at all), avoid various forms of pan-informationalism, etc., but the effort is arguably the only substantial idea in town as to what a semantical account would actually look like.

My difficulty stems from two sources. First, as I will discuss more in a moment, I believe that extant accounts of information are fatally dependent on undischarged ontological assumptions, and therefore cannot serve as a basis for a thorough-going phi-

itself-such as sets changing membership over time, a hole opening up in a topological manifold, or an Abelian group's multiplicative operator losing its commutativity.

[41]Contra such writers as Phil Agre (19■■), who argue that computing is merely a practice of building physical stuff.

[42]*Knowledge and the Flow of Information*, 1981.

losophical investigation. Second, however, and more immediately pertinently, I claimed that not all computing can be understood as "information processing," and therefore that the IP construal will not work as a general analysis (semantical or not).

In a nutshell, the problem is that not all computing is about something *else*, as the notion of information would imply, but actually deals with (as it were) "things in themselves." When my email client tells me that I have new email, it does so by representing that fact (in English) in a window on my screen. When I go to retrieve the email, however, the computer *actually delivers it to me*; it does not merely provide me information about it. Should your message fail to arrive, it is not that I did not receive *information* about your communication. Rather, what is true is exactly what I say: "*I did not receive it*." Information-processing is not a strong or general enough notion to deal with the genuine encounter and engagement in our lives that computers and other meaningful mechanisms (like people) manifestly exhibit.

And so while I am as supportive as anyone of pursuing work in the philosophy of information,[43] and even teach a graduate seminar on the topic, I believe the vast reaches of an encompassing and adequate theory of the semantic dimension of meaningful mechanisms remain largely unexplored.

6 "External" Prospects

Things are serious when (i) the "mechanism" side still needs theorizing, but we lack an account of physical types on which to base it; and (ii) the "meaning" side remains almost wholly unreconstructed. You might also think, since I have spent all these years struggling with the issues, that I would have something positive along these lines to propose.

In a way I do, but at best a cursory sketch. A hint of the reason is contained in the fact that we lack a theory of physical typing. After taking this long journey through the computational landscape, the most sobering result of all is to realise that the most serious problems to be addressed, in developing a theory of meaningful mechanisms, are *ontological*, not just mechanical or seman-

---

[43]Which is not to say that I believe that a theory of information is there to be found. We will have to see.

tical. Even more seriously, the ontological issues involve an inextricable mix of mechanical and semantical concerns. Ontology, that is—by which I mean an ontological account of the entities necessary in order to give an account of computing and other meaningful mechanical systems—is inexorably tied into semantical or intentional or epistemological issues of meaning.

Space prohibits any real defense of this conclusion here, which is anyway too consequential to accept lightly—though the complexity of the subject matter revealed by a close look at all four dialectics may suggest some of the reasons. The magnitude of the impact, however, is not hard to see. Within traditional science and analytic philosophy, it is traditional to accept the following "division of labour": (i) to assume that the "ingredients" out of which an account will be constructed can be distinguished and identified in advance: the objects and properties and relations and sets and states of affairs and so forth; and then (ii) to develop the account of the meaning or semantics or epistemology in terms of them. Ontology, that is, is not only assumed to be separable from epistemology, but to precede it, in some logical or metaphysical sense. For example, if you were to write a computer program to control an elevator, you would first specify the world of elevators, floors, passengers, buttons, cables, etc., and then write the program in such terms. Requirements engineering pretty much assumes this.

The conclusion I have come to is that this approach will not work in the long run. As argued by legions of philosophers of a more literary stripe, ontology (what the world is like, in any intelligible sense) and epistemology (how we take the world to be) need to be reconstructed together. If we use 'metaphysics' to name that conjoined effort, then the answer to the original question about developing an adequate account of computing—i.e., as we can now see, a comprehensive theory of the meaning/mechanism dialectic—involves nothing less than a full-fledged assault on constructing an appropriate metaphysics.

So be it. For a bit more discussion, see Smith (20■■) and for an inchoate stab at what such a metaphysics might look like, Smith

(1996).[44]

In a way, the conclusion is not too surprising. Think about persistent online worlds—and the vexed questions that come up about whether avatars, into which people pour thousands of hours of devoted labour, are: (i) prostheses, (ii) beings in another reality, (iii) names or representations, (iv) identical to the player (warranting the common use of the term 'I' in such phrases as "I am going to slay the demon"), etc. Is a "crime" committed in such a world as innocuous as those investigated by Hercule Poirot, as serious as a "real-life" version, or something in between? And if the last, what existential or ontological conception of what is going on is strong enough to found such a ethical regime?[45]

What is striking is that ontological challenges are not just "out at the level of use"—i.e., where people manifestly enter the picture. They permeate the entire subject matter. Even accounting for the identity conditions on a file outstrip the capacities of any known account. Is/are the file in the file cache, the one on the backup tape, the one I sent to you by email, the same? Or a copy? Sometimes it is convenient to think of it one way, sometimes the other. But is identity dependent on how we think of it? Maybe-but that's no innocent complication.

Similarly, any simple distinction between a sign and what is signified (name/named, description/described) is too blunt an instrument to deal with even simple computational systems. Does the ASCII version of a visual program *represent* the program, or *is* it the program, or is it a *translation* of the program? And what about the file I thought I lost, last night—but then realised that I did not, after all, because I had made a backup just a few hours before? Sure, I lost a few hours of editing—but still, I found "it." Says who? Says I. Which means that "the file," for me, as a singular term refers not to a particular physical copy, or even to a simple type of physical copy, or perhaps even to a more abstract single individual that the physical copy "realizes" (what's the differ-

---

[44]Note too the extent to this conclusion, wrung from an allegedly technical subject matter, resonates with claims made in feminist epistemology, science studies, and other poststructuralist initiatives. Not evidence for anything, exactly; but not sheer coincidence, either.

[45]See Kevin Eldred's forthcoming doctoral dissertation for an in-depth analysis.

ence between those two ways of putting it, anyway?), but to something yet more abstract—something whose identity conditions are more like the identity conditions on proofs we rely on to decide whether a young mathematician should be awarded tenure for their discovery of a "new proof" of a known result. In both cases, I would hazard, identity cannot be established independent of meaning—and perhaps even purpose. A sensible enough claim—but again a seriously expensive metaphysical result.

The problem, of course, is that once this gate is opened, and we take a step through, we enter a terrain of virtually unlimited grandeur and scope. The foundations of a great deal of what we consider science fall away, replaced by metaphysical and epistemological questions of almost unutterable consequence-and, needless to say, surpassing difficulty.

Objects, and in fact all of commonsense ontology, need naturalising, for starters—as much as any semantical or intentional notion. It is not obvious where semantical or intentional notions will come from, either, since there will not yet be any stable ontology on top of which to build them. The semantic notion of *information*, for example, will not be able to speak of counterfactual dependencies between entities—at least not if it is going to play a foundational role on which those entities are going to depend. Or perhaps objects and information will arise together, with objects being patches of the world understood or "parsed" ("coarse-grained," as those in AI would put it) at a level or degree of abstraction that facilitates the kinds of counterfactual correlation that in turn allows us to track them. That is: maybe it is not that reference allows us to refer to objects, but that objects are that to which we are able to refer? Who knows? It is not a crazy idea, even if how one would make good on it is not exactly obvious. It is not just objects that we need, either; the same goes for properties, relations, sets, etc.—to say nothing of "truth-makers" for non-conceptual content, be that Strawsonian feature-placings or whatever. Norms, too, or something to fill their role (perhaps fundamentally dynamic?) should be added to the list.

And so on. All I want to emphasise here, however, is the role of computation in this vast enterprise. For many years metaphysics has been viewed with huge suspicion—one of the few things on which both modernists and postmodernists agree. I am claiming, in contrast, that we are not going to understand computing—or meaning and mechanism more generally—unless and until we get over that suspicion, and take up the metaphysical gauntlet for real. Crucially, as I will argue elsewhere, that does not mean we need fall prey to any of the ways of doing metaphysics that have convinced a few centuries of philosophers that it is a hopeless and hapless enterprise. Interestingly, moreover, but consequentially, and something else that will need careful explanation, I believe that we can do so empirically, using computation as our laboratory—and not just metaphysics, but an indissoluble mixture of metaphysics, ontology, and epistemology; and not just theoretically, either, from an armchair or with Platonic detachment, but in an engaged, constructive, probably quite messy and concrete way.

Computers are not a subject matter, but as I said above, laboratories of middling complexity—vastly more complex than the atoms and frictionless pucks and pendula of simple mechanics, but vastly simpler than anything even reminiscent of human cognition. Whereas I identified four major challenges for future research "internal" to the study of computation, this is the one challenge—or rather, opportunity—I would name from the more serious and more sobering external perspective: that we recognize the first hundred years of computing as something of an Alchemical precursor to the intentional or meaningful sciences, and, with unswerving focus, parlay our computational experience into a finally successful metaphysics.

One final point, to bring the story full circle.

The term 'material' is the adjectival form of 'matter' in both of its senses: 'matter' as a noun, as in "slurries are a form of matter studied by geologists"; and 'matter' as a verb, as in "it doesn't matter whether you call me or not." When we speak of material objects, most people assume we are using the form derived from the *noun*—that a material object is something that weighs something, that occupies space, into which you might bump in the night. A material argument, however, of the sort a judge might deny you had raised, is of the other kind: an argument that does not matter (to whatever issue is at hand). How the two forms of 'matter' came apart could be argued, but suppose we lay it on Descartes. Then one way to describe the project laid out above is that of developing an understanding of a material object as a "spatio-temporal chunk of reality that matters"—thereby healing a temporary 300-year rift between *matter* and *mattering*.

And with that we can finally answer the questions with which I started. Can studying computing help us do to the richness and complexity of the human condition? Yes, but not in the way that I thought, back then. Computers can help by serving as a laboratory in terms of which to explore issues of intentionality, embodiment, semantics, meaning, mechanism, interpretation, etc., so long as we let go of the conceit that they are computers—or anyway, the conceit that that their being computers is theoretically relevant. Can they be understood with all the power and insight and elegance of the sciences? Well, no—not if elegance requires formality. But formality has lost its sheen, for me at least, and I find more reward in exploring the metaphysical depths that these seemingly innocent devices have opened up in front of us. So yes, in a more grown up way—a way I was nowhere nears up to, at the time. Finally, as an added bonus, the time was not wholly "off" from physics, after all—as maybe my professor knew, and anyway is betrayed in the etymology. Maybe metaphysics is just physics, pushed harder. Hard enough to unleash meaning.

Time for second semester.

— Had this page been blank, that would have been unintentional —

2 — Requiem for the Computational Theory of Mind[†]

Preamble

Thanks to Terry Horgan, for that very gracious introduction. Thanks too to Güven Güzeldere and Stevan Harnad, for an excellent anniversary program. Thanks to the ever-gracious Betty Stanton, also, for taking the time to return to what we all know is her real home. And thanks to Ken Taylor, for the heroic job of local arrangements.

I also want to thank CSLI—the Center for the Study of Language and Information—for their financial support and help in hosting this meeting. I cut my teeth at CSLI; it is good to be home, and to give it public credit.

At first, in fact, I wondered whether I should not speak on a topic related to CSLI's theories of situated language. I had in mind the double-indexical theory of truth, which mysteriously seems to have escaped John Perry's attention: why we say "here, here"[1] to mark propositions with which we agree; "there, there," to mark those with which we do not. You might think *proximity* signals *approval*. But that cannot be right, since "now, now" indicates protestation, not approbation. The phenomenon must be related to that famous Morganbesser quip: that "yeah, yeah" shows how

[†]Presidential Address, 25th Anniversary Meeting, Society for Philosophy & Psychology, June 21, 1999. The talk is reproduced exactly as given, orally, complete with introductory remarks, in order to preserve the decidedly informal character of the (after-dinner) setting.

[1]It was fortunate thing that this was a talk.

two positives can make a negative. Beyond that, illumination awaits.

I store my notes about these deep issues in a file in the corner of my hard drive, under the heading "Questions of great philosophical significance in which I have never been able to interest any philosopher." Here is another. Suppose one refers to the sun. How long does it take reference to get there? At this very institution, I once asked the question of no less a luminary than Alonzo Church, who (without batting an eyelash) said that reference "travels at the speed of logic" (I take it that is supposed to be fast). Problem is, I thought physicists had dispatched the notion of simultaneity to the same dust heap as the luminiferous ether. So there is an issue to be resolved. Is Brentano's arrow of directedness genuinely superluminary? Do we need Bell's theorem? Maybe Penrose missed his calling. Haven't you sometimes wished he had studied semantics?

Fortunately, I will not talk about these issues, either, since I know nothing about relativity. Rather than speculate about things I do not understand, in fact, I will confine my remarks to something I do. That is: I will return to my home field of…*psychology*.

Now you may not have realised that I am a psychologist. Neither did I, till I was so generously offered this position. I have certainly never taken any psychology courses—as my colleagues at Indiana will readily attest. But it is a tradition for the presidency of SPP to alternate, with philosophy presidents on even years, psychology presidents on odd. Last year (1998) we had Bob McCauley; next (2000), Terry Horgan—both with intensionally-correct PhDs. The odd years are reserved for us scientific pretenders.

For that is what really matters: not what I am, but what I am *not*. I fall in the "other" category. And that seems just right. For I have always viewed SPP as something of a philosophers' "at home"—a party, hosted by philosophers, for scientists interested in the mind. The philosophers have a chance to find out what is actually true; we scientists undergo a little conceptual grooming. And the food is great.

From that perspective, coming from computer science (my real home) seems appropriate enough. For nothing, arguably, so unified cognitive science over its first twenty-five years, and the philosophy of mind along with it, as the computational theory of mind. Or nothing did unify it. Paradoxically, as you have undoubtedly noticed, computational presence in cognitive science has recently been on the wane. Where I teach, telling students to read about the computational theory of mind these days is rather like recommending they listen to Mantovani (or even Monteverdi). And I am not even fifty.

So that is what I really want to talk about tonight. What happened to the twenty-five-year-old conversation between computing and the philosophy of mind?

1 Talking to philosophers

I will get to the content of that conversation in a moment. First, though, a word about talking to philosophers.

As anyone from the provinces—oops, sciences—will tell you, talking to philosophers is a little strange. They talk a lot, first of all; that much is hard to miss. They use odd constructions—hypotheticals three layers deep, with patently untrue premises, somehow convinced that the conclusions still matter. And they talk fast. Except, as I discovered, it s a very special version of fast. Not just 217 words per minute (I tried that, once—no one understood a word). Rather: *fast at clause boundaries*, when the situation is vulnerable, lest someone take the floor. As long as you are manifestly *mid-clause*, however—grammatically safe from coming to the end—you can slow down to a reasonable speed, so that everyone can follow. Arvind Joshi should study the stuff.

More seriously, different norms apply. In computer science, papers *report* on research; in philosophy, they *are* research. When I first showed a paper to a philosopher, I felt as if I had given them a map of buried treasure—only to have them respond: "Great; such an exemplary map! Such good lines. All the labels are so well arranged. You can see these nice paths." Good of them and all…but, well, it was the *treasure* I was trying to interest them in. This difference must in part be because computer science is essentially engineering. Its methods are neither theoretical nor empirical, but *synthetic*. Never forget this fact: it is what we build, not what we say, that matters.

All these things can be learned, and partially accommodated-though as we will see, they have more impact than one might expect. And they are laced with another distracting issue: of vocabulary.

As everyone knows, philosophy, like physics, uses ordinary English for its technical terminology-a trap for the unwary. As do all other cognitive sciences. It might not be so bad, except-have you noticed?—they all use the same 200 words, but with different meanings. *Object, concept, variable, function, model, class, meaning, semantics, reference, identifier, interpretation, binding, type*—every one of these terms has a demonstrably different meaning across SPP's constituent fields. It makes me wish that interdisciplinary conferences would provide U.N.-style instantaneous translation into different disciplines. Go to a psychology talk, put on headphones, dial the switch, and presto!—you would hear it translated into linguistics. Feminist cultural theory, translated into analytic epistemology! On-the-fly translation between analytic and continental! Twenty years ago, not entirely facetiously, I suggested to Harry and Betty that Bradford Books publish a dictionary of these 200 words, with entries for how each is used in each cognitive discipline. When we were setting up CSLI, we even considered appointing a "technical term librarian": anyone planning to use a term in a technical sense would be required to sign it out-and promise to return it on a pre-assigned date.

The vocabulary problem is particularly acute between computing and philosophy, for an interesting historical reason. Computer science took many of its technical terms—*language, semantics, variable, name, identifier, syntax,* etc.—from philosophical logic, but then, like any good science, proceeded to change their meanings. This has had the unfortunate consequence of allowing some philosophers to think they understand what computer scientists are saying. Consider Searle. While I doubt that he has spent much time hacking C++ code or writing Java applets, he nevertheless thinks he knows what computation is—in spite of the fact that I do not know a single programmer who thinks he even remotely "gets it." Why does Searle think this? *Because we describe our machines using his words.* Everything we say; *i*t all sounds so familiar, to him. *Searle would be right,* I tell my students, if computer scientists' own descriptions were interpreted on the assumption that the words mean what they do in logic (which is where we got them from).

And his case is not unique. Unrecognised overlaps in technical vocabulary continue to sew conceptual misunderstandings about the foundations of our interdisciplinary project. A critical example is going to play a role in tonight's story, having to do with the word 'computation.'

2 The computational theory of mind

So let's turn to that. As I said, the real excitement in cognitive science, back in the 1970s, had to do with this thing called **the computational theory of mind**. You all know the story: a set of internal words or sentence-like tokens—items in a language of thought—that (i) carry content (mean something) about situations or states of affairs in the world, but (ii) are manipulated internally, not in virtue of that content, but instead in virtue of their abstract shape or "form." As in logic (too much as in logic, as we will see). This catechism was chanted in the halls of every self-respecting philosophy and cognitive science department.

To understand this hypothesis, and the excitement it generated, we have to back up a bit.

The basic situation everyone was wrestling with—psychologists, philosophers, computer scientists alike—had to do with intentionality. And intentionality, on widely-shared metaphysical assumptions, involves an interplay between **meaning** and **mechanism**. This dialectic had occupied people's imagination for centuries: how a "mere mechanism"—naught but a lump of clay—could sit up and think: reason, wonder, dream, even be conscious.

Crucially, minds and computers are both intentional, and so both instantiate the meaning/mechanism dialectic. Take the mechanism side. That computers are mechanisms is obvious; that minds are mechanisms did not use to be obvious, but most academics believe it now. Similarly on the meaning side. Once again, it is obvious that minds involve meaning (meaningfulness is a prime candidate for the "mark of the mental"). That computers traffic in meaning is equally evident, if you look at practice. Computational discourse is rife with intentional words: *information processing*, programming *languages*, *data* bases, *knowledge represen-*

tation systems, *simulations*, *models*, *correctness*, and so on.

Moreover, it is only because both minds and computers are intentional that there was any reason to suppose the computers might be intelligent—that computers were relevant to mind, that we might be computers. In most other ways, after all (physical make-up, price, eating habits, etc.) computers and people are almost completely different.[2]

But the idea that excited people, back in the 1970s, was not simply that computers, like us, were intentional—that we were building, not just growing, intentionality. What set the world on fire was something more pointed: a belief that computers were *special*—that Turing (or someone) had discovered a distinctive way of merging meaning and mechanism—a way that had no one had understood before, had now been discovered, was embodied in Turing machines and Eniacs and Pentiums, and might even—who knows?—be embodied in minds. It was this "something special" that was called **computing**.

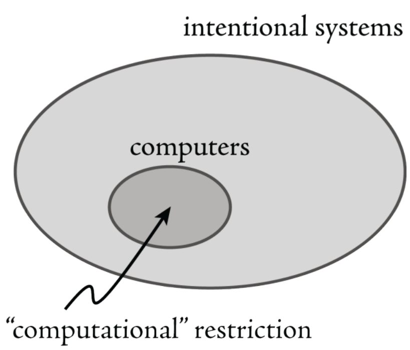


Figure 1 — Computing as special

It is critical to understand the importance of the "*specialness*"—this idea that there was a distinct, identifiable, scientific property "being a computer" or "being computational," which was: (i) uniquely true of all and only computers; (ii) would play a role in scientific laws; and (iii) might be relevant to understanding the mind. Fodor is perfectly clear about this in "Methodological Solipsism."[3] That people were general intentional systems (or even, somewhat more narrowly, general representational systems) he took to be obvious-and boring. The computational theory of mind, on the other hand, he took to be non-obvious, and exciting. It was non-obvious and exciting because there was something distinctive

[2]I once gave a talk arguing that the most important difference between computers and people was that computers have back-up tapes. It seemed innocent enough—until, when I arrived at the occasion, I discovered that it was being hosted at a senior citizen home.

[3]Fodor (19■■).

about being computational. *Computer science had had an idea*. That was Turing's brilliance.

Perhaps a picture will help (figure 1). The larger circle is the space of all (perhaps all possible) intentional systems; the smaller one is the space of all computers or computational systems. The "computing" idea we are talking about is the (characteristic) property of the subset.

So that was context in cognitive science mid 1970s—which, as it happens, was when I went to graduate school at MIT.

Now you have to understand what it was like to come to philosophy from computing, only to discover it was all excited about your field. It was a little like reading about your area of expertise in the New York Times: thrilling headlines—but a bit queasy-making, as soon as you started to read. In particular, it was not clear what philosophers actually knew about computing. In "Tom Swift and his Procedural Grandmother,"[4] Fodor said that "*providing a compiler for a language is tantamount to specifying a [procedural] semantic theory for that language*." *Compiler*? Did he maybe mean *interpreter*? (Compiling is just translating; it does not make things actually happen.)

So I went to talk to Fodor. "How do you know that computers work in the way you describe?", I asked. Now Fodor, modulo a certain bluster, is a pretty honest guy. "I made it up!", he said. "Look," he went on, "you are the computer scientist; it works in whatever way you think it does."

"If your understanding rests on my understanding," I thought, "then you are in trouble."

Seriously, here was the situation. Computer science—or computational practice, or something like that—was supposed to be supplying cognitive science with an *idea*: an account of the property "computational," on which the computational theory of mind could rest. The catechism—the claim that computation was formal symbol manipulation—was, I took it, philosophy's best attempt to describe that property.

I was supposed to represent computer science. Sure enough, I

[4] Fodor (19■■).

knew how to design programming languages, write compilers, build operating systems. But an idea? An idea good enough for philosophy? That is serious. I wasn't at all sure about that.

3 Formal Symbol Manipulation

So while most people in cognitive science busied themselves with their main project—to find out whether formal symbol manipulation was true of mind—I set out on a parallel, side path: to find out *whether it was true of computers*—specifically, what I call computation in the wild:[5] concrete, real-world systems. That is, to put it as bluntly as possible: I set out to determine the truth of the following radical thesis: *the computational theory of computing*.

The natural way to do this, you might think, would be to ask computer scientists, or look at computer science—at the discipline that builds and studies and theorises computing—and see whether it takes computing to be formal symbol manipulation. Not that their answer would necessarily be definitive; computer science could be wrong. But it is not often that a whole field is wrong. Anyway, it seemed like a good place to start.

So I went to computer science, and "asked it," as it were, whether it thought computers were formal symbol manipulators.

Computer scientists had never heard of formal symbol manipulation. They did not know what I was talking about.

Now this is a bit tricky, so bear with me. Computer science had heard the words 'formal symbol manipulation.' In fact they used the words—or rather, correlative words from logic—syntax, semantics, model, interpretation, completeness, etc. But remember what I said about vocabulary. They used the same words, *but they meant different things!* On the surface, it *sounded* as if we were talking about the same things. But if you pushed, in order to make sure that we were really on the same wavelength, it turned out that they were talking about something else.

What computer scientists had heard about was Turing. And Fodor was certainly right about one thing: Turing was one smart guy (I would definitely recommend him for tenure). But as for

[5]With an apology to Ed Hutchins.

understanding formal symbol manipulation, I sometimes think Turing was the villain. Because if you look hard at the original Turing papers, he almost drops the ball with respect to the critical issue for formal symbol manipulation: namely, the *symbol* part, the part having to do with *representation*. And make no mistake: formal symbol manipulation is an idea about the processing of semantically interpreted structures (as I have said elsewhere, if you are not interested in semantics, then you should call your thesis "stuff manipulation"[6]). Turing's 1937 paper[7] opens with laudable representational clarity: "the computation of functions whose representation as a decimal can be calculated by finite means."[8] But by the time you get to the middle—to all that stuff about universal machines and modelling and quadruples and what functions can be computed and so on and so forth—the difference between numbers and numerals has been stirred into oblivion. And as everyone knows, if you cannot tell a number from a numeral—if you do not mind your "uses" and your "mentions"—you definitely are not going to be invited to the next philosophers' party.

No, the theory of Turing machines, I have come to believe—and of effective computability, and complexity theory, and just about all the theoretical edifice of theoretical computer science—is *not about formal symbol manipulation at all*! It is about something else. What else? I will get back to you in a minute on that.

But back to the main plot. Here I was, trying to figure out whether computers were formal symbol manipulators. Theoretical computer science was no help. And so I had to do my own empirical study.

It sure took me long enough (going on thirty years). But I can now tell you the answer.

The answer is no.

There are many reasons. Tonight I will just quickly mention two.

The way to figure out whether computers are formal symbol

[6]Smith (19■■).
[7]Turing (19■■).
[8] «check, and refer»

manipulators is to see how formal symbol manipulation is *special*—and then to see whether real-world computers are special in that way. That is, one needs to see what essential restriction formal symbol manipulation places on general intentional systems—what exactly it is that formal symbol manipulation claims distinguishes the smaller inner circle from the broader outer one—and then determine whether the resulting restricted class coincides with the class of computers in the wild.

Now formal symbol manipulation, interestingly enough, is not itself a formal thesis, and so it is not exactly clear what it means. But after long study I have determined that, at least at a minimum, it places the following two restrictions on intentionality:

1. That there be symbols; and
2. That they be manipulated formally.

I will look briefly at each.

3a Language-like tokens

Start with the symbols. In this context, symbols are taken to be elements in an explicit, compositional array of language-like (representational) tokens. And on this characterisation, it sure seems as if computing involves symbols. Just look at an ordinary program-in, say, an Emacs buffer:[9]

```
if empty (paper-tray)
   then display-user dialogue-box("out of paper")
   else start-copying ... whatever
```

This looks paradigmatically symbolic. But appearance is distracting. There is no more reason to suppose that computing involves symbols, from looking at a program, than to think that car engines combust symbols, because there are symbols in the blueprints used to control the automatic milling machines.[10]

Sure enough, program ingredients are explicit and language-like. But *programs are completely irrelevant for psychology*. Programs are a convenience we use to constrain universal, general-purpose computing engines in order to flexibly implement fixed, task-specific architectures. Sure enough, you can mix and match

[9] «ref»
[10] See Smith (19■■).

identifiers and so forth, when writing a program-as compositionality and productivity require. But once the writing is done, the resulting program is held constant, during its execution[11]—invisible to the process it describes. It could as well be eliminated (and often is eliminated, by the compiler).[12]

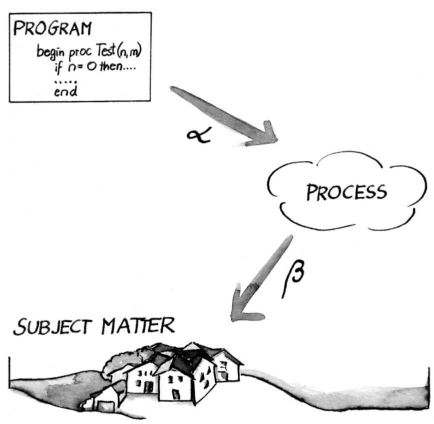


Figure 2 — Program, Process, and Task Domain

What matters to psychology is not *programs*, but *the processes and architectures those programs specify*: how they operate, interact with the world, modify their internal state (see figure 2). And the state-bearing ingredients inside such processes are called *data structures*. So the question we need to ask, with respect to the computational theory of computing, and thus *before* we take up any subsequent question about the computational theory of mind, is the following one: *are data structures explicit, language-like tokens?* The answer, in general, is *no*. (Do not be distracted by the fact that they are given names in programs. That is irrelevant; those names are not themselves the data structures; they (like all names) are *names*: they *denote* data structures—just as names of engine parts denote pieces of steel.) In the vast majority of cases, data structures are highly-constrained, purpose-specific, and non-generic. They have none of the properties that Fodor, Pylyshyn, Evans, van Gelder, and others think symbols have: of modularly designating arbitrary predicates or relations, that, modulo certain appropriateness conditions, can be algebraically recombined in systematic and productive ways.[13]

(In passing, I might note that the *program-process* relation, labelled a in the diagram, is what computer science calls "semantics." What we in computer science are interested in is the rela-

[11]Except for self-modifying programs, of which there are virtually none.
[12]Taken from Smith (19[redacted]), page [redacted].
[13]See for example Fodor & Pylyshyn (19[redacted]), Evans (19[redacted]), and van Gelder (19[redacted]).

tion labelled b in the diagram, between the thereby-specified processes and the worlds or task domains that the processes are about. To distinguish, we might call these program semantics and process semantics, respectively. This is one example of the sort of terminological confusion I mentioned at the outset.)

Admittedly, some computer systems employ symbols: theorem provers, expert systems—plus of course interpreters and compilers. But those, I venture to say, comprise no more than one percent of the programs that are written. Compared to them, there are hundreds of billions (if not trillions) of lines of commercial code that are *not symbolic*. These real-world programs—or rather, real-world processes—are tacit, implicit, non-conceptual. (Just try asking Windows NT why it put up a blue screen of death. Or Unix whether it likes thrashing so much. Or Deep Blue whether any of the arrangements of pieces of any of the configurations it examined reminded it of the face of its designer.)

And so restricting intentionality to symbol-using systems is vastly too narrow to capture computation in the wild.

3b Independent of semantics

OK, so that is the first reason that real-world computing is not formal symbol manipulation: there are no symbols (in the required sense). The second reason is that they are not manipulated formally. That is: data structures are not—at least in general—manipulated independent of their semantics.

Now that phrase—"independent of semantics"—is as recalcitrant as any in the philosophy of mind. Let me just say this about it. The systems where it is motivated—where there is some reason to think it is true—are those systems that are *entirely disconnected or detached from their subject matters*: theorem provers proving theorems about inaccessible cardinals, NASA simulation systems figuring out whether a rocket will venture outside the solar system—things like that. Where the independent-of-semantics mandate is *not* reasonable is where systems are thickly engaged, causally, with their subject matters. This is well-recognised: Devitt, Levine, Anthony, and others have pointed out that the formality condition is difficult even to *understand*, and almost certainly not true, in cases of transducers.[14] And there is Fodor's own telling comment: "Please don't ask me about transducers; I am particularly busy just now."[15]

So the formality condition is (arguably) true when systems are disconnected. Is that a necessary condition on computation in the wild? What about those trillion lines of commercial code?

The answer is interesting. By far the majority of those programs that are *good* programs—situations where computing systems have proved resilient and successful—are cases where computers traffic, directly, in their subject matters: network routers, compilers, window systems, document processing systems, email programs, web browsers, and so on. *Especially in those situations in which it is most* successful, computation in the wild, far from being detached, is highly involved in its subject matter.

In sum: once we set overlapping vocabulary aside, and look the subject matter squarely in the face, we are forced to conclude that real-world computing is not formal symbol manipulation.

It is all a bit ironic. It turned out that what I was discovering, on my side path, was the same lesson the main body of cognitive science was discovering, about people. Just as they were abandoning so-called "computational" (i.e., abstract, disconnected, purely logical) models of mind, in favour of embodied, engaged, interactive alternatives, so too I (and a lot of other computer scientists, I might add[16]) were, in our own way, *rejecting abstract, disconnected models of computers,* in favour of—you guessed it!—embodied, engaged, interactive alternatives.

Now this raises an interesting possibility. You might think, given all these results, that my brief would be to resuscitate the computational theory of mind. Maybe we can have a new computational theory of mind, one framed not in terms of the worn out formal symbol manipulation idea, but in terms of a new idea—of dynamic, embodied, real-world interactive computing. Tacit programming. Ready-to-hand software! Whatever. Then

[14]«Refs»
[15]«Ref; check with Murat»
[16]E.g., see Stein (19■■)

gramming. Ready-to-hand software! Whatever. Then the computational theory of mind could be true once again.

As is evident in my title, however, I have come to bury the computational theory of mind, not to praise it. So I still have some explaining to do.

4 Computing

I have said that formal symbol manipulation does not work as a theory of computing. But I also said that it was not computer scientists' idea of computing, anyway. So what is their idea. What does computer science take computers to be?

It turns out there are several answers—several standard candidates. They are all familiar: *information processors*, *digital automata*, *rule-governed systems*, *rule-following systems*, *physical symbol systems*, etc. Ideas are not the easiest things to count, but by my lights there are anywhere from half a dozen to a dozen different such stories. Some (even many) people think that these stories are all the same—because of various equivalence proofs. But that is an elementary, if common, mistake. Those equivalence proofs are extraordinarily coarse-grained, and *gloss over everything that matters for a theory of mind*. At the level we care about them, the ideas all differ, both intensionally and extensionally. A Lincoln Log contraption (so long as its parts were not information-bearing), would be a digital state machine, but not an information-processor. If continuous representations are possible, which seems not only possible but likely, then a formal symbol manipulation system could fail to be a digital state machine. And so on.

My real project, therefore, has been to assess not just the formal symbol manipulation idea, but this whole suite of other alternatives—ideas I call '**construals**' of computing. The plan, for each, has been to understand where it came from, what it says, and—crucially—whether it is true of real-world computers. That is, in terms of our graph, I have wanted to see whether any of these other alternatives could do better than formal symbol manipulation idea did in restricting the class of intentional systems to all and only computers.

I will not bother you with the details. Let me simply cut to the bottom line. There are three results, of increasing strength.

1. First, *none of these other standard construals works, either*. No one alone, nor any group in combination, is strong enough to delimit the proper computational subset of the full space of general intentional systems.
2. Second, I was not able to find or make up any non-standard construal that worked, either. So in the end my search for a conceptually sound and empirically adequate theory of computation-in-the-wild failed. After 30 years of looking, I have come up empty-handed.
3. Third, I did learn something, though. Not only did I fail; I had to fail. I had to fail because there is no such theory. There never will be such a theory. There is no theory, because *there is nothing there to have a theory of*.

There are not any computers! It is all a hoax, perpetrated by Bill Gates.

Seriously, of course there are computers. What I am saying is that the property computational or being a computer does not pick out a natural or scientific kind. It is not a property that will figure in scientific laws, or underwrite any deep or interesting scientific generalisations. Nothing of scientific interest holds of a computer in virtue of its being a computer, or of anything at all in virtue of its being computational.

In the end, computers turn out to be rather like *cars*: objects of inestimable historical and economical and social importance, the existence of which has and will continue to transform our lives. Lots of theories apply to cars: physics, thermodynamics, ergonomics, ecology, and so on. But no one writes equations with CAR(x) in them,[17] and very few universities have departments of automotive science. (I am embarrassed to say that MIT, where I came from, did have such a department—but hey, it is an engineering school, and anyway, my understanding is that they have since thought better of it, and shut the place down.)

For "computational" to be a scientific property—for there to be a theory of computation—as I have said since the beginning, there

[17]Except of course for John McCarthy and Doug Lenat.

must be something *special* about computers. It is that "specialness" that a theory should capture. And the result of my analysis—the reason why 'computational' is not going to survive as a scientific property—is that *there is not anything sufficiently special*. In spite of the press, real-world computers turn out not to be necessarily formal, or necessarily digital, or necessarily abstract, or necessarily context-independent…or necessarily any other property that has been suggested (or that I have been able to find). Rather, what computers are are dynamic, intentional systems-socially constructed[18] intentional artifacts, the best, at any moment in history, that we know how to build.

Period. No more and no less. That is all there is to say.

Now, for a computer scientist, you might think that this is a dismal result. On the contrary, however, I believe almost exactly the opposite: *the lack of a theory of computing is the most positive, optimistic, exciting possibility that anyone*—even the most unregenerate computational triumphalist—could possibly hope for. I do admit, though, that seeing things this way requires a change of perspective. That change of perspective is something I now want to explain.

On the old view, computing was taken to be an autonomous, distinctive subject matter—warranting its own theory, its own academic department, its own vocabulary. A subject matter whose name could be chiseled into the facades of twenty-first century university buildings, alongside physics and mathematics and literature and maybe even economics. What I am claiming is that it is no such thing. Rather, what computing is is an *historical occasion*—a laboratory of middling complexity, between the frictionless pucks and inclined planes of physics, and the full-blooded complexity of the human condition, in which to see issues of meaning and mechanism play out.

To put it in a slogan, computation is a **site**, not a **subject matter**.

[18]Socially constructed not as a meta-philosophical standpoint, but in the literal sense of being constructed by groups of people.

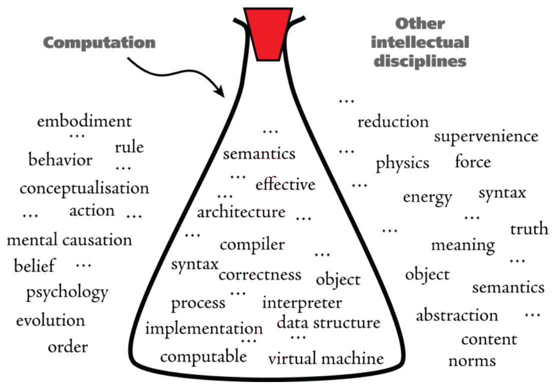


Figure 3 — Computation as closed

A pair of figures may indicate why this is a good result, not a bad one.

Figure 3 indicates the traditional view we have been working with, in which computing is taken to be an autonomous discipline—a subject matter with its own theories, vocabulary, insights, phenomena. Instead of drawing it as a simple circle, I have indicated it this time as a stuffed Erlenmeyer flask. Inside are the properties and relations classically taken to be computationally specific: *implementation*, *abstraction*, *effectiveness*, *complexity*, and so on. Outside, this time, I have put the rest of the intellectual map: philosophy, psychology, economics, whatever.

The picture I am recommending is given in figure 4. What I am indicting, as the result of my 30 years of study, is not the contents of the flask, but the flask itself—the bottled-up corker of an idea that there is an interesting property of "being a computer" that separates what is inside from what is outside. So what I have done is to peel back the flask—undo the conceit that computational things are theoretically distinct. What this allows, as the picture shows, is that phenomena that have been studied as computation-internal are *now allowed to join up with their appropriate partners that have been thought to be computation-*

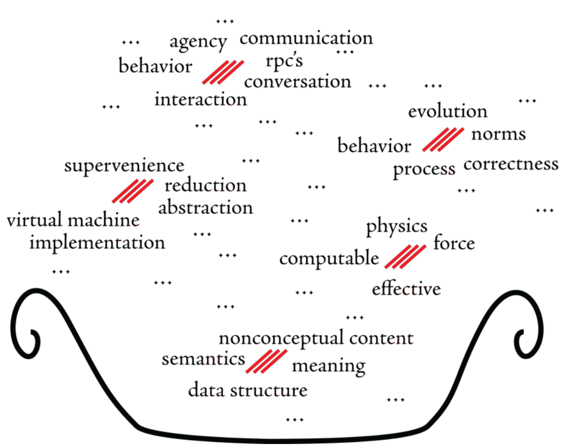


Figure 4 — Computation as open

external.

Take just one example: the question of how a system, described at one level of description, relates to that same system, described at another (say, lower) level of description. In computing we call this *implementation*; there is perhaps no more critical and powerful a notion. All sorts of issues arise: of *virtual machines*, of *abstract data types*, of *implementation boundaries*. There is very interesting work going on at the moment breaking down the idea that such abstraction boundaries are opaque, instead seeing how properties of the underlying implementation inevitably "shine through" at upper levels.[19] Various labels are used for this- "grey box" or "glass box" abstraction, for example, in place of the prevailing idea of "black box." But of course the very same issues are studied outside computing—for example in philosophy, under notions of type- and token-reduction, local and global supervenience, non-reductive physicalism, etc. And there is interesting work going on there, too—for example in the literature on emergence. What the idea that computation is a distinct phenomenon has done is to keep these two discussions apart. Somewhere in another corner of my hard drive I have another list: titles and abstracts for "PhD theses needing written."[20] One of them is on bringing these two disciplines together—which after all are talking about exactly the same thing.

In sum, my original misgivings, in response to Fodor, were right. *Computation does not give us an idea.* What it gives us is something else, entirely—something hugely valuable, I believe, even a sine qua non without which we will never come to understand the mind. But it is a thing of a completely different *kind*. What computing gives us is *experience*: insights and intuitions and practice and knowledge about the very same intentional phenomena that are being studied everywhere else. Phenomena, I might add, about which I do not think we yet have any very good theories of (but more on that later).

Now I have presented this picture a few times to computer scientists, and I have been stunned by their response. They are surpris-

[19]«References (e.g. to 'grey box abstraction,' IRL work, Kiczales et al., etc.)»
[20]With apologies to the Pennsylvania Dutch.

ingly agreeable! "It makes sense!" they say. And I agree: there is something quite deliciously relaxing about it. For one thing, it caters to their computer scientists' egos: it means that computation is not just taking over the world: it *is* the world (they like that). But there is more serious agreement. It makes sense, for one thing, of the daunting and seemingly limitless complexity of computational practice—and the fact that, as the field matures, more and more kinds of training, more and more kinds of practicioners, are being drawn into it—from theatre designers to anthropologists to novelists to quantum physicists.

And yet, needless to say, the reconception I am recommending is no small change. Just so that we know what is at stake, let me list just five of its most important consequences.

1. It renders vacuous all statements of the form "computers can (or cannot) do a" (for any a). Will computers be intelligent? Sure—as soon as we figure out what intelligence is, and how to construct it. Will computers be our friends? Yes; if we end up figuring out how to build friends. And not, if not.

   And so on.

   Sorry, Bert.

2. It evacuates the computational theory of mind of all intellectual substance. To say that the mind is computational is nothing more than to say that it is a materially-embodied intentional system. Which we have known for thousands of years.

   Sorry, Jerry.

3. More strongly, it removes the term 'computational' from all interesting theoretical discourse (except, perhaps, from historical studies of engineering).

At this point a personal computer was lifted up from underneath the lectern, placed onto a table, laid it on its side, and draped with a black cloth. At which point the lecture continued…

4. It implies that the mathematical theory known as the "theory of computation"—the theory of Turing machines, effective computability, complexity, etc.—must either be (i)

discarded, or (ii) recognised as in fact being a theory of something else.

Sorry, Dana.

5. It challenges the integrity of computer science departments.

   It is good I have tenure.

These are strong conclusions, but I think they are conclusion we must embrace.

5 Conclusion

There are many more things to say. Consider those construals of computing, for example—all those ideas about what characteristic property identified the computational subset, about what it was in virtue of which computation was *special*. I have claimed there is no such subset to be identified. So they failed in their original purpose. Does that mean we should throw them away?

No, they can be resurrected—this is a requiem, after all. But as befits the occasion, the they need to be transformed, in the process.

We do not have much time too look at them now. It turns out, if you do look at them (I cannot resist a few comments!) that each construal rests on a basic, seminal insight—an idea or intuition into the nature of (all or some) intentional systems. But in each case, the construal formulates its insight in particularly stringent terms. In particular, it formalizes or "absolutizes" its insight: forces it into black-and-white, all-or-nothing form. In each case, the absolutist formalisation turns out to be too strong. But if the black and white nature of the formalisation is relinquished—and the goal of identifying a computational subset of intentionality dropped—then the aboriginal insight can be recovered and used in the only remaining project worth doing: developing a general theory of intentionality.

The insight underlying formal symbol manipulation has to do with the tension between the semantic and the effective-in particular, with how systems have to use what is local and effective in order to behave appropriately with respect to distal situations that they are not causally engaged with. In that form, this is an unbelievably general and important insight, that we should never lose sight of. In formal guise, though, it ends up claiming that the semantic and the effective are independent—which as we have seen, is far too strong.

Even more interesting is the insight underwriting the "theory of computation" so favored in theoretical computer science. What it turns out to be, on this reconstruction, is a theory of pure mechanism. It is not a theory of computing—not just because there is no such thing as computing for it to be a theory of, but also because it does not deal with the "meaning" half of computing's fundamental dialectic. Rather, it is a theory of the *flow of effect*—of how states and state changes, arranged in architectures and processes, can be affected by and themselves produce behaviour. What we call a theory of computation or computability, that is, is neither more nor less than a general mathematical theory of causality. It does not look like a theory of causality, because the quest for formality has led it to be formulated in a way that is totally abstract. But once we let go of that conceit, we can see that what it is really doing is dealing with is the architecture of cause and effect, at a slightly more abstract level than in terms of the physics of concrete devices.

But as I say, these stories—and the work they involve—must be left for another time. For make no mistake: adjusting our theories to accommodate these changes will occupy us for at least another twenty-five years. Students should rest assured. As far as I can tell, most of the intellectual work remains to be done.

What I want to close with is a theme that has been with us since the beginning: there is more in the blood and bones of working programmers than has yet been formulated in language that other cognitive sciences can understand. In a way, I think of twentieth century computing as semiotic alchemy: a rag-tag bunch of practices, thick with inarticulate pre-theoretic knowledge, rich and disorganised-practices that, in spite of their distraction of turning web pages into gold, nevertheless contains within them the seeds for revolutionary theory-practices that, like fourteenth and fifteenth century alchemy, will, once those theories are developed, be largely forgotten, perhaps even shunned, until sometime, around the twenty-third century, someone in science studies writes a doctoral dissertation arguing that in fact we knew more back here in the twentieth century than the twenty-first and twenty-second centuries thought.

Buried in these practices are powerful intuitions—intuitions about architecture and implementation, about ontology and abstraction, about programs and processes and behaviour and state, about mechanism and effectiveness and how to exploit the tiniest otherwise irrelevant portions of atomic structure to set up long-distance correlations with distal states of affairs. Synthetic practices—not of studying or theorising, but of building. Practices sobered by the humility that comes not from merely thinking you understand something, but from actually trying to construct it-and thereby encountering the fact that you understand virtually nothing at all.

These practices, still under explosive development, lie in wait for philosophers and psychologists to mine. Just do not think they are a subject matter; just do not think we are doing something special. *Forget the original c-word!* For ironically, the idea that computing was really supplying an idea—that computation was a legitimate, autonomous, subject matter, that the term 'computational' denotes a scientifically interesting property—has been the major impediment blocking our appreciation of the vast intellectual importance of these synthetic developments.

Or maybe I can put it positively.

Only if we understand that there is no such thing as computing will we be in a position to appreciate computing's monumental impact on our intellectual life.

B • Metaphysics

— Were this page blank, that would have been unintentional —

How Can I Miss You
If You Won't Go Away?

Brian Cantwell Smith*
University of Toronto

*Coach House Institute, Faculty of Information, University of Toronto
90 Wellesley St W, Toronto, Ontario M5S 1C5 Canada

Last edited: November 14, 2009

Comments welcome
Draft only (version 0.01)
brian.cantwell.smith@utoronto.ca

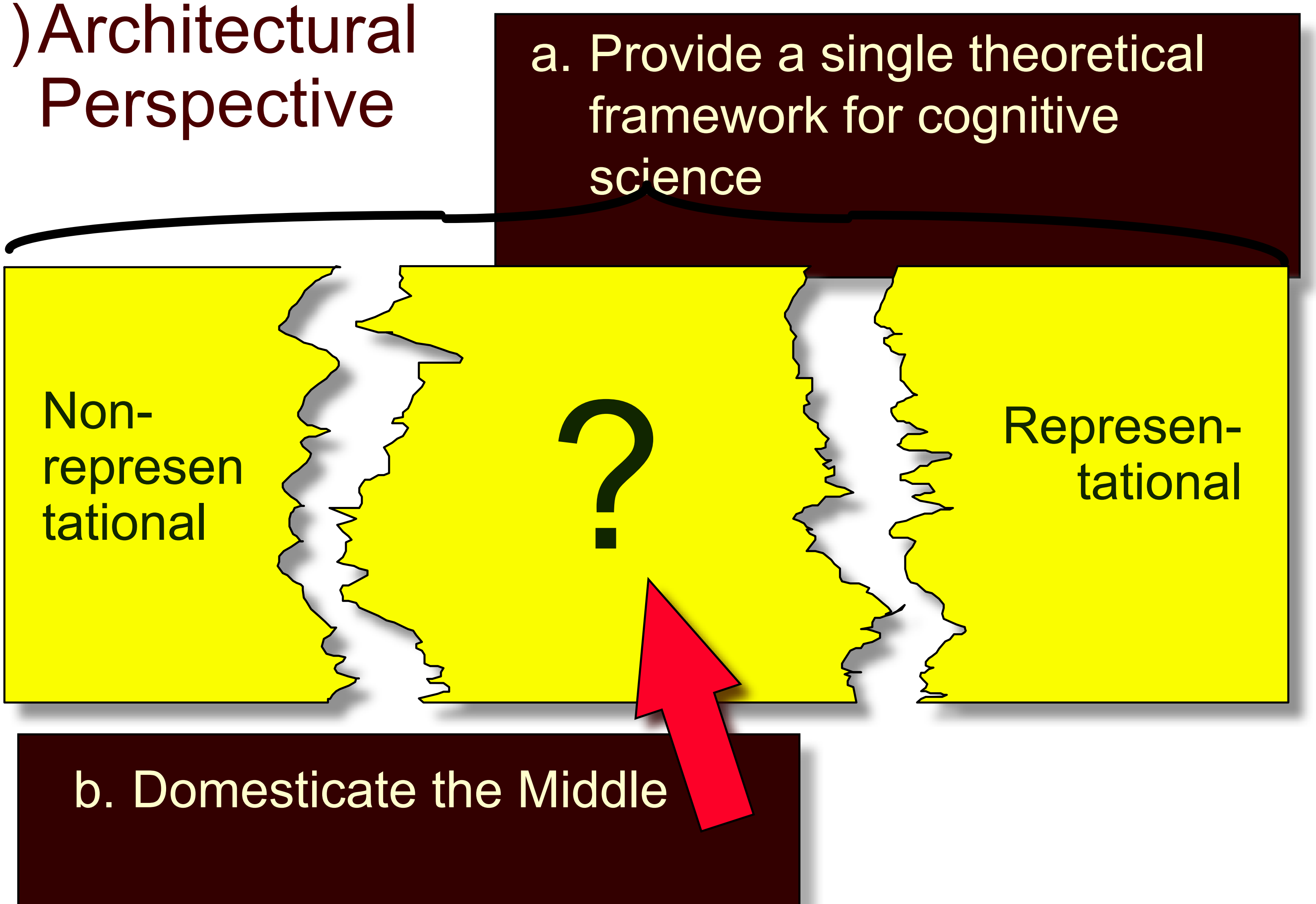
How Can I Miss You If You Won't Go Away?

Brian Cantwell Smith

Cognitive Science, Computer Science,
and Philosophy, Indiana University

bcsmith@phil.indiana.edu

1) Architectural Perspective



bcsmith@phil.indiana.edu

1) *Architectural Perspective*

Representational	Non-representional	
Slogan	"dynamic, embodied activity"	"formal symbol manipulation"
Paradigmatic systems	dynamics connectionism	logic, GOFAI ("computation")
effective		
1) local behavior	whatever happens	(syntactic) inference
2) character	continuous	discrete (?)
3) temporality non-effective	dynamic	static (?)
4) non-local coordination	———	semantics
5) structure	single-level	double-level
6) theory	differential equations	proof & model theory

bcsmith@phil.indiana.edu

Strategy

Non-representational	Representational
a) local/effective (behavior): whatever happens	a) local/effective (behavior): syntactic inference
b) continuous	b) discrete (?) …
c) temporal	c) static (?) …
d) ——	d) non-local & non-effective: semantic coordination
	e) two-level
e) single-level	g) proof / model theory
f) diff. eqs.	

bcsmith@phil.indiana.edu

2) Explanatory Perspective

1. "Extreme" naturalism
   a. Start with field-theoretic interpretation of (classical) physics
   b. *No presumed ontology!*
   c. Ask how "content" can rest on (develop out of) such a base
2. I.e., equivalent to a major metaphysical question:

   How can a bunch of physical "stuff" develop an (objective?) conception of (grasp on) the world around it?

3. Thesis:

   Need long-distance coordination (from the representational side)

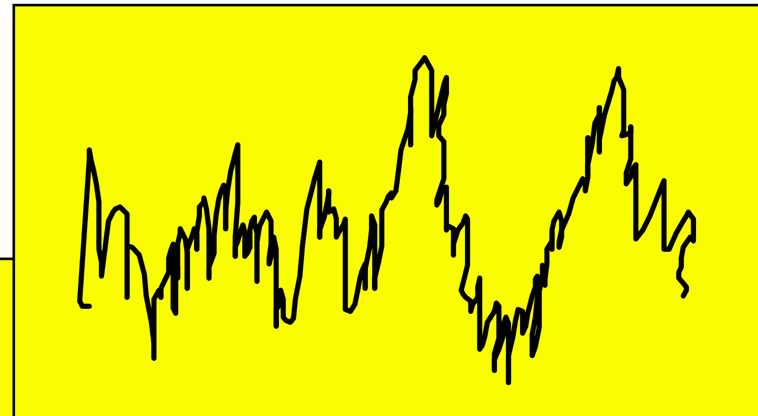
bcsmith@phil.indiana.edu

Part I — Tracking

bcsmith@phil.indiana.edu

A. Background

1. Continuous sea of fields
   — cf. Newton/Maxwell
2. May be turbulent, as in dynamics
3. Modelled by differential equations
4. Note: *incredible amount of fine detail*



bcsmith@phil.indiana.edu

B. Flex & Slop

1. The world is characterised by an underlying “flex” or “slop”

   — unlike gear world

2. Claim: this flex and slop …

   a. Establishes the problem that representation and intentionality solve

   b. Provides the wherewithal for its solution

bcsmith@phil.indiana.edu

C. Tracking — Connected (effective)

1. In simple cases of tracking (and other interactions), behavior can be *causally driven by the environment* (via the impinging stimulus)
   a. Sunflower
   b. Watching birds
   c. Thermometer
2. Paradigmatic case of being “coupled to the environment”
3. Perfect for (standard) dynamical analysis

bcsmith@phil.indiana.edu

D. Tracking — Disconnected (non-effective)

1. Problem: can't be affected by things *beyond effective reach*
2. Examples
   a. Rabbit: watching coyote behind a rock
   b. Super-sunflower
   c. Error correction codes
3. Strategy

   Use a *local mechanism*
   in order to *maintain coordination*
   across a *gap in causal coupling*

4. I.e.,
   a. *Exploit* local freedom (small, local motion is cheap), to
   b. *Compensate* for lack of coupling with what is remote

bcsmith@phil.indiana.edu

E. Intentionality & Representation

A way of …

1) Exploiting local freedom or slop in order to
2) “Orient towards” (establish and maintain coordination with)
3) What is beyond effective reach

→ ***All representation is*** **disconnected**

bcsmith@phil.indiana.edu

F. Disconnected Tracking (cont'd)

What happens when the coyote (sun, file) disappears?

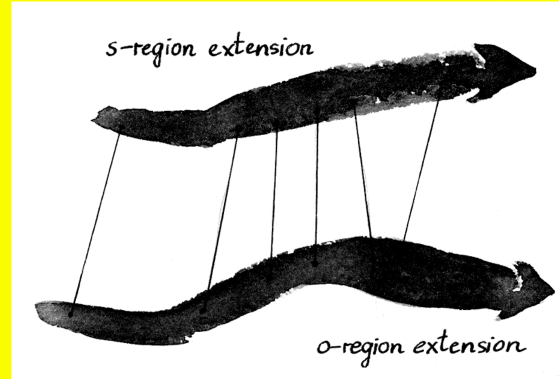
1. Need to *find a mechanism*
    a. To maintain coordination
    b. "Take over responsibility" from environment / disconnected object
    c. Compensate for what environment no longer (effectively) provides
2. Opens up (asymmetrical) split between "subject" & "object"
    a. Makes room for a difference between:
        i. What the subject is doing (tracking coyote) &
        ii. How it is doing it (with a neural mechanism)
    b. Possibility for "error"
    c. Origin of syntax/semantics
3. Needn't do it alone — cf. tools, equipment, civilisation
4. Important examples: memory & prediction

Part II — Stabilization

bcsmith@phil.indiana.edu

A. Stabilisation

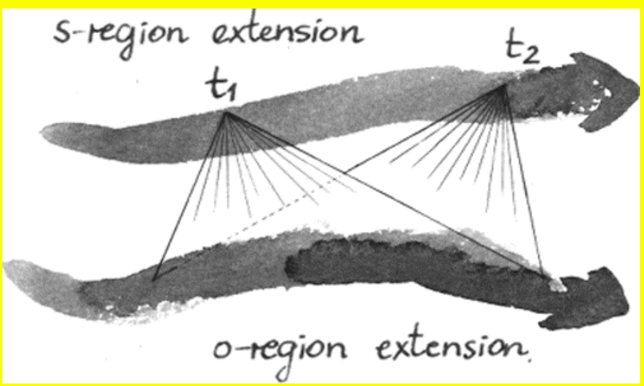
1. So far
   - Compensated for separation; not yet exploited it
   - Done nothing to stabilise the object as a unity



2. Problem
   - So far: established relational stability with “object”
   - Now: need to push stability outwards onto that object
   - I.e., subject must “extricate itself” from the object
3. Solution: abstraction, requires detachment
   a. “Let go” of variation in object (avoid buffeting)
      – Cf. jack-hammer, insecure sycophant
   b. Can’t be too far away, either
   c. Like vision:
      hold in the

middle distance

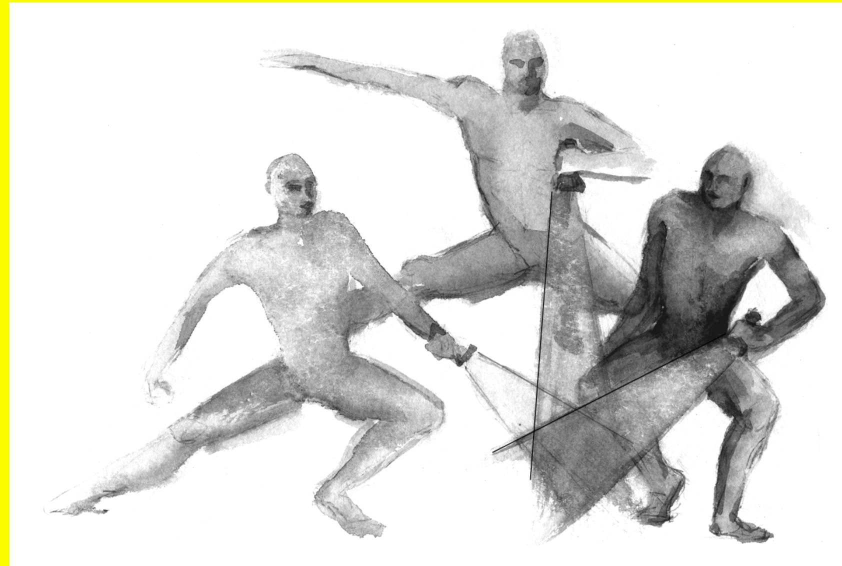
4. Establish cross-c
   ✘ Point-to-point correlation (like physics)
   ✔ Each-to-all



B. Intentional Acrobat

1. How are patterns of cross-cutting extension achieved?
2. Subject must

Compensate for its own contribution to the (deictic) relation in which it stands to the object



a. I.e., it must deconvolve the deixis

b. In order to stabilise the object as an object

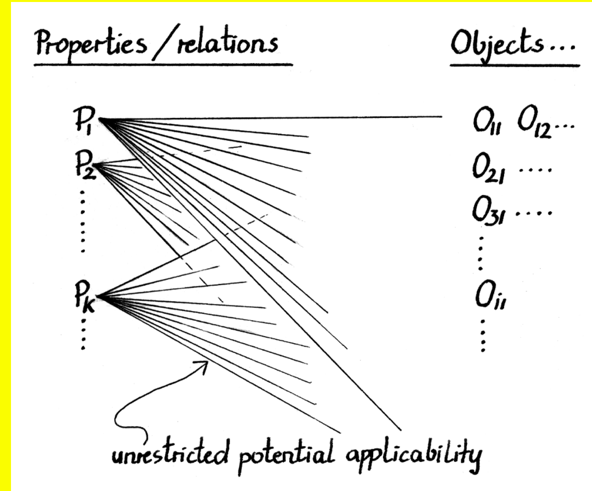
3. Cf. intentional acrobat

bcsmith@phil.indiana.edu

C. Preservation of Reference

1. Adjust indexical representation
   - "Behind me" ⇒ "In front of me" (turning around)
   - "Tomorrow" ⇒ Today" ⇒ "Yesterday" (upon passage of time)
2. Allocentric ⇔ egocentric coordinates
   - "Bar at the corner of Park & 3rd" ⇒ "Half a mile ahead on the right"
   - "Meteor at 37°16'n × 122°11' w at 2/24/97 19:51:04 cst" ⇒ "Meteor *right here, in 5 minutes!"*
3. Implicit ⇔ explicit
   - a) 5-3788 ⇒ 855-3788 ⇒ 812-855-3788 ⇒ +1- 812-855-3788
   - b) "Turn right!" ⇒ "Turn to *your* right"
4. Change in facts
   - "The tallest person in the class" ⇒ "The second tallest … "
5. Change of interest or belief
   - "Best paper I've ever read!" ⇒ "One of Jones' earlier tries"

D. Other parts of the story



1. Intentional dynamics
   a. General framework for talking about (temporal) coordination of
      i. local, causal, connected behavior
      ii. distal, non-effective, disconnected
2. Subjectivity / indexicality
   a. Claim: “deixis” fundamental to physical laws
   b. ⇒ *All representation initially indexical* (1st person)
   c. Challenge is to show how 3rd-person reference arises!
3. Variety: range of representational contents, from fields & features to concepts
4. Ontology
   a. “Stabilization” of objects ⇒ *registration* of objects
   b. Reference-preservation” strategies are general *stabilization* practices
   c. (Metaphysical) picture: objects (registration) wrested from the flux

bcsmith@phil.indiana.edu

E. Philosophy of Presence

Q: Why (re)presentation?

A: To deal with what's not around!

Q: What's the final word on representation?

A: Forget the 're'!

What is it to be present in the world, and to have the world present as world ?

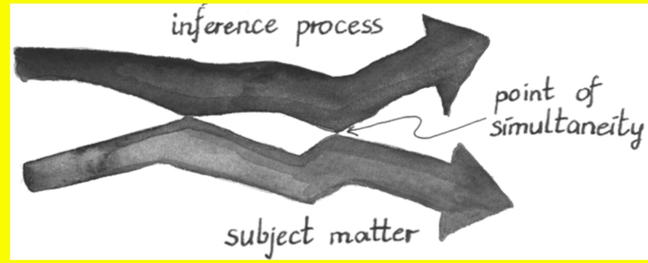
An Homage to Separation and Loss

bcsmith@phil.indiana.edu

Extra Slides

bcsmith@phil.indiana.edu

B. Intentional Dynamics



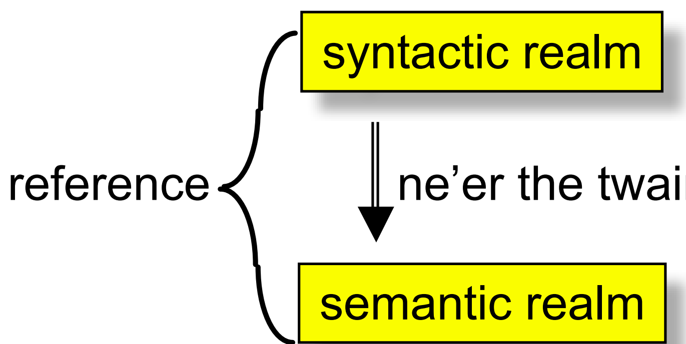
1. Generalised inference?
2. Would fit, under Etchemendy's characterisation
   a. Soundness: wanting what you get
   b. Completeness: getting what you want
3. Extended from normal inference in many ways:
   a. Effective (causal) connection with subject matter, not just intra-symbolic manipulations
   b. Apply to terms and predicates and features as well as to truth-bearers (sentences)
   c. Deal explicitly with deixis and generalised context-dependence
   d. Handle on-going temporal (& other forms) of temporal overlap & real-time coordination
      – e.g., clocks

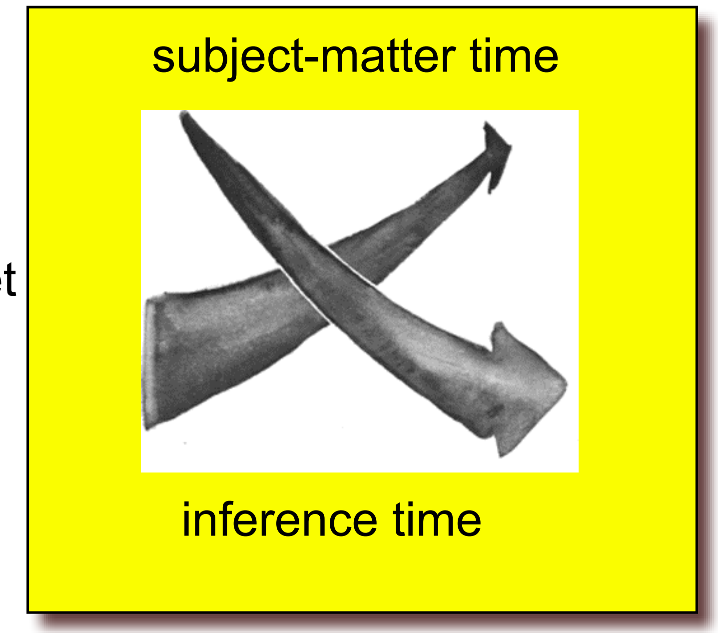
Call the result

Intentional Dynamics

bcsmith@phil.indiana.edu

C.d) Are logic etc. static? No!
Rather: Two realms are *independent & disconnected!*

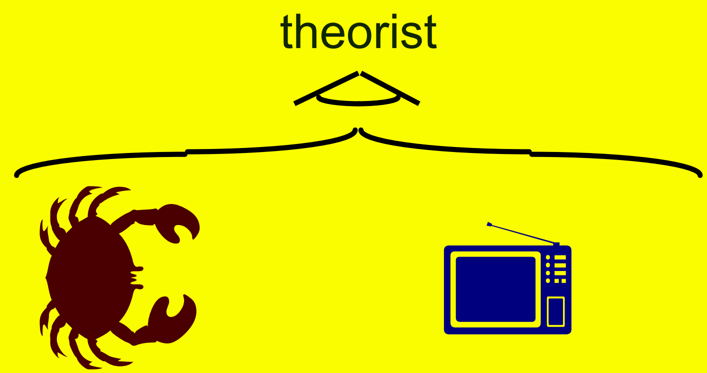




C.f) Formal Ontology

Distinguish

1. World for *theorist*
2. World for *critter*



Traditionally, ontology assumed to be

1. Given — presumed in advance
2. Shared — by critter and theorist
3. Conceptual — objects × properties, sets, types …
4. Uniform — everything captured in the type
5. Unproblematic — no strife, maintenance, negotiation

} Higher-order digitality

bcsmith@phil.indiana.edu

— Were this page blank, that would have been unintentional —

4 — Subjectivity & Objectivity

1 Introduction

Two properties are required, for a creature to be intelligent—for us to extend to it a sense of "we":

1. It must be conscious, "awake"—there must be "someone home" behind its…cameras;
2. There must be a "world out there," for it: a world it knows, moves around in—even: is curious about, respects.

Subjectivity and *objectivity*, in others words—as long as we are generous about the meaning of the words.

Subjectively (as Nagel says), there must be "something it is like," to be it. It must have an "inner" life; a point of view; a perspective on the world that originates from a knowing, feeling, conscious[1] self—a self that it, and it alone, owns, occupies, and knows the world from. Objectively, it must make claims about—have confidence in, rely on—a world beyond its senses, distinguishable from rank hypothesis and flights of fantasy. A world about which to be right…and to be wrong.

This is one way to characterise my project: to understand, enough for us to build it, what is involved in having authentic subjectivity, and authentic objectivity.

[1]Prereflexively original

1 Acrobatics

An example, before delving into analysis—an analogy, on which the rest of the discussion will depend.

Imagine a dimly lit stage, in a darkened theatre, blanketed in a fine mist. On stage is a single character: an acrobat, wielding a flashlight (figure 1). Silently, the acrobat leaps and dashes about, twists and turns, flings herself through the air in crazy contortion. All the while, as she weaves about, she adjusts the orientation of the flashlight, so that its beam—a dagger in the mist—passes unerringly through a constant point in space, three feet off the floor, front and center stage. As the beam angles and glances, that single point of illumination fairly glows.

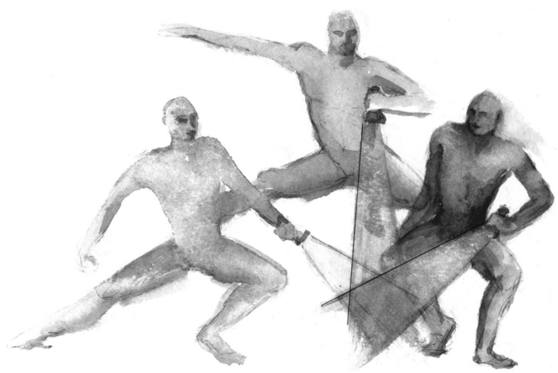


Figure 1 — The Intentional Acrobat

Six observations:

1. Performing this feat would take immense skill, but, as regards *orienting the flashlight*, very little energy.
2. Two motions are relevant: (a) of body (legs and torso); and (b) of wrist. In a sense, wrist motion "compensates" for body motion, with respect to stabilising that single point of illumination. Wrist motion is not opposite of body motion; nor is the wrist held constant (as if epoxied to space). Body and wrist are both lithe—they dance. The only thing epoxied to space is the target glowing fireball.
3. All motion—everything that "happens"—happens in the vicinity of the acrobat. What holds still is the point of illumination: 10, 20, … 100 feet away. From the acrobat's perspective: what is local, is not stable; what is stable, is not local.
4. Though what is stabilised is distal (to acrobat), it is the fluid, dynamic acrobat, not the stable illuminated point, that determines the boundaries of the point of illumination. (This will matter, for constructivism.)

5. Both (a) an explanatory theory of the acrobat, from the outside, and (b) the norms governing the acrobat's performance, from the inside (who was chosen to dance; whether the performance rates a 9.8) would refer to the stability of the distal, illuminated point.
6. If the acrobat stopped, the point would disappear. Nor could a photograph—a single time-slice—capture the phenomenon. Sans activity, nothing determines what position, along the beam, is stilled. Only through activity is the point triangulated upon, and thereby given (durable) identity.

Second example. Turn up the lights; get rid of the mist. This time, instead of a flashlight, the acrobat wields a pointer—a baton. Body movements are just as wild; twists of wrist, as ingenious. There is no illuminated spot, but the audience's attention—*your* attention—just as uncannily, would be riveted to the spot at which the acrobat irrevocably points.

Just two comments, this time:

1. In this version, *really* nothing happens—nothing measurable, at any rate—at the designated point. No energy is transferred, no causal connection exists, between acrobat and point (locus of activity and locus of stability). If the acrobat danced at the back of the stage, we could enclose both acrobat and pointer—the entire behavioural dynamics—in a "black box", across the boundaries of which no energy flowed, which excluded the pointed-at target. That is, we could contain the acrobat, movement, and pointing inside a "closed system." But even if we did this, (a) "what the acrobat is (intuitively) doing; (b) how the acrobat's dance would be judged, and (c) the "invariant" to which an explanatory theory of the acrobat's activity would necessarily advert, would remain external to the energetically bounded region.
2. No amount of local investigation at the point to which the acrobat is pointing—no microscopes, X-ray crystallography, exquisite sensors—could determine, up here at the front of the stage, what was being pointed at. In this (ad-

> mittedly limiting) case, there is a sense in which there is "nothing there," here: no independently-delineated entity or spatial region to serve as the "object" of the pointing. Yet that fact in no way undermines the existence or reality of this spot's being the spot *towards which the acrobat is pointing*.

A third and final example—or rather, suite of them. This time, replace flashlights and pointers with words. Yesterday, I thought to myself: "I have a talk to give tomorrow." Today, I refer to that day as 'today.' Tomorrow, I will think of it as 'yesterday.' Each night, there is a motion of my body—of my self—through time. Each night, compensating for this temporal shift, I adjust my brain: my mentalese. Neither my body, as referenced to the world, nor my mental or brain state, as referenced to my body, remain constant. *Constancy is death!* What is stabilised is once again distal: the day on which I gave the talk, the referent of my thoughts.

Similarly: I say "to my right"; you think "to his left." Last week I referred to "the tallest person in my class." This week, a basketball player signs up. Now, to refer to the original student, I have to say: "the second tallest person in the class." Or suppose I ask: "Would you hand me my glasses, there, on the table behind you?" You turn around. Now where are my glasses? *In front of you*. You unconsciously adjust your egocentric thoughts, to compensate for the movement of your body, in order to stabilise the in-the-world place where my glasses rest. (A good thing, too; without updating, you would go round and round forever.)

2 Logic

These examples illustrate what I call reference-tracking, or the preservation of reference.

Who studies such things? No one!

Actually, that is not quite true.

There is a discipline that studies something similar. Namely: logic—one of the intellectual success stories of the 20th century, and forebear of computer science, cognitive science, and artificial intelligence. Logic does not study reference-tracking per se, for reasons we will get to in a moment; what logic does study is *pres-*

ervation of truth, through streams of inference. And while (pace Frege) I do not consider truth a form of reference, truth and reference have similar properties. Whether a claim is or is not true is not a local, energetic property of it, either. Microscopic investigation of the font will not tell you whether a contentious sentence is right. Whether dinosaurs were warm-blooded is settled—"located," we might even say—a long way away from here. If you submit a term paper containing an especially dubious sentence—unlikely to be true, or perhaps spectacularly shallow—my suspicions will not be relieved by X-raying the paper, to see how deep the claim's ink penetrated the surface of the paper. Truth is not a local property; it (almost always) involves *distal relations*.

Inference, in contrast, and the sentences or mental structures on which it operates, are, in a crucial sense, *local*. They are also *concrete*: physical, subject to causal forces. In order to construct an inference engine, theorem prover (or computer; more on that in a moment)—in order to build a device that can move, dynamically, from one statement in a proof or argument to the next—you must fabricate a local, dynamic mechanism, using small but nevertheless real bits of energy, not unlike the slight rotations of the acrobat's wrist. Crucially, however, (i) the *norms* or *specifications* that govern the transitions, in terms of which we evaluate the theorem prover, and (ii) the scientific theory that explains what the inference system is *doing*, both advert to the non-local, non-causal semantic *interpretations* of the sentences.

Mechanically, that is, because of powerful constraints arising from the physical world, inference is a local, concrete, dynamic activity. Constitutively, in contrast, in terms of meaning, semantics, and *interpretation*, logic is non-local, because the applicable norms—truth, falsity, soundness, completeness, validity, inference to the best explanation, etc. (all technical terms defined in logic)—make essential reference to the (stable) distal interpretation. Put it all together, and we can see what logic is: a *fundamental theory about the dialectical interplay of meaning and mechanism*.

A moment ago, I claimed that the development of logic was one of the great intellectual achievements of the twentieth century. This characterisation shows why: understanding the interplay of meaning and mechanism is one of the most challenging intellectual problems of all time. The problem, however—and it is very grave, especially as regards AI, cognitive science, and the present interest in subjectivity and objectivity—is that logic dealt with this interplay of meaning and mechanism in an *extremely narrow way*. In particular, it studied the interplay solely within the confines of metamathematics. As a result, it ends up with esoteric and narrowly restricted results, formulated in abstruse technical theorems.

Underneath these theorems, however, and buried in the tacit conceptual frameworks in terms of which they are framed, lies a treasure trove of insights that apply, much more generally, to arbitrary meaningful or (as philosophers say) "**intentional**" phenomena: things like thought, language, representation, etc. Moreover—this is why I have been going on about these things—these buried results are profoundly relevant to our topic of subjectivity and objectivity.

Let me name just three of these buried insights:

1. Meaningful or intentional processes (including thinking) operate under *extraordinarily severe physical limitations*. It is virtually miraculous that evolution, first, and now science, in its footsteps, have figured out ways to exploit an almost vanishingly modest set of resources, so as to enable powerful intellection, so as to give people an astonishingly powerful abstract grasp of the world in which they are embedded.

2. The dialectical interplay of meaning and mechanism is a "relationally emergent" level of coherence or regularity in the world. It involves a fundamental collaboration between the two aspects, and cannot be reduced either to pure meaning, or to pure mechanism. This, too, will have consequences for the prospects of developing scientific accounts of subjectivity and consciousness. The non-reducibility of the intentional is apparent even in logic. Gödel's incompleteness results, Turing's non-decidability theses, theories of computational complexity, and a number of other scientific results (including results from chaos theory and non-linear dynamics) demonstrate that intentional processes (such as thinking!), can never, except in

the most trivial situations,[2] be reduced to proximal physical or mechanical dynamics.

God read Browning. Meaning's reach always exceeds its grasp.

3. The "subject matter" or "content" of a representational structure—what a symbol or thought is about—is not, by and large, the proximal structure itself, nor any of the local activities in which it plays a causal role, but rather the distal situation towards which it (or the agent) is directed. Not *flashlight* or *pointer*, but *what flashlight is pointing at*.

Think about…oh, Sarajevo, ice cream, a long-lost friend. What occupies your mind—what your thoughts are focused on, what you are emotionally directed towards—is a *city*, *food*, or *person*—not a pattern of neural behaviour, not even an image. If you do not believe it, try coming at it from the other direction. Your boss emerges from a meeting with the dean about downsizing the department. You ask whether they discussed you. You would hardly be satisfied by the reply: "Not at all! I only talked about patterns of activity in my neo-cortex."

My brief, here, is not to applaud logic. In fact my very first philosophy paper[3] denounced its limitations, rehearsing any number of ways it is inapplicable to human thought. My only point, here, is to highlight some of its more tacit, but ultimately most powerful, accomplishments—what I will call its core insights. Because what happens next, in intellectual history, is very interesting. Those fundamental insights get lost.

3 AI & Cognitive Science

Logic, as I have said, has two preeminent offspring: computer science, and AI/cognitive science—both fields in which I work. I will describe them separately, because they abandoned logic's insight in separate ways.

In computer science,[4] logic's intentional vocabulary has been redefined to refer to something else: namely, *purely causal rela-*

[22]«Quote Barwise on why completeness proofs are a sign of failure, not success.»

[3]«Ref 1978»

[4]As we explored in colloquium last week /«ref».

tions—between programs and the processes they engender; and between and among states of machines. There is a long story to be told about how this came about, but the bottom line is that computer science has *projected all intentional vocabulary back onto pure mechanism*. By doing so, computer science can honour science's general methodological predilection for causal accounts—but at a severe cost: genuine (non-causal) intentional directedness is thereby "disappeared." And as we have seen, to do that is to "disappear" the subject matter, since no mechanical reduction can do it justice.

The situation in AI & cognitive science is completely different. Rather than re-appropriating logic's vocabulary, AI and cognitive science have largely overthrown the logicist paradigm on which they were founded—for reasons we have already seen: its being massively too restricted and brittle to account for the sheer ingenuity of on-the-fly human behaviour.[5] What is at stake is often displayed as a list of opposites. Logic is seen as committed to a conception of cognition as: *individual*, *rational*, *abstract*, *disengaged* (from the world), *explicit*, *static*, *discrete*, *generic*, and *context-independent*. The alternative—often called a "situated" approach—rejects all of these assumptions, in favour of a claim that cognition (all human activity) is: *social*, *embodied*, *concrete*, *located*, *engaged*, *dynamic*, *continuous*, *particular*, and *context-dependent*. Something like *improvisational navigation*, rather than *rational intellection*, is taken as paradigmatic human competence. The new view also waxes hugely enthusiastic about the idea that human behaviour is "emergent": not the sharpest tool in the shed, in my view, but a term generally used to index the intricate, seemingly non-reducible patterns of (sometimes self-) organisation, as exemplified in modern non-linear dynamics and complexity theory.

I will label this wholesale (infamous) transformation in cognitive science the **situated sea-change**. I should say that I am completely in favour of it—it is something I have argued for for more than twenty years. If anything, I want something even more radical. For another assumption of logic, less often included in

[5]Not that it is logic's fault; that field, after all, developed as a theoretical effort to put the foundations of mathematics on rigorous intellectual footing, not to explain how, in finite time, you can work your way across a crowded Tokyo subway station.

the laundry list, is its commitment to what I will call "*formal ontology*"—an assumption that task domains consist of neatly individuated, discrete, unambiguous objects, properties, and relations. As will become evident in a moment, formal ontology is not my favourite, either. I want to press for a more constructivist alternative.

In sum: I want to see the situated movement, and raise it one.

Problem is, in discarding logic, cognitive science not only rejected its untenably narrow restrictions, but threw out its core insights, as well—in fact, discarded its entire project: of understanding mind as instantiating a fundamental dialectic of meaning and mechanism. In particular, it threw out what is so distinctive about the acrobat: the centrality of constitutive, non-causal directedness towards a distal object. That is: it threw out what is most important about meaning, semantics, interpretation.

Was this rejection necessary? No. There is nothing about intentional directedness in general, or even about representation in particular, that runs counter to the situated sea-change. Representation—or perhaps intentional directedness, something like representation but far more general—not only can, but must be "rehabilitated" from within a situated perspective. Not only that, such rehabilitation is a necessary precursor to understanding either subjectivity or objectivity.

4 Rehabilitating Representation

Here—very quickly—is how it would go.

Physical regularities—causes and effects—are, as I have said, *local* in essentially all relevant respects, both spatial and temporal. That poses a problem for cognitive creatures. All you get, if you are physically embodied, in terms of *resources*, are two things: (i) the effective arrangements within you, plus (ii) what is pressing in on you, right now, at the surface. You live in a laminar cocoon, with physical coupling limited to the immediate here and now. Moreover, the world is *sloppy* (only weakly correlated), so you cannot necessarily tell, from what is happening right near you, what is going on elsewhere—behind that rock, or back at home, let alone what went on yesterday, or will go on tomorrow.

Fortunately, however, that same slop—those local degrees of freedom—mean that you can rearrange your internal states with remarkable facility (if you are clever), without expending much energy. So what you do—what agents do, *what it is to think*—is to represent the world out there, beyond the periphery, by rearranging your internal configuration, and adopting appropriate habits and practices, so as to behave appropriately with respect to—so as to develop hypotheses concerning, so as to *stand in appropriate normative relation to*—that to which you are not, at the moment, physically coupled.

That is—to reduce it to one sentence—intentional systems:

1. Exploit what is *proximal* and *effective*, so as
2. To be intentionally oriented towards—i.e., to behave appropriately with respect to, to satisfy governing norms regarding—what is *distal* and *non-effective*.

I call this the **intentional mandate**. It is this mandate for which the acrobat is meant to stand as a metaphor.

As I say, the mandate contravenes none of situated cognition's tenets. Nor does it militate against even radically constructivist metaphysics. Remember the acrobat: what was delineated was established by the acrobat, not by any pre-existing boundaries or identity in the pointed-at spatial region.

5 Subjectivity

So that—glossing a thousand details—is the project: to (i) accept all the revisions of the situated sea change, while at the same time (ii) retaining the intentional mandate.

In what time remains I want to sketch the beginnings of a positive view, by seeing what it takes for subjectivity and objectivity to arise.

The route in is through consciousness. Consciousness has defied understanding for thousands of years because it seems so fundamentally different from things in the material world: sticks and stones, houses, galaxies and quarks. Consciousness has seemed especially inaccessible to scientific explanation, because of an apparently a vast, perhaps even unbridgeable, "explanatory gap"[6] between the two topics we are talking about tonight:

[6] «Reference Levine»

1. The private, **subjective**, inexorably first-person qualitative or phenomenological character of conscious experience; and
2. The public, **objective**, detached, third-person character of empirical science (from physics and neuroscience to cognitive science to scientific psychology).

It is the first-person character of subjectivity that most people think is the root of the problem. Scientific accounts of third-person perspectives are expected to be (relatively) unproblematic. But when they consider the first-person case, people are driven to say such things as "We haven't the remotest idea of how consciousness could arise, physically."

I believe this intuition is almost exactly backwards. In my view, contrary to received wisdom, for all physically-embodied creatures, it is first-person and second-person perspectives (or at least inchoate versions of them) that are easy. What is hard—what takes skill—is developing a third-person, objective conception. That is what requires acrobatics; that is what leads to the extraordinary intentional dance, to which I alluded at the beginning.

To see why, think back to the example of turning around to retrieve something behind you. The terms used—'in front of', 'behind'—indicate positions from an egocentric or oriented point of view. That is, their reference is defined in terms of a coordinate system established by the concrete circumstances of their utterance. Many words have this property: *here/there*, *you/me*, *now/then*, *come/go*, *bring/take*, … etc. Tense, too, is similarly egocentric. Linguists call such constructions *indexical* or *deictic*, because their interpretation depends on their use.[7],[8]

What matters, for us, is a special character of this use-dependence of indexical terms. In particular, what objects such words refer to are not only identified with respect to the use act

[7]Explain meaning/interpretation Δ.

[8]It is *active* use, not *static inscription*, that matters; Perry has an example of two deaf mutes, so poor they share a single card saying "I'm a deaf mute; can you spare any change?", which they alternately hand to passers-by. What 'I' refers to depends, dynamically, on *who hands out the card*.

itself; they are identified differentially, in terms of a "change" or "deviation" from the location and orientation of the utterance. Thus "bring" means to transport something *towards the location of the speaker*; 'you' refers to *the person to whom the speech act is directed*, etc.

The differentially-defined character of indexical terms is important to our story for three reasons:

1. Reaching upwards, it underwrites subjectivity;
2. Reaching downwards, it stems directly from physics.
3. It is what makes achieving objectivity hard.

Indexicality and subjectivity thus serve as something of "bridge" phenomena, connecting higher-level objective understanding with underlying physicality.

5.a Physical deixis

Look downwards, first, towards physics, and consider an example. What does a magnet say to an iron filing? "*You—come here, now! You—come here, now!* ..."—on and on and on, forever. Every single linguistic element ingredient in this baldly anthropomorphic projection (all four words, the present tense, the imperative voice, etc.) is indexical. In general: if we give voice to what happens in physics, not from the outside, as in standard physical theories, but—like ventriloquists—*to the physical entities themselves*, so that the content of their communication mimics the operations of the governing physical laws, we are *forced to put indexicals into their mouths*.

This, I argue, is because *physics itself is indexical*. Not physical theory—physics-qua-epistemic-inquiry—but the actual concrete physical world: the force fields, the flux, the underlying plenum. Admittedly, no one talks this way (except me—yet!). But the indexical, or, as I prefer to call it, deictic structure of the physical world is tacitly recognized in our theoretical frameworks. Magnetic and gravitational attraction, bumping and shoving, flashlight beams, etc.—all physical regularities—are: *local* (in space-time), *incremental*, and *differential* (they govern how things change, not how things are).

This deictic, differential structure of the plenum matters *im-*

mensely. It means that as physical signaling pathways well up from, or shade off into, direct physical engagement with the world, they must take egocentric form. Your stomach sends a signal to your brain saying: "hungry!"—meaning something like "I am hungry, now!"; not "Brian is hungry on Monday night." As you leave the room, your cortex will instruct your motor routines: "turn right, now!", not "turn north at 8:56 p.m.!" Our entire physical existence, in the end, is grounded in such egocentric, deictic signals—symbols trembling on the verge of mere mechanism.

This deictic structure of physical coupling is not subjectivity—yet. But it will underlie it. For consider, taking this deictic physical coupling as the base case, what it is to start on the long and difficult road of beginning, not to stabilise one's relationship to the immediate environment, but to stabilise the things related to—the world out there, beyond the incoming signals, beyond the press of that 1/r2 enclosing causal envelope.

The issue has to do with the relational nature of the egocentric, deictic physical connections. There is no problem if you are being causally driven by something in the nearby environment: then, if your circumstances change, your state will change, too, correspondingly—i.e., you will be "updated by the world." Sunflowers, for example, can *track* the sun by being *driven* by the sun—using a simple servo mechanism. The difficulties arise not when you are engaged with what you care about, but when you are *disconnected*—when you leave. Then, when your circumstances change (position and orientation, say), your internal state, if you allow it to remain constant, will relate you to something new. A displaced magnet will attract new filings; a rigidly held flashlight (rigid with respect to the body of the acrobat) will wildly light up new spots. In order to get a fix on the world, what you want to stabilise is not the *relationship* you bear to it, but *what you are related to*. Consider a fixed entity in the environment, say, like home—if you want to hold that stable, through changes in your circumstances, then you must adjust your egocentric relation to it, to *compensate for what has happened to you*. If you rigidly held onto the idea that it is "four blocks left and nine blocks down," you will find yourself forever referring to a new spot—not what you intend. Think about the acrobat: as she moved, she had to rotate her wrist, so that the light beam could point in a new direction, and thereby arrive at the same point in space—instead of maintaining the "same" egocentric direction, which would have led it to point to a new point in space.

A less contrived example. Consider: the vestibular-ocular reflex—the fact, wired deep in your brain, that if you rotate your head, with your eyes open, you will invariably find yourself rotating your eyes, in their sockets, by an "equivalent" amount, in the opposite direction, so as to maintain a stable point of visual focus—on a distal chair, tree, whatever. With effort, it is possible to override this, and rotate your head *without* doing the compensating eyeball rotation (i.e., holding your eyeballs fixed in their sockets); in which case—*note this!*—you no longer see the world; everything turns into a blur. But if you allow your eyes to move, just the right amount, the blur suddenly vanishes, and—this is incredibly important—*the world snaps into focus* (not the incident optic array, but the distal world itself).

Objectivity is going to be something like that: *having* (allowing) *the world to snap into focus*.

In some ways, the vestibular ocular reflex is misleading, as an example of intentional acrobatics, because vision can be causally-driven, at least in part, by what you are looking at (like the sunflower, vision can "servo in" on a scene). What is distinctive about long-distance intentional or semantic directedness, in contrast, of the sort objectivity requires, is that it *cannot be "driven" by what it is directed towards*. In order to maintain objectivity—in order for their to be a world, for them—agents (and communities) have to shoulder responsibility for keeping themselves appropriately "pointed."

In spite of these differences, however, two facts about the vestibular ocular reflex are revealing:

1. Like the acrobat, it involves a process that I call **deconvolving the deixis**—compensating for the contribution that changes in your circumstance make to the deictic physical relations in which you stand to the environment. You *wash out your own contribution*—and thereby let the world be stilled.

2. When you "lock onto" a distal object, what is stabilised (via these processes of deictic compensation) is seen or looked out onto *from a point of view*. The very conditions that allow the world to snap into focus entail, as a consequence, that the world is "looked out onto," by someone, from some place. Not only that; this "someone" is not anonymous. The "looker" is the very same "you" whose contribution you just had to erase, had to wash away, in order to allow the scene to emerge—like a developing Polaroid picture—before your very eyes.

It is a consequence of physics, in sum, that objectivity is only subjectively achievable.

6 Discussion

Needless to say, there is an extraordinarily long story to tell—a science's worth—about how these deconvolution processes, starting from deictic physical engagement, but gradually "letting go" of what is immediately connected, in order to stabilize what is progressively far away, can ultimately lead all the way to objectivity. In what time remains I will enumerate three final points, to convey a flavour of how that story goes.

6a First, second, & third person

The first set of comments has to do with relations between first, second, and third person perspectives.

You cannot update your *entire* physical state, every instant—it would take too much work. So not all representations can be deictic and/or egocentric. On the other hand, you always have to do some updating, because you need egocentric representations in order to mesh with your basic physical (mechanical) capacities. So in general—if you are going to develop a whole conception of a world—you need an efficient, balanced set of representational strategies, that trades off between local detail, necessary for action, and long-distance stability, necessary in order to "still" the world.

This is where full-scale intentional and representational practices comes in. All sorts of strategies are employed: relatively less perspectival or indexical signs, established landmarks, relying on others, the whole framework of "representational scaffolding" that Clark and others talk about in cognitive science.[9] Think about how you get to a national park. Relatively non-perspectival maps get you to the right region; from there, much more indexical signs— "straight ahead," "two miles (from here)," right or left—take you from there to the gate; from there, direct (Gibsonian) engagement with the physical environment guides you the last few feet. Even the road itself—including the markers at the edge of the lane—can be viewed as a final, deictic sign that allow local, connected guidance.

Third-person perspectives do not emerge as so much "from nowhere," on this picture, as "from anywhere"—at least "from anywhere within a broad range" (nothing is entirely deixis-free). Moreover, they are invariably grounded, connected to the world, by being seamlessly connected to progressively more indexical or first-person signs, shading imperceptibly into direct (deictic) physical engagement.

In sum, long-distance conception requires third-person stances, in order to manage the complexity (through abstraction and a degree of perspective independence). Being part of community requires second-person representation, so as to be able to communicate with other deictically-embedded agents. Local navigation requires first-person, deictic directedness, for coupling to the physical plenum.

Two crucial points, about this range of perspectives:

1. Developing an *objective* conception of the world—one that includes the knowing subject—requires not just one of these perspectives, but the full integrated set: skills to move back and forth seamlessly, between and among them (as you did, when turning around to pick up the glasses).

2. All intentionality—directedness—must be anchored to the world via grounded, first-person, subjective skills—skills that ultimately mesh, without a trace, into direct physicality.

[9]Regularities that hold over a full *range* of experience, do not need to be "updated" if one moves around within the range. You need not think about time zone if you stay on the East Coast; or that "to the right of" is a three-place relation, not two, unless you spend a lot of time standing on your head.

6b Ontology

The second set of comments has to do with ontology—with *world-making*.

Physical fields are stupefyingly complex: a maelstrom of superimposing and crashing waves and vortices and turbulence—a little like falling overboard at sea, and finding yourself drowning in a buffeting array of turbulence and spray—except without the "you".

The world is so overwhelming in detail, in fact, at the physical level, that if we tried to react to it as genuinely physical (i.e., as field-theoretic), we would be completely swamped. No finite creature could *begin* to deal with all this detail; to say nothing of the fact that huge amounts of it—almost all—is of no especial interest, as regards ascertaining what lies beyond.

So what does the creature do? It *abstracts*—or, as I like to say, "**registers**"—the world, simultaneously (i) failing to attend to a lot of what is there, by ignoring the vast majority of the detail or "information" with which it is presented, and (ii) attending to what is *not* really there, by imposing conceptual structure on what remains, so as to render it modestly intelligible.

A particularly important form of abstraction is what we might call "conceptual":[10] the staggeringly reductive simplification of the world into the familiar cast of characters—objects, properties, and relations. This is the most "third-person," disconnected, non-engaged form of registration. In keeping with its being necessarily disengaged, it is highly non-detailed. It is useful for long-distance coordination. As one moves in to physically engage with the world, the need for—and adequacy of—conceptual abstraction (i.e., the need to parse the world into discrete, reidentifiable individuals) falls away, to be replaced by fine-sensory and motor coupling with inexpressible detail.

Objects, properties, and relations, in other words—products of conceptual registration—are, as I once put it: "the long-distance trucks and interstate highways of normative, intentional life. They are undeniably essential to the overall integration of life's practices—critical, given finite resources, for us to integrate the vast and open-ended terrain of experience into a single, cohe-

[10] Maybe talk about Evans/McDowell Δ.

sive, objective world. But the cost of packaging up objects for portability and long-distance travel is that they are thereby insulated from the extraordinarily fine-grained richness of particular, *indigenous* life—insulated from the ineffable richness of the very lives they sustain."

There is no reason to suppose—indeed, every reason not to suppose—that, in these conceptualist projects that parse the world into discrete objects, that how the world is registered is "independent of the subject." That does not make the view entirely *relativistic*—i.e., does not imply that ontology (the stabilised world) is wholly dependent on the subject (or the subject's community). Neither limit case is tenable—one hundred percent world, independent of subject; or one hundred percent subject, independent of world. Both are ideological manifestations of exactly the sort of absolutism for which we banished logic. Like gardening, real-world ontology is a *collaboration*—between subject, environment, community, and thereby-stabilised world.

Four notes:

1. This constructivist sentiment is not an extra-theoretical metaphysical assumption; it is demanded by taking field-theory seriously, as an account of the physical basis of existence, and computational complexity arguments seriously, about what finite creatures can do with limited resource. That is: an intermediate level of social constructivism is *naturalistically forced*.

2. This forced view—to which a serious commitment to physics inevitably leads—is one of *ontological pluralism* on top of *metaphysical monism*. I am prepared to argue that the resulting picture does simultaneous justice to what matters, to realists, about realism—and also to what matters, to constructivists, about constructivism.

3. I have said that ontology is not independent of subjects, because it is based on abstraction. The metaphysical world is not independent of subjects, either, but for a different reason: subjects are part of the world—and parts are not independent of the wholes they partially constitute.

4. A note for philosophers: part of what is being said, here is, that ontology needs naturalisation as much as semantics—since (I claim) there are no reidentifiable objects in science. And not just ontology: but abstraction, as well—its epistemic warrant.

6c Normativity

Third and finally, a remark on normativity.

I have not said much—have not said anything, really—about what establishes the non-effective links that relate subjects (i.e., us) to the entities towards which we are intentionally directed. This directedness, as I have suggested all along, include "truth" and "reference", but that is not all; intentional directedness also involves caring, respect, love and hate, curiosity—and awe.

The bottom line is that the links are normative—as long as we are generous about the meaning of that word. If we had time, I would draw distinctions between (what I call) "statical" norms—norms that govern states, in the way that truth and reference and the values of science and the like have traditionally been conceived—and "dynamical" norms: norms that govern processes. And I would go on to say that dynamical norms, including full ethics, how to live, living in the truth, etc., must be in the driver's seat.

More particularly—to draw out the most important moral—it is the dynamical norms governing people's ongoing projects that warrant the abstractions in terms of which they ontologise the world. Whether there is a single mountain over there, or three mountains—which of the claims "There is one mountain" or "There are three mountains" is *true—depends on what you are doing*, depends on your normative stance towards the normative status of the object of your claim. More generally, there are no norm-free empirical claims. On this the feminists, among myriad others, are right.

Moreover—there is something wonderfully ironic about this—this normative dependence of ontology on life practices is *especially true of conceptual claims*, involving traditional ontology. Direct physical engagement—the local, microdetailed physical engagement with the world, being coupled to the immediate surround, is not nearly so subject-relative. Rather, it is third-person, framed in terms of high theoretical abstractions—including the claims of science—that are the most norm dependent.

The point can be framed terminologically. At least since Descartes, the word 'matter' has been split in two: between a noun, meaning physical stuff, and a verb, meaning something like "is important." On the view I am defending, the two sides rejoin. A "material object," I claim, is *a chunk of reality that matters*. ("To whom?" you ask. That is right: that is the right question.)

Put it this way: my aim is to heal the temporary 300-year schism between matter and mattering.

7 Summary

So that is the picture, in a few thousand words. Radically embodied agents who, in virtue of (i) their singular and collective particularity, (ii) the unutterable contingencies of their existence, and (iii) the enabling constraints that derive from being embodied in physical and social fields, are able to *parlay their material freedom into a commitment to the world*—a world that contains them but transcends their grasp; a commitment that allows them to take it as world, populated with what exists.

Because that, in the end, is what objectivity is: a commitment, on the part of subjects, to take the world to be world, host of objects (and everything else); and objects, to be entities hosted by the world.

Subjectivity and objectivity, in other words, far from being independent or at odds, are inextricably intertwined, for at least three reasons.

1. Objectivity requires subjectivity, because of those underlying physical constraints; there is no other way for reference to be authentic—no other way for a theory to refer to anything at all—except when it is grounded in an agent's subjective experience, and deictic physical engagement.

2. Truly objective knowledge requires recognising one's own subjectivity, in order to take the world to be the world. You cannot be objective, that is, unless you recognise your embodied participation in, responsibility for, and effect upon, and subjective contingency of the conceptual scheme in terms of which you understand, the world about which

you care, to which you extend commitment.[11]

3. Finally, being conscious, being awake, being aware, in its ultimate sense, in turns requires objectivity. In this the meditative traditions are perhaps right. True consciousness requires commitment to, but openness towards, the self-transcending world.

8 Epilogue

Two projects, I take it, motivate a university:

1. Understanding who we are—our character, our practices, what we are made of, what tests our mettle; and

2. Understanding, without limit, the world that hosts us.

Subjectivity and objectivity, in others words—as long as we are generous about the meanings of the words.

Except now we have expanded beyond the confines of ai, cognitive science, and philosophy, to incorporate history, literature, the humanities and social sciences—including social studies of scientific inquiry as a particular (often privileged) form of human practice. This time, that is, we are concerned, not about what it is to build a creature, what must be true in order for us to extend a sense of "we," but about the human condition, and the world we find ourselves in—how to understand the "we" *that we already are*. This time it is not a case of constructing scientific accounts, from the outside, of subjectivity and objectivity, but of developing stories, from the inside, of what our subjectivity and objectivity are like.

Except that to divide things up this way is untenable, for two reasons—*both of which we reached as results, above*.

1. First-person ("inside") and third-person ("outside") understandings are not separable like that. No third-person account, for starters, *has ever referred to anything at all*, except

[11]Fragment: It also requires admitting that one's attempts to describe or conceptualise it invariably make reference to one's inescapably embedding normative projects—from which it follows that there is no single story, no master narrative. *That* one is an embodied subject, however, and therefore that there is no single story, no master narrative, neither voids talk of reference, nor commitment of object.

as grounded in concrete, particular first-person subjectivity. On the classical image, scientific understanding is claimed to be third-person; but we have just argued that such a view must be wrong. It is incompatible with physics. Like all understanding, scientific understanding must rest on an integrated suite of first, second, and third-person skills.

Similarly on the other side. No subjective social science or humanities understanding of our place in the world can have the requisite detachment and dispassion to warrant the university's imprimatur, unless it is also objective—unless it assesses our situation fairly, from the outside. Committed detachment—passionate dispassion—require subjective-cum-objective integration, not fracture.

So dividing the university by who studies from the inside, and who studies from the outside, is a mistake.

2. Second, because abstraction is normatively grounded, empirical studies cannot be separated from normative studies. The "takes" on the world that science uses are grounded in the normative projects of scientific inquiry (and perhaps the societies that do it).

All of which has consequences for how we should proceed.

Look at it this way. Science, in the form of logic, ai, and cognitive science—to say nothing, increasingly, of neuroscience and biology—is *theorizing theorizing*—constructing knowledge of what knowledge is, developing accounts of how accounts are developed. Gradually but inexorably, science is growing reflexive—bringing the "doing" of science within science. Though still inchoate, this process of self-incorporation is bound to accelerate. Scientific accounts of consciousness and first-person subjectivity, just now starting to appear, will soon be *de rigueur*.

Increasingly, that is, science will domesticate questions that have classically been viewed as extra-scientific: not just subjectivity and consciousness, which we are already seeing, but also: the relation between models and reality; the character of scientific evidence; the role of the observer; which of formalism, Platonism, relativism, constructivism, or any of a host of other realist and irrealist metaphysical positions is right; the normative status of world-making; what it is to care. Questions like this, which used to be relegated to bars and late night conversations, to conviction and personal faith, are moving into the heart of this reconstituted science. (They are even starting to be talked about in the middle of the day!)

Problem is, though I find this all intellectually thrilling, I am not politically sanguine about these developments. I do not think we can ask science, present-day science anyway, to shoulder this whole burden. I am still troubled by that erasure of logic's central insight, in computer and cognitive science—by the weight of the methodological pressure to reduce everything to mechanism. Because as we said at the outset, to take on these ultimate questions is to address issues of meaning as well as mechanism—the original dialectic. Symbols, interpretation, sociality, normativity—to say nothing of even deeper issues, such as wonder, creativity, and awe. To do this right, we need collaboration from disciplines that study those things: not just the arts, as well as the sciences, but politics, law, religion. The full gamut. For it is with their insights, about contested stories, pluralism, etc.—as I hope even today's bit of a sketch suggests—that these considerations from science (even: robotics) most naturally mesh.

So maybe this reflexively reconfigured inquiry should not be called science, but science^prime^. Or perhaps (since we are studying meaning): *reënchantment*. Maybe even: colleges and universities themselves. "*Eruditio* et *religio*", one might even say—if that phrase were not already taken.[12]

How this all works out, we will have to figure out together. All I know for sure is that the intellectual consequences of (i) the rapidly converging trio of computing, biotechnology, and something we have not much mentioned, nanotechnology, in conjunction with (ii) the theoretical disciplines that undergird, nourish, and are in turn buffeted by them…the intellectual consequences of those two things, once we realise they have our being in their sights, are immeasurable. This is why I said, when interviewing for this job, that what matters about STS—about the social impact of science and technology—is not just the spread of the tech-

[12]'*Eruditio et religio*'—learning and religion—is Duke University's motto.

technology itself, its impact on our practices and routines. What matters even more are the ideas on which these technologies are founded: ideas go straight to the heart of our conception of what it is to be us.

5 — The Nonconceptual World[†]

Over the past twenty years, the notion of non-conceptual content has played a prominent role in philosophical discussions of the relation between thought and language, on the one hand, and perception and action, on the other. In spite of its importance, however, the nature of non-conceptual content remains remarkably obscure. The fact that it is negatively defined is not helpful. Just about the only thing on which both proponents and detractors of nonconceptual content agree (beyond the fact that non-conceptual content is not conceptual content—though as we will see it is very likely that they disagree on what conceptual content is) is that the predicate 'non-conceptual' should be understood *epistemically*, as a predicate on content-bearing mental states.

What distinguishes nonconceptual content, it is thought, is the structure of the belief, or the attitude of the believer—not the world thereby believed in. I will argue that this purely epistemological focus is mistaken, in the following sense: that the phenomena that have driven at least many advocates of a notion of non-conceptual content to embrace the notion are not, in the end, best understood from an epistemological point of view. Instead, I claim, the character of and ultimate warrant for nonconceptual content is ontological.

There are two parts to the claim, to put it most baldly. First, what is explanatorily fundamental about nonconceptual content, I will argue, is neither how it is used, nor the epistemic role in plays in the agent, but *how it takes the world to be*. And second, the world as truth-maker for such nonconceptual content bearing

[†]…

mental states is …

Any distinctive epistemological characteristics of nonconceptual content is a consequence of that ontological commitment.

I will also argue that nonconceptual content is, in a specific sense, more accurate—more faithful to the world in detail—than conceptual content. This raises problems for the analysis of conceptual content, including for issues of realism and truth. Conceptual content, I argue, involves a form of abstraction, which in turn ties both semantic issues of truth and reference and ontological issues of objects and properties to dynamical human norms. At the same time, the story remains fundamentally realist, illuminating the metaphysical ground underlying the intimate relation among perception, thought, and action.

The overarching theme of the investigation is that epistemic issues of experience, representation, and thought, and semantic issues of truth, reference, and content cannot be solved without tackling fundamental metaphysical questions about the nature of objects, properties, relations, and the founding world. Only by understanding the mind against this metaphysical ground, and investigating both conceptual and nonconceptual representation's ontological commitments can we understand what these various kinds of content are like, what these various kinds of content are for.

1 The Nonconceptual Content Debate

The notion of non-conceptual content has played a major role in recent discussions about the relation between thought and language, on the one hand, and perception and action, on the other. Yet in spite of its importance, the underlying nature of nonconceptual content has resisted trenchant analysis. The fact that it is negatively defined does not help. It has also been recruited in diverse ways: some writers focus on nonconceptual content in judgment or thought, often perceptually based; others, on forms of intentional content that play a role in action. Some have asked whether nonconceptual content can be exhibited in a creature without conceptual skills at all, or whether instantiation of mental states with nonconceptual content requires a prior or concomitant mastery of concepts. Across this range of issues there is no doubt that the literature contains a number of intriguing insights and provocative proposals. Yet it is probably fair to say that nothing approaching a comprehensive theory of nonconceptual content has yet been presented.

The aim of this paper is to redress that situation. I will start by considering the case of non-conceptual content in judgment (including perceptual judgment): judging that the world is a certain way-but not a way that finds ready or even possible expression in conceptual form. Later I will extend the analysis to action, but even in the case of judgment substantial theoretical issues arise, in need of disentangling.

I will start by considering two major lines of argument: Gareth Evan's classic defense of nonconceptual content (Varieties of Reference [VOR]), and John McDowell's now equally classic rebuttal (Mind and World [M&W]). Evans and McDowell are talking past each other, I will argue—and therefore (to put it a bit anachronistically) are missing each other's point. Their views are held together only by an implicit assumption—an assumption it will be helpful to put on centre stage, in order to subject to challenge.

1a Evans vs. McDowell

Though Evans' avowed concern was with reference, truth, and objectivity, in fact he was almost equally concerned with the epistemic role that concepts and conceptions played in the mind of agents. As is especially evident in parts II and III of VOR,[1] for Evans a semantic account of a concept involved, among other things, *explicating the role that the concept played in the mental life of a person who possessed*. His discussion of indexical concepts, for example, and his recruitment of notions of information delivered by the senses, are as much an analysis of their role in cognitive activity as they are of traditionally semantic issues such as truth. Indeed, one of the signal contributions of VOR is Evan's attempt to show how reference is achieved, not just what reference is.

Evan's epistemic or even cognitive bent is well kept in mind in understanding his analysis of conceptual content. For him, conceptuality has first and foremost to do with the semantic or propositional form of a judgment. As codified in his **Generality Con-**

[1]Unless otherwise indicated, all Evansian references are to VOR.

dition, Evans takes a judgment to be conceptual just in case it is of the form *a* is *F*, and the agent not only thinks that *a* is *F*, but is also capable of entertaining the thought that *b* and *c* are *F*, for any *b* and *c* of which it has a conception, and that *a* is *G* and *H*, for any *G* and *H* of which it has a conception (modulo various appropriateness conditions). Thus conceptuality, according to Evans, consists of content with something of an algebraic or compositional constituent structure.

It is content of this (at least potentially) recombinant form, according to Evans, that figures in linguistically articulate judgments and propositional attitudes—i.e., that is conveyed by such embedded English sentences as that there is a war in Serbia, that the sun is rising, that three people are standing at the door.

For McDowell, in contrast, conceptuality has first and foremost to do, not with the internal structure or form of a judgment, but with the fit of the judgment into a overall conception of the world: not a complete conception of the world (which, if even meaningful, would be impossible for a finite agent to achieve), but, as it were, a conception of a complete world—a cognitive grasp on the world's being whole—exhaustive, entire, complete in all details. What leaves McDowell unmoved by cases of perceptual judgment that lead others to embrace a notion of nonconceptual content is his (correct, in my view) recognition that we understand even spectacularly diverse arrays of colour as part of the world of our experience-as located, objective parts of the one comprehensive reality.

These at least superficially different concerns—between the internal form of a judgment for Evans, and its role in undergirding our grasp of reality in toto, for McDowell—are betrayed in a number of passages, in M&W, that to an Evansian, might otherwise seem perplexing: his unproblematic embrace of indexical concepts and indexical judgment, such as "that red". This is conceptual, à la McDowell, because, in any situation in which it is uttered, it plays an unproblematic role in our overall cognitive grasp of the world. For Evans, indexical judgments are not so obviously conceptual, even if they have the structural form *a* is *F*, because two different thinkings or utterances of 'that red' could have such different contents. If, like Evans, one is concerned to explicate the role that thoughts play in a rational agent's cognitive economy, saying that a thinking of 'that red' refers to a particulate shade (I will get to the issue of individuating shades presently) doesn't do half of the work that is required. The problem is that one such thinking may connote blood, occult ceremonies, and Burgundy wine; another, lipstick, Orlon sweaters, and Muffy at a prep school dance. For an Evansian, they are different judgments—and the differences matter, as regards conceptuality, because how they differ is not obviously conceptually explicable. For McDowell, of course, they are different judgments too—but with respect to the defense of conceptuality, the differences don't matter—because both locate their referents in the world, in a way that can be comprehended objectively, and because of that fact, are, in McDowell's sense, conceptual.

There are others cracks suggesting that Evans and McDowell may be focusing on different aspects of judgment, if not on outright different phenomena. One is McDowell's comment, in passing, about the problematic nature of just-noticeable-differences (JNB) in colour perception.[2]

The only conceivable argument that McDowell's criterion entails Evans' criterion—i.e., that a grasp of the world must consist in judgments all of which have what Evans would call conceptual form—rests in part on what is ultimately an ontological assumption—an extremely common one, yes, but not something the student of nonconceptual content should blithely assume:

A · The world is exhaustively constituted of objects exemplifying properties, standing in relations, configured in situations or states of affairs, and gatherable together in sets, and so forth, with at least some of those objects (the concrete ones) spatio-temporally located or related.

As is evident from the Generality Condition, it is just this sort of

[2]The phenomenon of a just-noticeable difference, or JNB, well-studied in psychology, arises in cases where, for example, of three shades *x*, *y*, and *z*, human subjects, even in excellent viewing circumstances, cannot judge that *x* and *y* are different, nor that *y* and *z* are different, but *are* confident in being able to see a difference between *x* and *z*. In this situation (of having a non-differentiable neighbour in common) *x* and *z* are said to exemplify a "just noticeable difference."

world that Evansian conceptual content represents. Content meeting the Generality Constraint is not only itself articulated, that is, and potentially recombinant; it also attributes an articulated and potentially recombinant structure to the world it represents. For issues of modal realism aside, most would agree that the objects, properties, and relations thereby represented are metaphysically distinct from each other in part because they, too, could have been differently combined. If *a* is *F*, then God—or another world, or even this same world, at another time and place—could have made it the case that *b* is *F*, or that *a* is G, for appropriate *a*, *b*, *F*, and G.

The picture is thus relatively clear. If A is true, and the world thus consists of objects exemplifying properties and constituting states of affairs (etc.,), then it would be natural to conclude that McDowell's completeness requirement could be met by entertaining judgments that are conceptual in form, according to Evan's characterisation. This is not the realm of logical implication: nothing guarantees that even if A is true, all judgments of its being this way need satisfy Evan's condition; and conversely, nothing guarantees that meeting Evan's condition in and of itself need give the agent a comprehensive grasp of the world as whole. Rather, the point is that it is A that allows people to think that Evan's condition and McDowell's requirement characterise the same sort of conceptuality—and hence to conclude that McDowell and Evan disagree. For unless A is true, it is not clear that Evans and McDowell's construals of conceptuality are even compatible. If A is *not* true, in particular—as I am going to argue it is not—then an agent able to entertain only conceptual content in Evans' sense would not be able to meet McDowell's requirement: hence would not be able to have conceptual content in McDowell's sense at all.

My strategy, therefore, will be to take as conceptual any content that has this ontological character: that is, any content that represents the world in terms of objects, properties, and relations in the standard way. That is, I differ from Evans (and most other writers) in focusing: (i) neither on the articulate structure of the content itself—i.e., qua sense, proposition, meaning, or other "intensional" entity; (ii) nor on the structure of the expression or representational vehicle; (iii) nor on the conceptual capacities of the agent—in the sense that the agent could rearrange the content, or the mental state that bears it, by substituting other pieces of appropriate type; but (iv) *on the structure of the world thereby represented*. This ontological focus does not automatically imply that conceptual content so defined is without distinctive epistemological character. Rather, what I want to argue is that conceptual content so defined is strong enough to entail Evans' Generality Condition as a consequence.

For ease of discussion, I will extend standard usage and say that conceptual judgments take the world to be conceptual when they take it to consist of discrete, extended, concrete, reidentifiable objects, exemplifying properties, standing in relations, arranged in states of affairs, grouped together in sets, etc. That permits the following simple definition: conceptual content is content that takes the world to be conceptual.

I want to argue for this "ontologising" of the conceptual/nonconceptual distinction from considerations in cognitive science, my home field. I will have more to say about the specific character of cognitive science later; for now, it is enough to say that it is based on a broadly representational theory of mind, taking mental life to arise out of semantically-warranted causal transitions in material substrates. That is: the mind itself it taken to be physical instantiated (and hence, in a sense, to be a physical mechanism), but nevertheless to be distinctive, among physical mechanisms, in trafficking in representational, or semantically evaluable, states (and hence to be not "*merely*" a mechanism). What especially matters is that the transitions between and among these states are normatively governed. The most familiar—though, as we will see, not the only—norm is that the transitions should be semantically sensible (e.g., truth-preserving).

Within this context, the foregoing characterisation of conceptual content yields something like the following (familiar) image. An agent with a mind is taken to be a causally-realised, normatively-governed creature inhabiting and interacting with the world around it. Semantically-evaluable information derived from the world—that x is F, that y is G, etc.—is encoded in causally-efficacious representational vehicles that lead the creature to act in (normatively) appropriate ways towards that same world. In perception, for example, an agent's encountering a situation of *x*'s being *F* would lead it into a representational state the (conceptual) content of which would be that *x* is *F*. If the creature had inferential powers, and believed (for example) that all *F*s are *G*, it might then conclude that *x* is *G*. Or something like that.

As I say, it is a familiar picture. And if that were all there were to it, then perhaps all content would be conceptual—at least all content in experience or judgment.

But in spite of its familiarity, what thirty years of cognitive science have shown is that it doesn't work.

It doesn't work because the world doesn't do its part.

There aren't any objects out there.

2 A world without objects

Some of you may fear that, in denying that the world contains objects, I have taken leave of my senses. On that I ask you to withhold judgment until the paper is done. What I can assure you now, however, is that I have not abandoned realism. Perversely, in fact, it is exactly in order to preserve realism that the story I am telling must be told. It is not me, but the person who clings to objects, that is, as it were, "unrealistic" (though I admit that the nature of the real is going to come in for something of a beating, in order to see why that is true.)

Now in order to defend this strong a claim—not that there aren't any objects, which would be false, but that there aren't any autonomous objects independent of subjects—we need an industrial-strength theory of what the world is like. Where do we find such a thing? There are at least three candidates: (i) commonsense, and the deliveries of introspection; (ii) science; and (iii) our experience constructing contentful systems. I will consider each, in turn.

2a Commonsense

Start with introspection and commonsense. It certainly seems as if the world contains objects. *Just look! Lo: a table! a chair! a person sitting on a chair! a person sitting on a chair at a table!* Nothing, most people think, could be more fundamental to unreflective, lay experience.[3] Perhaps that is right. But to say that commonsense judgments take the world to consist of objects exemplifying properties and standing in relations is just to say that at least the sorts of commonsense judgment that philosophers bring forward have articulated conceptual content. We know that.

Problem is, the fact that the natural attitude takes the world to be conceptual begs the question: of whether the world is autonomously conceptual, independent of and prior to our so taking it. So we have not made any progress.[4]

2b Science

So look at our second source: science. How could science have possibly succeeded, to the extent that it has, unless the world really and truly consisted of objects exemplifying properties, independent of our so taking it?

Thing is, I am not convinced there are any objects in science—at least not objects of the right kind. At the very least science doesn't provide a theory of objects—a theory of concrete, extended, reidentifiable, particulars, in Strawson's sense. It doesn't provide a theory of identity or individuation, in particular—and an object is not an object without identity or individuation conditions.

What science does give us is theories of *properties* or *types*: such as trees. But ask whether that clump of redwoods is one tree, or seven, or thirteen—and you'll find that science is of no help. Or suppose an amoeba splits. Did the old amoeba die, and two new ones emerge? Or is the old amoeba still with us—just spatially distributed? Or is one of the new amoebas the same as the old one, with the other new one having just been born? It is not just that biology doesn't provide any answers; *biology doesn't care*. This is because no scientific regularities, I'll wager, hold in virtue of ob-

[3]Actually I am not so sure of this. Explain ...

[4]This is too simple. Many people—from poets to painters to phenomenologists—deny that it is intrinsic to the phenomenological character of experience that it "objectifies" the world, as one might put it. I strongly agree. But that only strengthens the conclusion we are aiming at: that it is not the world, on its own, that presents in terms of objects, properties, etc., but that that is an epistemic way of taking it, with very particular merits and demerits.

ject identity. Scientific regularities, as I've said, care only about properties.[5]

More seriously, consider physics. I don't know much about quantum mechanics or relativity, except to know that they are surpassingly strange. It seems wild to suppose that they might provide the theory of individuals that classical physics does not. And it is by no means evident that there are any individuals in classical physics. For think of the world that physics depicts: a four-dimensional manifold of continuous spatio-temporally extended density, charge, force, mass, energy, etc. The best way to understand the ontological commitments of classical physics is field-theoretically: a stupefyingly complex superimposition of interpenetrating waves, vortices and fields and quiescence and turbulence, vibrations from glacially slow to blazingly fast, forces continuously impinging, forces falling continuously away. Imagine falling overboard in a storm at sea, surrounded by nothing but crashing waves, stinging spray, and undulating currents, as far as the eye can see-and then subtract you. That is approximately what the world is like, according to physics—except a zillion times worse.

The investigative practice of physicists, of course, does deal in concrete, discrete objects. Consider a high school physics prob-

[5]There are other objections. (1) Fields are nothing but space-time points, someone might argue; and a space-time point is a paradigmatic object. But I am not sure it is right that space-time points are paradigmatic objects. Space-time points are exactly not what ordinary objects are: extended. In fact it is not fully clear to me that we can even genuinely understand-can genuinely conceive of-space time points. We can understand representations of space-time points: temporally-durable points in a representation, such as a graph, in which time in the subject matter is represented by a spatial, not temporal, dimension in the representation). And we can arguably imagine space-points persisting in time. But no matter; perhaps I am alone in having trouble conceiving of a space-time point directly. The more serious comment is that, in virtue of not being extended, space-time points are intrinsically not subject to reidentification. There is no such thing as "encountering them again." Nor do they have paradigmatic properties of physical objects, such as being the common-cause of multiple effects, or the common effect of multiple causes. (2) What about chemistry? Surely chemistry deals in objects, such as molecules, or atoms? Or biology: what is a cell, if not an object? But is that really so? What are the individuation conditions on cells>

lem: a mass of 3 kg traveling at 4 meters per second slides off a table 1 meter high; where does it land? There are at least two objects in this situation: a mass of 3 kg, and a table. Or think about calculating the gravitational attraction between two masses m_1 and m_2. In such a calculation, those masses are likely to be taken as discrete objects. But of course they are not objects to which physics is ontologically committed. As objects, they are *idealizations*, are *approximations*. The true physical nature of the situation involves only a continuum of point-to-point forces and fields—a continuous manifold of physical disturbance. Yes, *qua* physicists, we traffic in discrete objects—but only for epistemic reasons, to make our calculations simpler-or even to make them possible.[6] (Calculating the gravitational attraction between two extended objects, without this idealisation, would require solving a double triple integral.)

Some may agree that basic physics should be understood field-theoretically, but claim that ordinary individuals derive from (rise up on, emerge from, etc.) this field of physical forces by abstraction or idealization. That might be right. But—and in a way this is the point—'abstraction' and 'idealization' are terms from *epistemology*, not from physics. No natural science theorises abstraction as such (along with force, mass, and valence). No scientist writes "*abstraction*(x)" in their daily equations. On the contrary, like the discrete objects mentioned above, abstraction is part of the epistemic practice of scientists.

It follows that abstraction stands in need of explication. It especially needs explication if one is committed to anything like a naturalistic account of mind. In fact naturalising abstraction is one way to understand what we are doing here.

To make this clear, some terminology will help. By **physical** ontology I will refer to the world as theorised in physics; not the epistemically simplified version that permits calculation, but the ontological version to which the fundamental equations are committed. By **material** ontology, in contrast, I will refer to the

[6]We may also treat higher-order objects as individuals. That is: it may be that physics cannot itself be formulated without objects (though I don't know for sure). But that's okay; my point is only that there are no first-order, discrete, concrete objects.

furniture of everyday life: tables, chairs, dirty dishes, continents, détente. The claim I am exploring is that the objects of conceptual ontology are in the first instance *material*, not *physical*. That is: individuality, identity, being one as opposed to being none or being two, are not issues that any purely physical science can address.

2c Experience with constructing contentful systems

The third place to look for evidence as to the nature of the world is nonstandard, but I believe very significant: efforts in cognitive science and artificial intelligence to develop representational systems. That is: efforts to build semantic or representational systems that honour constitutive norms by representing the worlds they inhabit. This has proved to be an unbelievably sobering experience.

For several decades, Artificial Intelligence (AI) tried to build systems that took the worlds they inhabited to be (what I am calling) conceptual: to consist of well-behaved objects, properties, and relations, in the standard way. And AI failed. Virtually no one in the field any longer believes that the route to intelligence is via this kind of logical or conceptual representation.

Two developments are especially relevant. First, even in the early years, sobriety overcame students of perception. If you place a camera (or other sensor) on a robot, and show the signals to a first-time observer, they are stunned. The world does not "present" remotely as well-behaved as we imagine. Let me simply say, to any readers who have not worked with empirical data, that these results are unambiguous, repeatable—and unbelievably humbling. It is a bit tricky, of course: what you must not do is to display the incoming data in another medium that recruits the same sensor or perceptual apparatus that we would normally employ to deal with the world in that modality. For example, it does not work to display the signal coming in from a camera on, say, a TV monitor—because then our faculties of visual perceptual simply "parse" the image on the monitor in the same way that they would have parsed the scene that the TV camera is recording—giving us a false illusion that the world is well-behaved. But as long as you look at them in a different modality, or—better—construct algorithms to deal with them according to you think are the patterns holding them together, you are forced to conclude that *the messiness and partiality of the world outstrip anything untutored intuition would ever have imagined*. And this is not simply a result of poor instruments—noise in the signal, problems of transmission. The neatness is simply not there, in the world, for the instruments or signal to record. (Note that the problem is not just that the world recorded by instruments does not come precategorised. It is that the objects to be categorised cannot be *distinguished*, cannot be *segmented*.) The idealisation—and it is an idealisation—that the world is made of well-behaved, discrete objects has to be imposed on an unruly underlying reality.

The difficulties are not limited to perception—which leads to the second development bearing on these issues: the collapse of traditional or "symbolic" AI and cognitive science, quite apart from issues of perception. For its first few decades, the cognitive sciences operated on a model of mind in which the task of perception was to recognise essentially conceptual arrangements in the world, and to encode the results in explicit representations; the task of "mind" was to reason and solve problems with respect to these representations, and the task of "action" was to take representations of desired states and bring them about. That is the project that failed, and notoriously so. The reason normally proffered for this failure—a reason you still hear in the hallways of cognitive science centers—is that these systems failed because they were *representational*. At Indiana, for example, where I teach (you may not believe this, but I swear it is true) to claim that people represent the world is thought by many colleagues to be a recidivist, backwards view (about as popular as admitting that you listen to Mantovani and the 101 Strings). What is happening is that the traditional models are inexorably being replaced by a variety of non-traditional alternatives: connectionist systems, embedded agents (à la Brooks), systems that search high-dimensional state-spaces, dynamic systems, etc.

For several years, advocates of these new systems claimed that they were better because they did not represent (Rod Brooks, a champion of the new view, wrote a paper famously entitled "Intelligence without representation"). Indeed, antirepresentationalist tracts still regularly appear. On reflection, however, it is becoming clear that this characterisation—that the old school systems used representations, and the new school systems did not—is an inadequate way to characterise the sea-change. In spite of the press, that is, the transformation has not been a shift from representational to non-representational systems. Rather, the situation is better described using the terminology laid out above. What was characteristic of the old school was that it used *representations with conceptual content*—that is, representations that represented the world of the agent to the agent in terms of what we are calling conceptual ontology: discrete well-behaved objects, properties, relations, states of affairs, and the like. What is characteristic of the new systems is not that they completely eschew representation. On the contrary: on the very general representational model adumbrated above, of normatively-governed, causal, contentful behaviour, the new systems are still fully representational. But these new representations represent the world in other-which is to say, in nonconceptual-terms.

2d Summary

In sum, neither science itself, nor our synthetic experience constructing systems to represent the world, supports the idea that the world au fond, consists of material objects. These two facts, in conjunction with the spectacular failures of early AI and cognitive science, suggest that we do well to be cautious in extrapolating from the content of our lay conceptual judgments to any view about the autonomous structure of the world.

3 Feature fields

But if not objects, then what?

My aim in this paper is to take a first, provisional step towards answering it. Not a step that gets us all the way. But a step that opens up a host of potent questions that any stronger answer will have to address. (Note: I am being conservative. I do not think that steps are discrete—and that we should take a lot of steps down this...plank. But bear with me; even this small step will have resounding consequences.)

In particular, with an eye on the fact that science was interested in types but not concrete individuals, I will follow Strawson, Cussins, and others, and characterise it in terms of what we might call a **feature-space**.

By a feature, imagine something like a property, but logically simpler, in that it does not require an object for its exemplification. That is: features are logically simpler—more basic, less committing—than more complex objects and properties. As a paradigmatic case, consider how we describe the weather. "It's raining," we say; or "It's foggy." But as everyone knows, the 'it' in such sentences does not refer; there is no thing, *x*, such that *x* is raining, or that *x* is foggy. Rather, as various writers have put it, "It's raining" means something like: "Rain, here, now!" or simply "Raineth!" Features do not characterise particulars; they do not serve as sortals. "Feature-placing," as Strawson called it—this is the crucial point—does not commit one to any discrete, concrete, extended thing that can be reidentified.[7]

So that is the image I want to examine: an extraordinarily rich, four-dimensional world (of time and space), that instantiates a bewildering array of features—colours, smells, textures, fogginess, whatever—without any concomitant commitment to individuals or particular identity.[8] And remember that this is the structure of the world I am talking about; not the structure of the incident sensory array, the press of local causes at the agent's periphery-not a manifold of sensation, received at the sensory organs of the creature, or a manifold of energy, pressing in on the creature's skin. Rather, it is an arrangement, laid-out in space-time, of differentiation-not yet "grouped" into the rather large-scale, coarse, "synthesized" or "abstracted" individual objects of conception.

In particular, the suggestion is that the world presents to embodied intentional creatures as a (literally) unutterably-rich spatially and temporally continuous array of spatio-temporally instantiated features, sans identity. The question is how those creatures manage it-how, given finite physical resources, they orient towards it, get around in it, in ways that satisfy the governing semantical norms.

Part of the answer (as is clear from modern neuroscience and biology and cognitive science and AI) is that they have different strategies, for different purposes—even if in our own case we do a

[7]For a discussion of feature-placing see Cussins "Content, Embodiment, and Objectivity: the Theory of Cognitive Trails."
[8]Say something about how this is only a first step ...

stunning job of putting them together in a seamless whole.

The alternative ways create something of a tension. At the local level, the richness of the real-time feature array is a boon: necessary in order to control the fine-grained detail of action and perception. If you want to place your finger *just there*, if you want to track that pattern of animal motion against a background of grasses waving in the wind, if you want to slip just so through this crack in the wall, then the rich detail provides critical resolution for fine-grained action. The superfluity of detail is not so good, however—exactly because it is *so* detailed—for long distance inference, reasoning, planning.

Some of the difficulties are epistemological (as we saw in the case of science): managing that complexity would swamp any finite computational mechanism. Some of the difficulties are ontological: long-distance correlations are often not sufficiently strong to warrant being framed in such detail. Fortunately, these two limitations conspire together: it is a good thing that we can refer to far-away objects without needing all the fine-grained detail, because by and large we do not know what that detail is. If we couldn't refer to Pompeii except by laying out the spatial configuration of every inhabitant, we could not refer to Pompeii at all, since we don't have any idea of exactly how many people lived there, let alone where there all were distributed. By the same token, if I couldn't remember you unless I was able to represent the exact position of your arms, I similarly could not remember you at all, since most of the time (unless I happen to be facing you) I do not know how your arms are arrayed. This is all banal, of course—no one doubts the ubiquity or utility of abstraction. My point is only that this abstraction is not something that happens *to* objects; rather, abstraction *enables* objects; it goes on *underneath* them. That is: objects are related to ur-reality by synthesis and abstraction. And synthesis and abstraction are *intentional*. To say of a patch or region of the world that it is an object already involves an intentional—and, to up the ante a bit, even a purpose-relative, which is to say, normatively governed—abstraction or synthesis away from the underlying sea of features.

This then is the picture I want to explore. Representing or otherwise dealing with the local, ever-so-rich feature array (the feature array in the world) is good for controlling action and responding to fine detail; it is bad for long-distance inference and generality. The non-local, long-distance abstractions underwriting conceptual ontology are good for long-distance inference and generality, but not very good as a way of capturing the unutterable fine complexity of local circumstances. Achieving a realistic picture of the world requires mutual support of both. Intelligence involves being able to move back and forth between the two, flexibly, plastically-correctly-as appropriate.

I believe this is a sensible story, making sense of a variety of truisms: why you have to "be there," in order to fully understand a situation; why there are such limits on "book knowledge," and the like. I will also want to claim, ultimately, that it has various overarching theoretical benefits—giving us for example the wherewithal to begin to steer an appropriate path between realism and social construction. Nevertheless, the picture needs defense, if for no other reason than because it flies in the face of the "well-entrenched" (to put it mildly) intuition that there really are autonomous objects out there, independent of us.

I will start on that defense by considering, respectively, what nonconceptual content, and what conceptual content, come to, on such a view.

4 Featural content

… I am not sure whether this section has been adequately written. I have the following set of notes as to an outline of what may go into it, but then there is the text that follows—which may or may not instantiate the outline. All needs to be checked …

- A. Summary
  1. Don't have time
  2. Investigate one specific kind: featural content
- B. Three major questions
  1. Featural content: really content?
     - a. Yes, because eminently revisable
     - b. I.e., can serve as reasons
     - c. Cf. McDowell's argument against it
     - d. He doesn't understand *semantically-warranted* (normatively governed) *causal transitions*

2. Featural content: subpersonal?
   a. No; perfectly accessible to consciousness
   b. Cf. Lowe slide
3. Featural content: really conceptual, after all?
   a. No; not the same content
   b. Doesn't figure in right generalisations
   c. Would swamp finite mechanism
      i. Cf. trillion lines of commercial software
      ii. Illustrates all points (normative governance, etc.)

C. Connection to action
1. Not just a question of richness of details
2. Also: indexical, differential, closer to physical coupling
3. Therefore: shade into this kind, in order to control activity
4. Tie to field-theory, differentials (deixis, etc.
5. That in turn generates why more modality specific
   a. Cf. Cussins: motorcycle, Evans "behind you", etc.
   b. And non-generality: drive as fast as hit the tennis ball

Nonconceptual content is negatively defined, hence weak. That is not its only problem; I am not convinced there is just one kind of nonconceptual content-that it names a single, unified phenomenon. For the conceptual/nonconceptual distinction to be taken seriously, we need positive accounts: of rich, delineated, self-standing kinds of content-to compare and contrast with conceptual.

In this section I want to consider perhaps the simplest such possibility, generated by the picture of finite embodied agents operating in a feature space. I will call this species **featural content**-content that takes the world to be feature-instantiating. But even it is a bit of a grab bag. I will not here (though a proper account should) examine the space of features—*egocentric*, *allocentric*, more or less physical, etc.—nor say much about how different kinds of featural content can figure in different kinds of agent activity. A genuine theory awaits such detailed cartography. Here I must limit myself to some broad remarks.

If, as suggested above, the world in itself is more featural that conceptual, then the semantics of featural representations are in a sense simpler than the semantics of conceptual representations (which we will examine later), because of the more direct "fit."

The basic idea is similar to Peacocke's scenario content: featural representations take the world to consist of a three-dimensional array of feature instantiations.

Because of their potential for richness, featural representation is paradigmatically indicated when an agent is in close contact with the world's details, presumably including perception and action. As I will explain in a moment, featural content is also the easiest kind of content for a physical mechanism to attain—it is of a sort that the laws of physics can underwrite rather directly. Intuitively, this makes sense: think of simple instruments, such as thermometers, light meters, etc. Setting aside for the moment the (critical) question of whether such devices really have content, or are only interpreted as such, it is nevertheless intuitive that, when we interpret them, a featural reading is most natural. Thus a thermometer indicates that "it is 90° here, now"—where, as in the case of weather statements, no ontological commitment is made to any object's "being 90°."

For essentially engineering reasons, it is likely that fine-grained motor control will want access to the rich, detailed, pre-conceptualised (non-objectified) structure of its environment that featural representations are ideal at conveying, and that sensory mechanisms will yield these at the onset of perception. But no logical commitment is being made, in this account, as to whether any of perception, motor control, and conscious perceptual experience use the same representational vehicles, or even represent the world in compatible ways. It is certainly no part of the view being expounded that either perception or sensation must go with featural representation; conception with conceptual.[9] By the same token, there is no logical requirement that featural (or other sorts of nonconceptual) states can only be engendered by direct encounter with the world. It is a benefit of an ontological approach that, instead of building such claims into the theory itself, it provides the wherewithal for giving such claims genuine empirical content, if they turn out to be true.

Rather than go architectural issues, however, what I want to do here is to consider three questions that will naturally be asked:

[9]Unless of course one were to stipulate the difference between perception and cognition in such terms.

(i) whether such featural states warrant the label content at all; (ii) whether they are not intrinsically sub-personal; and (iii) whether, as described, they are not really conceptual, after all.

4a Is featural content really content?

What warrants the claim that nonconceptual states—in particular, any states that arise in a creature with featural content—can legitimately be said to have content, at all? That is: how do I know that my describing states as nonconceptual isn't merely word-play?

There are two facets of essentially a single answer. Overall, nonconceptual states are governed by the same normative/semantical considerations that apply to conceptual content. As such, they can serve as rational (or at least normatively appropriate) reasons for an agent's action. There is no implication, from that critical fact, that they must thereby be able to be given "conceptual voice." It is not an *a priori* truth, after all, that if a content-bearing state is a reason for an agent's doing something (bending its arm in a such a way, feeling that danger impinges, whatever) that, if asked, the agent must be able (even potentially) to articulate its reason. In fact commonsense, lay experience, artistic sensibilities, psychoanalysis, cultural anthropology, and just about every other form of human study suggests the opposite. That is not to deny that it is philosophically common to assume that reasons must be articulable; the point is only that, in the current context, to presume that would be empty. All that that assumption comes to is a claim that the only genuine form of content is conceptual—exactly what is being denied.

Moreover, featural content is eminently revisable. There is nothing about nonconceptual content that means that it just is what it is, independent of semantic constraints from the world it represents. Nor—it must be emphasised—is nonconceptual content in any sense "given." By analogy, consider photographs. The content of photographs is surely derivative, not original—but qua derivative content, it is also surely nonconceptual. The photograph itself is a two-dimensional spectral density array, representing something like a three-dimensional reflectance array that laid out in front of the camera. Suppose, while looking over some photographs, you come upon one which does not look right, and say "Oh dear; this one is a double exposure!" Why do you say that? Because the world represented by the photograph is not possible, or anyway so unlikely as to warrant extreme doubt. By the same token, imagine a depiction (painting or photograph) of dense shrubbery surrounding a path across a rocky slope, with wisps of fog swirling up from a valley below. Now that description is necessarily in language (this is a philosophy paper), and will therefore have conceptual content;[10] but what I am asking you to imagine is one of the infinite variety of visual scenes from which that conceptual abstraction is possible. Now ask what that scene would look like if one were to take a few steps forward. There are constraints on what would be a legitimate answer. Some depictions are incompatible, some depictions are wrong—depictions that, if they were delivered to you in consciousness by your perceptual system, would make you rub your eyes and look again (or perhaps rub your eyes and try to wake up).

Conceptual representation, in sum, has no patent on revisability. On the contrary, featural and other nonconceptual states are just as amenable to semantically-warranted transition—and hence are just as capable of semantically-unwarranted transition. So they can serve as reasons.

4b Is featural content subpersonal?

Some may argue that if there are states with featural content, in the way I am proposing, then they must be "subpersonal." But I believe that is manifestly false. For nonconceptual states as I have described them are available to consciousness. Indeed, it is (perfectly realist) consciousness of the non- or pre-conceptualised world, I believe, that has led legions of philosophers to believe in a spate of such ontologically unfortunate entities as sense data, visual fields, "experiences," opaque mental states that can be reflected upon independent of their content, etc.

Another example: suppose, visiting in California, you say "The fog has come in again." Suppose I, in turn, say to you: "You have a PhD. In saying 'the fog has come in,' do you mean the same instance of fog? or a new instance of fog of the same type?" There is no appropriate reply. There is no reason to suppose that, in mak-

[10] I don't quite believe this, of course ...

ing your original statement, you are epistemically committed, or have made any ontological commitments to, an object that is "the fog," Rather, you have merely judged that the feature "fog" is reinstantiating itself again, around here, around now.

But that is a linguistic example—which is distracting (since we are largely associating conceptual content with articulated language). A better example is given in figure 1. This image (by Adam Lowe) can be "parsed." It depicts the painter's studio, with a door on the right, a waste can on the floor to the left of the door, a cheap phone attached to the door frame, also on the left, and a person's body moving towards the door, from the right.

Pictures such as this—and even more so, pictures that look roughly like this, but that cannot be so readily parsed, or that cannot be conceptually parsed at all—are paradigmatically labelled abstract. I think that labeling is exactly backwards. It is so-called "representational" pictures that are abstract, by my lights. They discard the rough and tumble of the world—"over-neaten" it, pull out and present to conceptual judgment, in imagistic form, what our conceptualising faculties do to the world, in normal perception. It is Lowe's painting, in contrast, that is concrete-concrete in the sense that it depicts the concrete world as it is.

Admittedly, this image may be slightly exaggerated—but only a tiny bit (remember those robot cameras). That is because what the painter is trying to do, with this image, is to bring to our conscious awareness an inkling of how the world presents to our unconscious or artistic sensibilities.[11] Forget the image, therefore, and reflect on the world. The image is merely intended instrumentally: to bring us to reflect, consciously—perhaps even with conceptual supervision and commentary—on the pre-conceptualised reality that is what we regularly look out upon, whenever we open our eyes.

[11] If one were mathematical, one might say that if the conceptualising faculties of judgment that conceptualise the world transform it according to some function f, then Lowe has painted this image to be something like f^{-1} of what the world is really like, so that what arrives into conscious conceptual judgment is, as closely as possible, the nature of the pre-conceptualised world.

4c Isn't featural content conceptual after all?

One final objection must be dealt with, which is likely to be raised to the picture developed so far: that there is nothing nonconceptual about the sorts of featural representations being discussed. "Surely," such a person might say, "a nonconceptual featural representation of the world can be expressed, conceptually. Simply take each point in space and time as an object, and predicate, of that space-time point, a property that corresponds (isomorphically) to the there-placed feature. For each point in the depicted region, one can simply predicate the appropriate colour value of that point."

There are number of things to say, by way of reply. First, there are problems of continuity. Conceptual representation seems to involve a certain degree of digitization, and featural representation, at least as I have presented it, can be, and perhaps most often is, continuous. That suggests that one would need to digitise the image first. And imperceptible digitization is no simple affair—as decades of work on graphic displays and audio in the computer and entertainment industries attests. In particular, there is perception's notorious non-transitivity of indiscernibles: the fact that there can be three colour patches, *x*, *y*, and *z*, such that a subject cannot tell the difference between *x* and *y*, or between *y* and *z*, but *can* distinguish *x* and *z*.[12]

---

[12]For example, someone might suggest that one could digitize an image, at some level of resolution finer than the optical resolution of the eye, and store colour values for each pixel, making just as many distinctions as are required to meet "JND" properties of the visual recognition system. This is not such a simple thing to do. The point of JNDs is that there can be three colour patches, *x*, *y*, and *z*, such that a subject cannot tell the difference between *x* and *y*, or between *y* and *z*, but *can* distinguish *x* and *z*. If one were to record colour at the level of resolution of the eye—i.e., one bit per JND—then it is clear that *x* and *z* should receive different codes. But what about *y*? It cannot be given the code for *x*, or the code for *z*, and reproduce the same phenomenology.

This means that the digitization of recording must be finer than the resolution of the sensory system. How much finer is a matter of intense debate. The case of digitised audio is instructive. When compact discs were first produced, it was widely believed that the then-standard digitisation standard—16 bits of information 44,100 times a second—would suffice, since the upper cutoff Nyquist frequency was above the range of human hearing, and the resolution of a 16-bit amplitude (1 part in approxi-

But in a way that misses the point. For while in some abstract sense it may be possible to construct a conceptualisation that has the same information content as the original image, it does not follow that it would have the same content. It will not figure in the same semantically-warranted generalisations. For imagine: take Lowe's image, given above, to be an image of his studio, taken from a perspective a few feet away from the door. How would the image change, if one were to step a few feet to the left? Nothing in the brute-force conceptual recording of the image gives one any help with that.

Moreover, there is no reason to suppose—and every reason not to—that we represent images in such a way as to have ready access, meshed with our more general conceptual powers, to such a detailed recording of our visual experiences. The data and processing load this would demand would be overwhelming. And if cognitive science has taught us anything about the architecture of the mind, it is that considerations of computational complexity, even in a mechanism as dauntingly impressive as the brain, are of the utmost importance. Moreover, as mentioned above, a conceptualisation that relies on space-time points (i.e., on non-extended, non-durable objects) is conceptualisation in letter only; it does not deal with any of the issues—of synthesis, abstraction, reidentification, life-time variability, etc.—that are constitutive of objects in real life. Space-time points are not really objects.

In passing, it is perhaps worth pointing out, in this vein—for people tempted by such conceptualised recording—that essentially all modern computer software design can be viewed as implementing semantically-warranted causal transitions on nonconceptual representations.[13] I estimate that something like a trillion

mately 32,000) was well below the audible JND of volume discrimination. It is now widely believed that those standards were insufficient: CDs are audibly inferior to the resolving power of the human ear. The professional audio industry is now moving to a more informationally-dense standard of 24 bits of information sampled 96,000 or even 192,000 times a second. But the data implications are enormous: uncompressed, that implies something approximately 1 gigabyte for 40 minutes of music, which even today is substantial storage.

[13]Except for the vanishingly small number of AI expert systems.

lines of computer programs have been written, to date; virtually none of it uses conceptualised representation. Because of the egocentricity, purpose-specificity, and contingency of the tasks that computers are up to, it is vastly more efficient and reliable-indeed, it may be the only possible way to get anything done-to use purpose-specific representations. So no one should think that purpose-specific algorithms, such as "abstracting" algorithms, in order to get to the long-range conceptualised skeletons of the world's nonconceptual detail, are in any sense precluded.

4d Connection with causation

To conclude, set objections aside, and consider one final argument in favour of featural representation.

Two things that I have not emphasised here, but that need to be dealt, include:

1. Nonconceptual representational vehicles, including featural vehicles: i.e., the structure of the representation that bears nonconceptual content, viewed as a causally-efficacious entity (the problem of mental content, after all, stands as need of solution in the nonconceptual case as in the case of conceptual content); and
2. The semantic (interpretation) relation between the vehicle and the (nonconceptual) world thereby represented.

But some suggestions along these lines have been implicit in various examples I have used, such as photographs, TV cameras, and recording instruments. This ties into the statement made earlier, about the field-theoretic nature of physics, and the suggestion that science is interested in types (or features).

According to the broadly representational theory of mind we are working under, an agent works, causally, but subject to governing norms in such a way that it is not a "mere" mechanism. That means that an embodied, embedded agent—as all agents must be—will be causally plugged into their environments (though not, again, merely causally plugged in). As I explore in detail elsewhere,[14] it turns out, given the way causation works—i.e., as a consequence of the nature of physical law—that the form

[14]«Reference O3; also "Who's on Third?"»

of representation that is easiest to have is nonconceptual: one whose features correspond, moment-to-moment, with the features of that with which it is causally coupled. That is why instruments—thermometers, microphones, photographs, cameras, etc.—tend to be such good examples of featural (nonconceptual) signifiers.

It also turns out, for similar reasons, that the content of the simplest form of representational mechanism will be egocentric or indexical—"deictic," as I have put it. This fact has profound implications for the nature of first-person reference and consciousness awareness; it also establishes the nature of the task that an agent faces in order to have objective content (be it conceptual or nonconceptual). For now, though, the point is that it is a consequence of the nature of underlying physical laws that the sorts of structure that must guide action must be vehicles with egocentric, nonconceptual content that structurally, qua vehicles, as they get closer and closer to the world, grow increasingly isomorphic or iconic to the content they carry.

5 Conceptual content

… Again, unsure about the relation between this outline and the following text … whether it is complete, etc. …

- A. Conceptual content
  1. Turn then to conceptual content
  2. In a way, even more interesting, because of lack of fit
  3. Once again, deal with three major topics
- B. Skip the first two
  1. Structural correspondence
     - a. If underlying reality is fields of features …
     - b. What is object reference (objectification) like?
     - c. Two cross-cutting algebraic kinds of correspondence
       - i. Property / object (as in Generality condition)
       - ii. Singular object reference
         - — Has to do with name ≈ type, object ≈ instance
         - — Not point to point
       - iii. Field theory of object reference

   d. Complicated by stabilisation, deixis, first-, second-, and third-person objectivity, etc.
2. Relation to nonconceptual
   a. Basic: non-conceptual anchors conceptual
   b. Not logically identified with perceptual / cognitive (()
      i. Allows cross-fertilisation (interpenetration)
   c. Make sense of many truisms
      i. Limits of book knowledge
      ii. "Had to be there" (walk in another one's shoes, etc.)

C. Norms
1. Third one has to do with norms.
2. Want to wrap up with this.

Turn then to conceptual content: what it is, how it could arise in a featural world, how it is tied to (or grounded on) nonconceptual content. In a sense, this is the interesting case, because of the lack of obvious fit. How, if the world is not (aboriginally) conceptual, can claims with conceptual content be true?

Conceptual content takes the world to consist of objects, exemplifying properties, standing in relations, grouped in sets. Rather than being metaphysically basic, conceptual ontology is "constructed" by intentional creatures, using processes of abstraction, out of an explanatorily and ontologically prior world (a world we are for now taking to consist of a vast array of instantiated features). The question is how those abstraction processes go.

I want to consider three issues, as a way of getting at the answer. The first has to do with the form of correspondence that conceptual representation bears to the world. The second concerns the relation between conceptual and nonconceptual content. The third, which is also the most consequential, has to do with the norms on which the abstraction processes are based.

5a Structural correspondence

We have identified two features of conceptual content: abstraction and recombination. A modal claim was made, about the (potentially) recombinant structure of conceptual ontology: that if *a* is *F*, and *b* G, then *a* might have been G, and *b* *F*. This potential for recombination must be reflected in the conceptual vehicles. It is a constitutive condition on conceptual abstraction, that is, that it eventuate in a kind of Evansian generality—guaranteed by appropriate patterns of rational inference.

The potentially recombinant world, that is, is reflected in potentially recombinant representation. It is not, of course, ultimately reflected isomorphically: negation, disjunction, numerals, and quantifiers all famously break any one-to-one correspondence between language and world. Historically, however, it is possible that it started out isomorphically:[15] "This is blue," "Pat is eating dinner." But isomorphism is broken the minute 'two' is introduced: 'two' represents duality with unity.

That claim, about isomorphic or partially-isomorphic mappings between representation and represented, is couched in terms of conceptual registrations of both realms. Things get much more interesting when we look at the correspondence (interpretation) of conceptual representations, but understands the represented world featurally. For what emerges is that the recombination of parts is only one form of vaguely algebraic coupling; another one, underneath the objects, is explanatorily more basic. One of the characteristics of featural representations, mentioned above, is that they typically (at least in the simplest cases) involves a point-to-point correspondence between vehicle and content: at time t, the sunflower points at the direction of the incident sunlight at time *t*; at time *t+1*, it points at the direction of the light at time *t+1*.

But think about reference to an object—say, with a proper name. Suppose the name 'Pat' refers to a person, Pat. We think of this as a one-to-one correspondence: one name, one person. But of course that is a distracting way to put it, since the name is a *type*, the person, an *instance*. There are instances of the name-uses or utterances, that typically occur at a specific moment in time. As our experience with indexicals has taught us, it is these temporally-specific uses that refer. Since objects do not exist in the world independent of being objectified, on this story, there is a also a sense in which there are "instances" of objects: namely,

[15]Something of the sort is suggested by Terrence Deacon—though since his semiotics is not very developed, it is hard to know whether the thinks that this was only an evolutionarily transitional stage.

those temporally-specific "manifestations" or "time-slices" of objects that also "occur," at different moments in time. Crucially, however—and this is the important point—*individual temporal utterances do not refer to individual time-slices of their referents.* When I thought of you last night, I did not just think of you-last-night, or (even less) of the then-occurrent instantaneous time-slice of you. Rather, I thought of how you were doing at school, whether you were prepared for today's exam, etc. Today, when I thought of you, I again did not think of today's time-slice, but again of you as a temporally extended entity. In other words—as depicted in figure 2—each instance of a proper name type refers to the full extended space-time worm (or whatever region of the infinite flux) you constitute. So the featural (or field-theoretic) structure of even a simple name-object relation involves various forms of *cross-cutting coupling*. All of this is required—is an achievement of subjects—in order to refer to an object as an object (and referring to objects is surely one of the most basic capabilities of conceptual representation). The complexity of this cross-correlational mapping, as compared with the simple form of point-to-point correspondence characteristic of featural representation, underscores the significance of the accomplishment that is intrinsic to conceptual abstraction.

5b Relation to nonconceptual content

Even on a classical account, few would deny that thoughts are anchored in perception and action—that human reference is grounded in our engagement with the world. But on the story being told here, a much stronger moral emerges, having to do with the relation between conceptual and nonconceptual representation.

We can get at this moral by noting two facts about conceptual content.

First, conceptual content involves loss. When we take the world to consist of objects exemplifying properties and standing in relations—when, that is, we "objectify" the world—we discard staggering amounts of information (the vast majority we are presented with, in fact). Remember those robot cameras; once again, computational experience is a sober reminder of the prowess-this time, the "forgetting" prowess-of the brain. And as I have said, it is fortunate that we shed this much detail. Given finite computational resources, it is only with the radically pared-down result that we have even a prayer of doing passable inference. (That's one reason conceptual representation is valuable; if one tried to compute with full featural maps, the computational load would be intractable.[16])

Second, conceptual representations are *disconnected from* their referents. That representation be able to be disconnected from its reference is well-recognised; it is that ability that allows us to hypothesize, to refer to things beyond the reach of our senses, to have a sense that there is a world out there, beyond the reach of our senses. Once one recognises that objects are an abstraction over the world, not part of the world's aboriginal structure, though, one is forced to realise that essentially all representation is disconnected-if for no other reason than that objects by and large are perduring-exist through time-whereas all that physics allows us to couple with (because of its locality) is the present moment.

In sum, conceptual content sees the world "through a glass, darkly." It "lets go" of the world, discarding vast amounts of detail, so as to support generalisation and long-distance inference. There is a worry, however, given the depth and human centeredness of this abstraction (more on that in a moment), that the conceptual content will take leave of the messy details of the world altogether, and float entirely free.

But of course that is exactly one of the roles of featural content: that it anchors the "abstracted" conceptualised objects built on top of it. Nonconceptual content is the "glue" that binds abstracted objects and properties to the pre-objectified world. It is what keeps the fact that even concrete objects are abstracted from implying that they take leave of reality. This is why I said above

[16]Why the world should be such that conceptual abstraction works is a non-trivial metaphysical question, which I will not address here except to note that what it seems to work best for are artifacts, which we build-perhaps with malice aforethought. Just as language has evolved subject to the constraint that the human brain can speak it, so too artifacts may have developed subject to the constraint that the human brain can understand them.

that nonconceptual content is necessary in order to retain what is right about realism.

What this consideration shows, however, is that abstraction is just half of the story. For as normally conceived, the term 'abstraction' refers to the processes of "letting go" of the world: to the discarding of the mass of featural (and other nonconceptual) information, so as to achieve a finite, compact, gloss on what is the case. But that describes perception; what about action? What happens when, on the basis of a conceptualisation of the world, we reach out to do something?

What happens, I believe (in part for the reasons cited above, about the closer fit between featural representations and causally coupled mechanisms) is that *our nonconceptual faculties enter into the equation so as to fill back in the requisite detail*. Thus imagine deciding to reach for a cup. You have that thought: "I will pick up this cup." And it may even be that in the conceptual thought is some indication of the size of the cup: perhaps it is a small latté, or large mug. But then, as your hand approaches the cup, your fingers adjust, through representational mechanisms (prior to contact), so as to be prepared much more exactly than they could have been, in virtue of conceptual content alone. The same for serving tennis, for leaning into a corner on a motorcycle, and so forth.

In previous work I used the term 'reconciliation' for this process that is the opposite of abstraction—this "filling back in" of the world's detail that is lost when one conceptualises.

… use: 'concretisation' …

In sum, conceptual abilities are required to conceive of an object as an object, to conceive of an object as conceptualised. Nonconceptual abilities are required in order to understand that which is conceptualised as an object. Only if you understand that an object is a conceptualisation of reality do you really understand what an object is. So nonconceptual content is thus not "optional"; it is a necessary ingredient to objectivity.

This conclusion contains strong lessons for AI. It implies that purely conceptual creatures have no chance of achieving objectivity, because they exactly lack the critical (nonconceptual) glue that binds their conceptual conceptions to the gritty stuff and substance of the world. No wonder "book learning" is limited-and eerily detached.[17]

5c Norms

Third and finally, consider norms. In a way, the point is simple: Those constitutive processes of abstraction, qua epistemic practices of rational, norm-governed agents, happen for a reason. It is the ontological character of that statement that makes it strong. It is not just that there is a reason people represent objects, in other words. That much is obvious—and anyway guaranteed by the normative character of the semantical story within which we are working. The point is stronger. Which abstractions a creature makes—and as a result, what objects there are in the world, for that creature—arises out of the constitutively norm-governed life that that creature leads.

Objects themselves, that is, not just their representations, have their existence in worlds governed by significance, interpretation, and importance. To put it in a two short words: **objects matter**.

This claim has a happy terminological consequence. In §1, I distinguished *physical ontology* (the strange world described in modern physics) from *material ontology* (the everyday world of human experience). Now it is common to suppose that the word 'material,' in English, has two distinct senses: (i) a more common one, meaning something like physical or bodily, as in "living in the material world," materiality, materialism, and so forth; and (ii) a less common, vaguely legalistic one, meaning something like impor-

[17]Note an irony to the story I've been telling. I started out saying that I wanted to "ontologise" conceptual and nonconceptual content-to push it out from heads into the world. Now, however, it seems that I am allowing the nature of what is "in the world" (particularly in the case of conceptual content) to slip back partway into the agents that inhabit it. There is a grain of truth in this blurring of the subject/object boundary. Still, this is absolutely not a story that devolves into pure idealism or vacuous relativism; that would only be true if the reality (the reality that for the moment we are characterising as featural) had no grip on the thereby-conceptualised objects' nature. But that is no implication of what is being claimed.

tant, as in a "material argument," or "material consideration."

In calling everyday ontology material, it may have seemed as if I was recruiting the first, roughly physicalist sense. But my intent was more devious. For what I am suggesting—which we can now see—is that there are not really two senses, after all. If, as I claim, the processes of synthetic abstraction constitutive of conceptual content are anchored in the norms governing the lives of conceptualising creatures, then material ontology (as I have defined it) is in part normatively derived. *To be an object is to be important* (to someone), in the world. It is because of this fact—that objects *do* matter, not just that they *are* matter—that I called them material.

Part of what we are doing, that is, in this project of naturalising abstraction, is healing the 300-year gap between matter and mattering.

To see what this comes to, let's bring it to bear on an issue that has been lurking in the background for some time: what it is (on this picture) for a conceptual judgment to be *true*.

I trust it is clear why this is a non-trivial issue, on the picture I am painting. I have claimed, after all, that objects, *qua* objects, are not wholly independent of people (of us). I have depicted featural content as closer to the "mind-independent" structure of the world than is conceptual content (though, to repeat, it is only *closer* to reality; I am not saying that it is reality—remember, we are just taking a first step down that plank). So it looks as if featural content has a better claim than does conceptual content on being true, perhaps even on being objective. But that cannot be. It would be perverse—even nihilistic—to deny to conceptual judgments the possibility of truth. Rather, the question we must ask (in a spirit of reclamation) is this: *what can or does truth mean, for conceptual judgments, on an abstracting, human-implicating picture?*

I want to get at the answer by going back to where we started: with the representational theory of mind. In broad brush strokes, I characterised that view as committed to a picture of mental life as involving semantically-warranted, normatively-governed, causal processes. At that very general level, I remain sympathetic to the view. It is (among other things) in the details of how the norms are treated that the view I am proposing radically parts company with standard accounts.

To see why, consider the classical (logicist) picture. It works as follows: one starts by distinguishing (static) states from (dynamic) processes defined over states. Given this distinction, the norms then attach in stages. In the first stage, semantic evaluation is defined for the states—in a way that is assumed to be explanatorily prior to, and independent of, their use in inference or reasoning. Then, with truth and reference in place, a second set of norms is defined for the processes, in terms of that presumptively prior semantic valuation. Once we realise that semantic evaluation is a species of normative governance (truth being better than falsehood, information better than misinformation, etc.), this classic view can be summarised as follows. What I will call the **dynamical norms** (norms on activity or use, such as on inference) are assumed to be explanatorily derivative on the **statical norms** (norms on states—such as reference, truth, etc.). This form of asymmetrical dependence underwrites all standard accounts of soundness, completeness, truth-preservation, etc.

Experience with real-world systems, however, shows us that this strategy does not work. It turns out to be impossible to assign semantic evaluation prior to and independent of activity. Rather, over the last few decades, in a perhaps unwitting endorsement of a vaguely Wittgensteinian doctrine of "meaning is use," computer and cognitive scientists have all come to lean in the opposite direction. They have shifted to the opposite form of explanatory dependence, with semantic evaluation, content, interpretation, etc., taken to derive from large-scale dynamic activity. To put it in terms of the terminology just introduced, it may not be recognized as such, but it is nevertheless virtually universally assumed that *statical norms derive from dynamical norms*.

This shift is *unimaginably consequential* (in spite of the fact that no one seems to be noticing it).[18] What makes it so important is that it requires, on pain of circularity, that something else (other than "preservation of the static norms") ground, or serve as the origin of, the dynamic norms.

What are some plausible dynamic norms? Several alternatives

[18]There seems to be a kind of "Road-Runner" effect: everyone has rushed off the cliff, but no one has yet looked down.

have been pressed into service: meeting a specification, maximizing an externally-supplied value, etc. These days, however, especially in cognitive science, philosophy of mind, and evolutionary epistemology (to say nothing of Artificial Life and the theory of complex adaptive systems), the dynamic norm most in favour is that of *adaptability* or *evolutionary survival*. (I have in mind for example Ruth Milikan, teleofunctional semantics, notions of proper function, etc.)

Whether evolution will prove strong enough to anchor the range of dynamical norms needed in order to explain the human condition is of course a matter of intense debate. I am not going to broach that topic here, except to say that I do not believe it. I do not think evolution is remotely strong enough; I am not even sure it has the right categorical form. For note how much is being asked. The full dynamic norms on the condition of a person will among other things include *ethics* and *significance*—how to live, what to do, what constitutes a meaningful life.

However those questions work out—whatever dynamical norms govern human (or humane) lives—what is relevant here, and this I do want to insist on, in the present context, is that they do not just apply to reference and truth; *they also apply to ontology*. If the story I am telling about conceptual content is right, that is, dynamical norms underwrite not only the semantical content of an agent's representational and intentional states, but also the constitutive normative standards in terms of which the agent forms its conceptual abstractions.

And with that we are finally ready to answer the question of what it is for a conceptual judgment to be true. A conceptual judgment is true when the thereby-abstracted situation satisfies the dynamic norms governing the lives of the creatures who perform it (i.e., the creatures who objectify its constitutive objects, delineate its constitutive properties, and so forth). Roughly, that is, a conceptual judgment is true just in case conceptualising the world in that way—including not only the patch of the world thereby conceptualised, but also the act of so conceptualising it—is a "success maker" for the objectifier with respect to the overarching norms that govern that objectifier's projects.

It is a theorem of this view, that is—a consequence of the recognition that objects matter—that (conceptual) truth is to a degree pragmatic. To put it in a slogan, conceptual truth (that is, truth applied to states or judgments with conceptual content) ultimately depends on living truly.

Some may still object that even if the choice of how to abstract or conceptualise is purpose-relative—perhaps norm or project relative—the "space of possible abstractions over real-world features" is nevertheless already out-there. And so, they might say, the story is realist after all. Conceptual ontology is not so much *constructed*, on this story, they might say, as *selected*. If one chooses to call this realism, that is okay with me. But I believe it is a misleading way to view things. Consider something as simple as a 16 × 16 array (the number of pixels in used for the cursor on your computer). The number of ways to colour them (i.e., the number of possible distinct cursors) is hundreds of thousands times greater than the number of electrons in the universe. The intrinsic choices for grouping are so vast, that is, that most of the nature of the choice derives from the constraints one obeys in making it—constraints (to return to the case at hand) that derive from the creature's normatively-governed life. The nature of the abstraction, that is—the abstraction that the object, *qua* object, must normatively honour—derives in part from the intentional practices of the representing agent. And that agent's practices are grounded, ultimately, on anchoring in the sustaining field of features—or more generally, in the nonconceptual world.

Moreover—to shift up one level—there is no guarantee that the norms that ground this conceptual abstraction will themselves be conceptualisable without loss. Moral realism, to take one striking example, does not imply moral effability.

6 Conclusion

More can be said—but time has run out. What have we learned?

The metaphysical morals are the strongest. It is not just semantics that needs naturalising; ontology needs naturalising too. That is the first lesson. At least material ontology needs to be naturalised: the familiar everyday world of objects, properties, and relations. The processes of abstraction that underwrite the grouping and individuation of nature into material individuals, essential to our understanding of cognition, are no less mysterious, no more secured by a mechanistic or causal scientific world view, no more automatically integrable with results in contemporary science, than any of the other challenging features of intentionality. To assume that the world of objects, properties, etc., exists independent of us, in fact, as naive realism and any commitment to the "natural ontological attitude" would suggest-to think that the world is autonomously conceptual-is an especially pernicious way of succumbing to the Myth of the Given.

In sum, this is a picture of *metaphysical monism*, but *ontological pluralism*. And to my sensibilities, that seems just right. It allows us to do justice to the humility that underwrites constructivism: the recognition that we are here, and have an impact on the worlds we inhabit. And it allows us to do justice to the humility that underwrites realism: the recognition that we are not all that is here, that ultimately we have to defer to that world of which we are a part, on which we have an impact, but that we do not control-and certainly cannot ultimately grasp.

What we do is to make our way as best we can—neither impotent nor omnipotent, neither ignorant nor omniscient. We live, that is, in something of a middle ground—in a continuous creative tension between the incredible richness and unabstractable detail of local coupling, on the one hand, and the long-distance utility of language, inference, and abstraction, on the other. When we engage directly with the world, we want to do the opposite of "abstract": we want to to *concretize*, to reconcile our ideas with reality, to let more of the world's ultimately ineffable details fill our representations, in order to be appropriately responsive-in action, in perception, in local, contingent reasoning—to the world's fine-grained, particular, structure. When we want to travel long distances—in order to conceive of the world as a whole, in order to create complex institutions, in order to do science—it pays to let go of that overwhelming profusion of local detail, and employ sparer, more efficient methods—methods purpose-designed for inferential travel. If we are clever (and surely we are clever) we can—in fact must—do both, in such a way that each props up the other, thereby allowing the nonconceptual representations to approach a kind of objectivity, and the conceptual representations to, in their own way, be true.

Perhaps the best way to summarise this is by an analogy. I sometimes think of objects, properties, and relations (i.e., conceptual, material ontology) as the long-distance trucks and interstate highway systems of intentional, normative life. They are undeniably essential to the overall integration of life's practices—critical, given finite resources, for us to integrate the vast and open-ended terrain of experience into a single, cohesive, objective world. But the cost of packaging up objects for portability and long-distance travel is that they are thereby insulated from the extraordinarily fine-grained richness of particular, indigenous life-insulated from the ineffable richness of the very lives they sustain.

6 — Dennett on Smith (and reply)[†*]

Hugh Clapin
University of Sydney

Brian Cantwell Smith holds a rare and valuable intellectual pedigree for a philosopher of mind. As a principal scientist at the Xerox Palo Alto Research Center (PARC) and founder of the Center for the Study of Language and Information at Stanford University (CSLI), he has studied foundational questions in computability and computer programming. Through this work he has come to the conclusion that the representational capacities of artificial systems such as computers raise profound metaphysical and epistemological questions.

In his "One Hundred Billion Lines of C++" (1997), Smith illustrates how misleading is the ordinary philosophical conception of computer programming. Standard programming practice is not (as is often assumed) committed to classical cognitive architectures. In particular, the processes implemented by executing programs have nothing like language of thought structure; none the less they make use of representations successfully to negotiate the world. They provide a rich resource of physical representation systems that are effective but don't fit the ordinary analyses of the philosophy of mind.

Smith's work may be aligned with the situated cognition tradition

---

[†]Originally published in Hugh Clapin (ed.), *Philosophy of Mental Representation*, Oxford University Press, 2002, pp. ■■–■■.
[*]Coach House Institute, Faculty of Information, University of Toronto
90 Wellesley St W, Toronto, Ontario M5S 1C5 Canada

due to Barwise and Perry (1983). This approach emphasizes the importance of context in determining meaning. The situated semanticist is inclined to begin her theory of meaning with indexicals and other radically context-sensitive representations. Tokens of 'I' have very little meaning independent of how, when, where, and by whom they are used. More generally, the situated approach to cognition places significant emphasis on the contribution of the situation of the organism to that organism's cognitive processes.

Smith argues that as soon as we register the world using a system of representation, we make a set of strong assumptions about the way the world is. His task has been to show the profound consequences of this insight for the study of systems of representation.

Smith makes use of an engaging imaginative strategy to draw attention to the theoretical moves required to explain the occurrence of representation using only the resources of a representation-free physical world. Smith urges us to consider whether we need to think in terms of objects at all. Might an ontology consisting only of Strawson's (1959) 'features' be sufficient? When we declare that 'It's raining' we are drawing attention to a feature (raining) without being committed to any particular object that has that feature. Smith suggests we begin by thinking of the physicist's world as populated not by objects but field-densities. This field-theoretic description can be comprehensive while admitting only of field-densities for a small range of properties (for example, gravitational fields, electromagnetic fields, etc.).

Smith suggests that the common-sense world of middle-sized objects is an achievement of our representational practices. Representation is achieved when one aspect of the mish-mash of fields is able to separate in a certain way from the rest of the mish-mash. This region, the 's-region', is (or is becoming) the subject-something that represents the world. Smith first emphasizes the distance required between the representation and the represented, and secondly the need for coordination between the two. This coordination is likened to the actions of an acrobat who dances around a stage, but keeps a torch beam focused on one spot. The torch must undergo dramatic changes in orientation to maintain its focus at one point. The intentional acrobat is similarly dynamic in keeping its intentional objects stably registered.

Smith builds on this fundamental picture to argue that all representation is partly context-dependent, or deictic. Smith is scrupulous about the reflexive morals which thus apply. Acts of representation *bring the world's objects and properties into being (as objects and properties), and any attempt to talk about the world will be an act of representation, and thus an act of object-making. This makes likely what Smith calls 'inscription errors' or 'pre-emptive registration'. For example, it is difficult to talk about the world except as containing objects with properties. But if this is due to the subject-predicate structure of language, then it would be an error to infer that the world must be so constructed.*

In short, Smith says that representation is an immensely complex, powerful, and sophisticated achievement of the physical world. We are so adept at representing that we are apt to neglect this point and think it an easy and simple procedure.

The paper 'Rehabilitating Representation' (forthcoming c) amply illustrates what Smith takes to be the more practical implications of his view. Both classical and embodied/embedded approaches to cognition misunderstand representation. The former places too much emphasis on formality and the non-semantic; the latter places too much emphasis on the causal, local interactions between the system and the world, underestimating the importance of disconnection to intentionality.

The rehabilitation required involves acknowledging that representation is about causal connectivity to the world, but not a direct, local, or simple connectivity. Representing subjects, by virtue of their representations, participate fully in the world (not just the skin boundary of the world), help constitute the world (by virtue of the entanglement of ontology with representation), but are able to maintain a separateness from the world.

Dennett, despite being a self-proclaimed 'reluctant metaphysician', is sympathetic to Smith's metaphysical project (though perhaps is not completely converted). His dispute with Smith concerns the role of evolution in explaining the difficult achievements of representation and objectification. Objectification, says Dennett, is an evolutionary 'Good Trick', which was likely to be stumbled on because it provides significant selective advantage. Dennett also objects to what he takes to be Smith's commitment to the determinacy of mental content.

6a — Brian Cantwell Smith on Evolution, Objectivity, and Intentionality

Daniel C. Dennett
Tufts University

1 An Original Account of Intentionality and Objects

Like the rest of us, Smith wants to steer between the Scylla of GOFAI and the Charybdis of Dynamical Non-Representational Systems, and he adds to the feast his own bounty of acute observations and tempting proposals about how such a rehabilitation of mental representations would go. But he and Haugeland, unlike the rest of us, are ontologists who think we need to reach way back and rehabilitate the whole of metaphysics in order to do this job right. Yikes.

What are the less radical alternatives? One might have thought we could safely presuppose the usual catalog of physical objects-ranging in size from sub-atomic particles through tables and mountains to galaxies-and their properties-mass, charge, location, shape, color…and then simply explore the question of which complicated organizations of such objects count as believers, or representations, or symbols…and why. That is the strategy that has worked so triumphantly for magnetism and metabolism, photosynthesis and jet propulsion. Why not for mental representation, too? If we can explain growing an apple, and eating an apple, why not seeing an apple and wanting an apple and reidentifying an apple?

Why not indeed? I have always been a reluctant metaphysician, and Rob Cummins and Andy Clark seem to me to have shared my optimism about the innocence of the standard inventory of what we might call the ontology of everyday life and engineering. We happy sailors on Neurath's ship resist the alarm calls of Smith and Haugeland.[1] Do we have to put on our life-jackets and jump overboard and get all wet doing a lifeboat drill? Maybe, and maybe not. But it can't hurt. A lifeboat drill is a great way to reassure ourselves that we know what we're doing. And actually going through with it-not just imagining going through with it-is the only way to get this reassurance. If we end up with pretty much the same inventory and explanations we thought we were going to use in the first place, it will be a sounder ship that continues the voyage. And maybe we'll discover something important that has been distorting all our other projects.

For anyone who shares my conviction that traditional or 'pure' metaphysics is a played-out game, a mandarin pursuit so isolated from the rest of human inquiry that it is extremely unlikely to find enough leverage to move us from our comfortable habits, Smith's project is apt to be appealing. Only somebody coming from outside philosophy, somebody whose driving problem is not philosophical but somehow more 'practical' (however abstruse and theoretical relative to farming or building bridges) could hold my attention in a metaphysical exercise, and Smith has been led to his metaphysical vision by decades of struggling with problems that are eminently practical-problems arising not just in the crypto-philosophical arena of AI, but in engineering, for heaven's sake, in the design of hardware and software for all manner of applications. His title "100 Billion Lines of C++" sings to me, then. If disk operating systems, word-processors, and web-browsers confront problems of reference and meaning that can only be alleviated by some revisionary metaphysics, I am all ears. But still, dragging my feet. Constructively, I hope.

Let's start with what Smith calls The Representational Mandate:

The Representational Mandate

1. Conditions
    a. A representational system must work, physically, in virtue of its concrete material embodiment (the role of effectiveness).
    b. But it is normatively directed or oriented towards what is non-effective-paradigmatically including what is physically distal.
    c. Being neither oracle nor angel, it has no magic (non-causal, divine) access to those non-effective situations; just caring about them is not enough (physical limitations bite hard!);

2. So what does the system do?
3. It
   a. Exploits local, effective properties that it can use, but doesn't (intrinsically) care about-i.e. inner states of its body and physical make-up, in interaction with the accessible (effective) physical aspects of its environment.
   b. To 'stand in for' or 'serve in place of' effective connection with states that it is not (and cannot be) effectively coupled to
   c. So as to lead it to behave appropriately toward those remote or distal or other non-effective situations that it does care about, but cannot use. (Smith, forthcoming c, hereafter: RR.)

I will be surprised if anybody here has any serious quarrel with Smith's Representational Mandate (though it is easy enough to think of absent theorists who would squirm or rage). But some of us may be taking Smith's Mandate and interpreting it down, understanding it in a less radical way than he would wish. In the hope of giving his vision of it a proper outing, I will first try to give a summary of what strikes me as the dozen or so main points in Smith's work that bear on the issues of mental representation. (Much of the most interesting stuff in his book I'm going to set aside, reluctantly.)

2 A Dozen Important Points

1. Why re-tool our ontology? If we don't, if we complacently (or opportunistically) cling to the standard inventory, we will commit what Smith calls inscription errors or pre-emptive registration:[2] a tendency for a theorist or observer, first, to write or project or impose or inscribe a set of ontological assumptions onto a computational system (onto the system itself, onto the task domain, onto the relation between the two, and so forth), and then, second, to read those assumptions or their consequences back off the system, as if that constituted an independent empirical discovery or theoretical result. (Smith 1996: 50, hereafter: OO)

Pre-emptive registration is a sort of metaphysical anachronism, back-projecting onto our vision of ultimate-or at any rate more fundamental-reality a category or assumption that is in fact the effect or artefact of some later, higher-level, more 'expensive' development. [[Discussion point 5.1]]

2. The granddaddy case of pre-emptive registration is imagining we can parse the universe primordially into objects, which may or may not be appreciated in their object-hood by any (psychological) subjects in the neighborhood. By objects, Smith means what we (now) mean by objects-things that have spatio-temporal boundaries (at least roughly), that have careers, that can be reidentified, and that can, on occasion, be present to subjects-as objects to be perceived, sought, remembered, thought about, moved, destroyed, gathered, and so forth. As he puts it in RR, the world doesn't come 'pre-parsed' into objects, properties, relations, and other 'formal' categories.

3. The antidote to this form of pre-emptive registration is hard to swallow, but Smith gives us lots of help with various imagination-aids, temporary ploys, and other delicious candy-coatings. If I understand him right, it is actually strictly impossible to describe the primordial state without committing some sort of pre-emptive registration, since words-any words we can use-already bias us in favor of objectification of just the sorts he wants to describe the birth of. If I understand both Smith and Haugeland (unlikelihood squared) on this matter, they both think one can tiptoe past this problem (of the apparent inevitability of inscription errors in our attempts to do metaphysics). Here is how I put it in my review of Haugeland's book:

The task facing any 'Heideggerian/Kantian' theorist is to do justice to the role of us in constituting the denizens of 'our' world without lapsing into awful relativism/subjectivism on one side or caving in to noumena, or a 'God's eye view' on the other. Haugeland's solution, which grows on me, is to show how and why it is hard to 'constitute' a world (that takes care of anything-goes relativism) but not because there is a privileged way that the world-the real world-has always been constituted. His view is a close kin, I think, of my view of the evolution of colors: Before color vision evolved on this planet, sunsets and cliffs and volcanic eruptions had the reflective properties they did, but it makes no sense to ask if those sunsets were, say, red-since that question has no meaning independently of a reference class of normal observers. We can of course extrapolate back from our current vantage point and fix and answer such questions, using ourselves as the touchstone for colors, but we must recognize that we are doing that. [That is, as it were, acknowledging the pre-emptive registration that you're doing, and just keeping track of the fact that you're doing it. You're keeping yourself and your own categories somehow as a touchstone to talk about something to which they're not really directly appropriate.] Were there dinosaurs before H. sapiens came along and invented censoriousness and then ontology so that dinosaurs could be constituted? Of course there were, but don't make the mistake of thinking that this acknowledges a fact that is independent of H. sapiens. [[Dennett 1999: 433–4]]

I don't see that Smith's view of this is different, and that's fine, since I think this is a good and defensible view. Discussion point 5.2

4. With that apologia (or is it a caveat?) in place, I can now (pretend to) describe the primordial basis, the out-of-which that objects find their origin in. It is (very roughly) a Heraclitean world of flux, dynamically flowing and concentrating and dissolving. What is it composed of? Well, you really shouldn't ask, barefaced, since any answer will involve registration that is to some degree pre-emptive; but since we must advance the discussion, let's just speak of features. Don't worry; this is just a temporary stopgap: 'That the distinction between features and properties and objects is not sharp, on the other hand-that logic is messy, not just finger paints-will not ultimately be a problem, at least not for us' (OO: 127).

Features, I take it, are ways one region can be different from the neighboring region. Here 'neighboring' means, constitutively, in effective interaction with. One of my favorite dicta in a work filled with arresting phrases: 'Distance is what there is no action at' (OO: 200 n. 11).

5. This idea of locality underlies Smith's account of another kind of pre-emptive registration, highlighted in RR: the family of errors that occur when we persist in casting what really ought to be a theory of effective processes (or just effectiveness) as a theory of effective computation. The idea is that the truly important phenomenon of effectiveness is not a particularly computational phenomenon-it is a sort of historical accident that our first intellectual grip on effectiveness came via the work of Turing, Church, and their kind. Smith suggests that all kinds of mechanisms are effective without being computational in the ways that foster spurious connotations (of semantics, of proof in formal systems, etc.). This then seduces us into further pre-emptive registration and more inscription errors, taking 'logicist' baggage along for trips where it proves worse than useless. The central idea of effectiveness, Smith claims, has to do with local, non-distal causation.

6. The importance of 'flex and slop': Interactive effects dissipate, diminish with distance and time, due to what Hume once called 'a certain looseness' in the world (Hume 1739/1978: 408 (ii. iii. 2)). If the whole universe were like a gigantic interlocked gear-world, in which nothing could move without propagating effects ad infinitum, nothing could be out of touch with, or inaccessible to, anything else; nothing could be alone, or individual. Nothing could 'keep its distance' without flex and slop, which is a heretofore unremarked precondition for intentionality, because it creates the distance that then creates the problems that the varieties of reference-negotiation solve. [[Discussion point 5.3]]

7. Particularity is not individuality. The primordial physics world is everywhere particular, but contains no individuals (OO: 124–5). [[Discussion point 5.4]]

As I said at the outset, what appeals to me about Smith's project is that he's coming to this from a career in computer science, not from a career of teaching metaphysics. I'm trying to reconstruct the head-scratchings in computer science that make this seem so attractive, and it seems to me that they are something like the ultimate Y2K problem. The Y2K problem was not having enough bits for the year-settling for 2 when you should have 4, or, if you want to take a longer view, 5, or if you want to take Smith's view, many many many more. That is, when you start representing the world, if you're using any sort of data structure, you stop short with n fields, and n fields is in a certain sense never enough for a concrete thing, even something as simple as a cup.

The reason we make something a cup is that we have to create our little Y2K problem. There are only so many fields that we can carry along in our representation of the cup. We realize that if we want to keep track of that cup, there are all sorts of futures that we're going to have real trouble tracking if the cup gets smashed and then reconstituted, if it gets sold, if it gets repainted…There are so many different things that can happen to that cup. If we want to have a data structure that refers to that cup-come what may-it's going to have to have too many fields. We just can't do it. This is Haugeland's point, I think: a description of a person can't go into everything that's determinate about that person. ⟦See Discussion point 5.4—ed.⟧ It simply leaves out a lot of fields. There's a lot of bits that just aren't fixed and there's no room to fix them. ⟦Discussion point 5.5⟧

8. Chiming a point also made rather differently by Cummins (1996a), Smith offers several arguments to show that reference, and semantic relations in general, cannot be effective or causal relations. We can refer effortlessly to things outside our light cone, for instance, and the whole point of having something local by which you keep track of something distal is to overcome (without guarantee) the non-effectiveness of all such distal relations (OO: 157, 210–11, 228; see also the Representational Mandate 1c, above). And, like Cummins, Smith sees this as providing the elbow room for error (OO: 223).

9. The sort of 'non-effective tracking' exhibited by Smith's imaginary supersunflower is the forerunner of semantics, the basis of intentionality.[3] It is not what Smith calls registration, but it is the competence out of which registration can ultimately grow.

In all these situations, what starts out as effectively coupled is gradually pulled apart, but separated in such a way as to honor a non-effective long-distance coordination condition, leading eventually to effective reconnection or reconciliation. There is a great deal more to intentionality than that…but in various forms these notions of connection, gradual disconnection, maintenance of coordination while disconnected or separated and ultimate reconnection or reconciliation permeate all kinds of more sophisticated example. (OO: 206)

10. 'The retraction of responsibility onto the s-region [forerunner of the subject] is the origin of registration's asymmetry and directedness' (OO: 223). (I'll have more to say about this later, mostly critical, but reluctantly so, since I love the pedagogical uses to which he puts this mythic image of the s-region parting from its partner.) Smith's ulterior aim in this imaginative theme is to highlight the importance of the perspective shift he advocates in the next point.

11. It is the emergence of dynamically coordinated variation-systems (illustrated winningly by the intentional acrobat's flashlight, and the 'columnar'-shaped 'sustaining physical field' that unites the frog to the fly, OO: 217) that explain 'why we see trees, not electromagnetic radiation'. I think this point is strongly related to some of Ruth Millikan's (1984: ch. 15; 2000b: §§7.1-2) observations on identifying the reference/function of something by finding what holds constant across occasions, when we 'turn the knobs'. It is not just co-variance but systematic co-variance-which won't be perfect since systems are costly and may have weaknesses-that underlies our identification of objects of experience. ⟦Discussion point 5.6⟧

12. There are a variety of instances in which philosophers have traditionally dealt with dichotomies and Smith shows us how to see these as extremal points along some axis of variation. Thus the philosopher's ideal of a purely non-deictic registration is a myth (OO: 249). We have cases that are halfway between implicit and explicit, halfway between 'pure' reference and intension (e.g. OO: 251), and so forth. These middle-ground cases are very important in Smith's larger scheme of things, since as he eloquently says (OO: 254–5), the main lesson to be learned is 'not to be seduced by limit cases'. (See also the end of ch. 8 (and of the book): Life-what matters-happens in the middle ground.)

These strike me as the main things I have learned from Smith's book and RR. Let me add what I take to be the main point of 'One Hundred Billion Lines of C++' (1997), to make a baker's dozen:

13. The productivity, the compositionality of programming languages (such as C++) should not lead us to suppose that in general the processes such programming languages permit us to design and implement are similarly compositional. The fact that the programmer can create indefinitely many identifiers (and indeed can create nonce-systems of compositionality on the fly as he goes) does not at all imply that the identifiers thereby created can be treated as manipulable, composition-friendly items by the program itself. The compositionality is in the syntax and semantics of the source code but not in the structures that then get built and then actually get implemented and then run.

This insight, restored to philosophy, shows not that Fodor's language of thought is not the way we work, but does show that there is nothing remotely like a plausible inference to the conclusion that there is a language of thought from the premise that the brain engages in computational-like processes whereby it extracts apt behavior from the information it extracts from the world.

Now I take the upshot of all this to be a multi-path attack on the 'classical' ideal of mental representations as well modeled by 'propositional' symbol systems that obtain their intentionality by composing something like Fregean Thoughts out of Terms with Extensions and Intensions. Every tractable theory has lots of idealizations and simplifications, but the idealizations of that family of theories are trouble-makers, not helpers-largely because of preemptive registration: they create the illusion of sharp distinctions where in reality there is something of a spectrum, from 'non-effective tracking' to the most intellectual of opinions (e.g. my opinion that the shortest spy is a spy). What Smith calls registrations—occupants of the right-hand region of this spectrum, you might say—only work in contexts of 'coordination conditions', in adjustment or compensation (what Smith calls 'intentional dynamics', OO: 262), processes that philosophers have tended to overlook or underestimate the importance of. Smith puts to good use one of my own images to skewer the false view: the classical system of uninterpreted symbols is seen as wearing a thin 'overcoat' of transducers and effectors as the interface between symbol and world. [[Discussion point 5.7]]

Now I want to offer what I take to be a friendly amendment, but I expect Smith will view it askance. If my expectation is mistaken, hurrah; if it's right, there is no question I am more than eager to explore than why he resists this (to me) obvious improvement.

3 The Origin of Objects?

For me, the ghost at Smith's banquet is—surprise, surprise—Charles Darwin. Evolution is hardly mentioned in his book, whose very title trumpets its likely affinity to Darwin's great vision. How can we have a story—a Just So story, in fact, eloquently brandishing its own unavoidable metaphors and anachronisms—of the origin of objects, of their emerging onto the contemporary landscape from some primordial scene in which they were absent-and not have it rely on the fundamental Darwinian principle of natural selection? What alternative shaping forces could do the work that needs to be done? Smith does a wonderful job of showing us the 'expensiveness' of objects and subjects; something has to pay for all this R&D!

Let me draw your attention to a few crucial points in his account where I, Darwinian Fundamentalist that I am, feel an irresistible urge to insert evolutionary considerations. Look again at the Representational Mandate: a representational system is 'normatively directed'(1b); it 'exploits local, effective properties' (3a) 'so as to lead it to behave appropriately' (3c). Smith's examples-the supersunflower, the frog, and (most important) the unnamed simpler organisms who pioneer the passage from proximal irritation to distal 'non-effective tracking'—all bespeak his interest in evolution, in simple minds and their successors, but he strangely eschews the evolutionary perspective. Why? Because, I think, he wants to avoid what he takes to be the pre-emption error of what we might call premature teleology, or premature function. But he overdoes it, methinks. He wants to introduce normativity in his way, not riding on the coat-tails of evolutionary normativity. But I think this is a mistake, too. All normativity does ride on Darwin's coat-tails. In trying not to be 'expensive' Smith goes too far here. Consider, for instance, his excellent summary (OO: 241):

> "The underlying spatio-temporal extended fields of particularity throw tufts of effective activity up against each other, and let them fall apart, fuse them and splinter them and push them through each other, and generally bash them around, in ways governed by the pervasive underlying

> (physical) laws of deictic coupling. [So far, no hint of teleology; this is all just Heraclitean flux, signifying nothing.] For a subject to begin to register an object as an individual is, first, for a region of the fields (the s-region) not to be connected to another region (the o- region), but in the appropriate way [[my italics]] to let go of it.…The coordination requires establishing appropriately [my italics] stable (extended in the s- region) and abstract (extended in the o-region) focus on the o-region, while remaining separate. The separation helps in maintaining the s-region from being buffeted by every nuance and vibration suffered by the o-region."

Notice how we end with pure engineering: protection of the s-region from buffeting, in order to maintain a 'focus'-on an appropriate object. The fact is that s-regions that happened to begin to register inappropriate o-regions (don't-cares) or to register suitable o-regions inappropriately (inefficiently, counter-productively, etc.) would not last long in the buffeting flux, not long enough to out-reproduce the competition in any case. Once we add this evolutionary point, we can emend Smith's account, adding what strikes me as its most important theorem: the world doesn't come 'pre-parsed' into objects and properties (just as Smith says) but objectification is what I call an evolutionary Good Trick (Dennett 1995a, hereafter: DDI), an elegant solution to the problem of staying alive in the world of flux, flex, and slop, a solution we would expect to find, for instance, in other galaxies in which life had evolved.

Will Smith want to go that far with me? I hope so. His pluralism is sane and temperate. By taking pluralism (and postmodernism more generally) seriously (and not just pre-emptively dismissing it with a sneer, as it is extremely tempting to do) he allows it to tame itself. Yes, there are real problems of pluralism, and yes, there is no guarantee at all of a single, pre-given ontology to which we can anchor all reference, but reference-preservation, or reference-negotiation, is a problem that we can solve, and routinely do solve. (Don't patronize the Others. You can be brought to understand their ontology and they can understand yours, with a little effort.) There is a Good Trick (maybe two or three, but we know of one for sure that works well) that has been discovered again and again by evolution, and Smith has a deeply insightful account of how it works to generate our ontology.

I think this evolutionary perspective on the birth of intentionality is preferable to the charming myth that Smith puts in its place: 'In all these situations, what starts out as effectively coupled is gradually pulled apart, but separated in such a way as to honor a non-effective long-distance coordination condition, leading eventually to effective reconnection or reconciliation' (OO: 206), which ignores the fundamental evolutionary facts: we only 'want' to be coordinated to the things that matter to us, and these are not necessarily things we used to be attached to. The food I hope to coordinate with has never been within hailing distance of me till now, but I pounce on it just the same. I love Smith's imagery—especially his Country and Western song sound bite: 'How can I miss you if you won't go away?'—but it reminds me, I fear, of another cool idea (Freud had a lot of fun with it) that we evolutionists have shown how to replace: what we might call the Siamese-twin theory of sexuality, which imagines a primordial time when male and female were happily united, later cruelly sundered, and spending the rest of eternity as 'halves' trying to reunite. The evolution of sexuality is a deep and fascinating problem, since it, too, is expensive and needs to be paid for, but we don't solve the problem by imagining that an m-region and an f-region gradually got pulled apart and are striving to reunite.

Smith says at one point: 'Third, the retraction of responsibility onto the s-region is the origin of registration's asymmetry and directedness' (OO: 223). This serves to balance his various claims about the shared roles of subject and object. As he says, the dance has two partners but is not symmetrical. By leaving out evolution, however, he leaves out what I take to be the deeper reason for the asymmetry. The sun doesn't give a damn about the sunflower, but the sunflower needs the sun. You need something more like predator-prey (or mate) asymmetries to make sense of the asymmetry of registration.

I think Smith ought to accept all of this, and in some passages he sounds just the right notes. For instance, he notes that 'a distinction takes hold between what the s-region is doing (tracking the coyote or incident sunlight) and how it is doing it. The former gets at a non-effective regularity; the latter, at an effective mechanism whose "job" is to implement or sustain it. Among other things, this split provides a toehold for normativity to attach its tentacles' (OO: 222). Exactly: An evolutionary toehold for normativity.[4]

What might be fueling his resistance, then? In his account of what he calls 'intentional dynamics', his name for the theoretical basis of situated cognition, he tells us he wants to keep the normative at bay (OO: 262): he doesn't want to build the normative condition into the name (by calling intentional dynamics something like 'rationality' or 'reason'). Fair enough; we need to understand the underlying physics, if you like, that any representational scheme, good, bad, or indifferent, must cope with, so we must be careful to describe not just the (presumably) optimal mechanisms, but also the junk that might be lying around interfering. Bad engineering and good engineering live in the same world, and that world should be clearly described without the bias of preemptive registration, if possible. I also think he wants to avoid what might be called 'premature agency' a sort of inscription error in which one breaks the world up into things doing things to things, as if this were the primordial catalog. See, for instance, his nice image of getting rid of the potter, OO: 270. But in the end, I gather—mainly from the strong claims in RR (p. 29) about a distinction between static and dynamic norms—that Smith's reasons for resisting an evolutionary treatment of representation come from…Pittsburgh. The 'dynamic norms' claims ring a Haugelandian, Brandomian, McDowellian bell for me, but I don't buy it. Not yet. I think I'll stand firm and ask to be shown what's wrong with my Darwinian fundamentalism, whose motto is All normativity is grounded in evolution and emerges from the cascade of Darwinian algorithms.

4 Coda: Three Reservations

1. *Indeterminacy of Content.* I see a tension between 'There may not be any compelling reason to believe there is even a metaphysical fact of the matter' (OO: 55) on the one hand and, 'We may not know what it is, but that does not mean God leaves the content indeterminate' (OO: 62) and, on the same topic: 'Somehow or other-and this I take to be the most important and difficult task facing the cognitive sciences—it must be possible to have determinate representational content, that is, for there to be a fact of the matter as to how the world is represented' (OO: 68). I ask, 'Why?' Smith says 'it will have to be an answer that does not depend on how anyone registers or individuates those mechanisms-again, for the simple reason that it happens in people, for example, without anyone doing that.' I don't see that as a good reason. This is like Cummins' similarly staunch line on determinacy of content, and I am not yet persuaded. Why can't God leave content indeterminate?

In this tug of war, I tug on the former side, of course. It helps us escape what might be called Cartesian (or 'from the inside') ontology, the view Quine calls the 'museum myth of meaning'. We must not assume that there will be an 'inner' perspective from which semantic facts of the matter can be mined. (See also Ruth Millikan's (1984) critique of 'meaning rationalism'.)

I don't see what's wrong with (my) perspectivalism about this. After all, it is flat true of some computer applications that they can be adopted wholesale for use in another domain (the old chestnut of the chess machine that can play war games, or whatever). See 'The Abilities of Men and Machines' (Dennett 1978b) for an ur-example. Why should it be different when we then look at animals, say? What if the fly-detector machinery is reused (exapted) intact in some later beast? I think Smith is right (and it's a good point) that the semantic/syntactic distinction is not the external/ internal distinction, but I don't think that this further point follows. [[Discussion point 5.8]]

2. *What about Animals?* In spite of all the good discussion about frogs (OO: 197, 216–18, and other places) and coyotes, we are left wondering: do clams register? do amoebae? do they objectify? (see OO: 149, 193, 232). Smith (OO: 195) says that larger corporations and communities may be implicated in intentional achievements, but he downplays the role of proper parts of organisms. Why? Smith's bias in favor of human beings is largely uncharted (see my 1999 review of Haugeland on the same topic). Yes, only whole human beings living in whole societies, with slathers of normativity laid on, ever really refer to anything, but then there is lots of quasi-reference. And Smith is the master of pointing to just these facts. I wish he'd said more about whether dogs reidentify individuals, for instance (a question I take up, and don't answer properly, in *Kinds of Minds* (1996: 113–16), where I explore the case of Ulysses' dog Argos, who seems to recognize him when he returns. Does he? Really? ⟦See also Discussion point 1.4—ed.⟧). In Smith's brief remarks on ethics (RR: 31), there is a clear link to my concern with Smith's silence on evolution. What if there were no people, only animals? There would be no ethics, I gather, but wouldn't there be lots of mattering? There would be lots of survival and extinction for cause, lots of biological norms.[5]

3. *C++ and Searle on Programs.* First, I give Smith's essay an A++, and express my main objection: he should have written it twenty years ago and saved us all from a series of dubious battles that have gained precious few insights as by-products. But I also want to add to his concluding point 3 on Searle, about which a bit of clarification is in order: 'Searle's analogy of the mind to a program is misleading. What is analogous to mind, if anything, is process.' Smith adds: 'it is unimaginable that evolution constructed us by writing a program, a syntactic, static entity, which specifies, out of a vast combinatoric realm of possibilities, the one particular architecture that the mind in fact instantiates'.

In his uncharacteristically ill-considered *Daedalus* article on AI, Hilary Putnam (1988) speaks of the Master Program—which is perhaps the closest anybody has ever got to imputing this view to AI or to anybody. (See my critique in the same issue of *Daedalus*, reprinted in *Brainchildren*, 1998.) Smith is right in what he says, but let's see what this leaves available: the mind is, as Smith says, process (or a bunch of processes conspiring together), and while there is—need be—no programming language that specifies that family of processes in nature, that plays the causal role played by the source code in the genesis of new processes inside computers, those processes may nevertheless be usefully specified as if they were implemented programs. That is, to take the Searle case very much to heart, Searle has claimed that whatever consciousness is, it is not like a program in this sense: take a brain that is unconscious, and make it conscious by installing/implementing that program on it.

Now I continue to believe with all my heart and soul that this is exactly what consciousness is! Consciousness is a set of behavioral competences that depend not so much on the organicity of the brain's neurons as on their global behavioral roles, so that you could in principle have live, healthy neurons by the billions subserving no consciousness at all—a comatose or otherwise utterly demented person—and you could turn that brain into the brain of a conscious person by 'simply' revising the behavioral micro-dispositions of those neurons, turning them into organelles and tissues that accomplished various 'computational' and 'communicative' tasks. In fact, when people recover from strokes, the resumption of various parts of normal conscious competence is very much a matter of the reutilization of healthy neurons to play new computational roles.

Moreover, of course, I've argued (and here is where my view is most radical, most embattled) that there is something that plays a causal role similar to that of source code in the genesis of much of this behavioral microcompetence: there are virtual machines that are installed by cultural imposition, learning, imitation, and memetic infestation, and whatever it is that hops from brain to brain is, in some no doubt hugely indirect way, a specification of a set of habits of thought. A bit like Java applets. Thus, you encounter Tetris, and find yourself executing shadow Tetris-moves involuntarily for some minutes or hours. Or you learn bridge, and find yourself putting yourself to sleep doing shadow-finesses, or you learn about agreement of adjective and noun in Italian and execute hundreds of agreement-checks…until it becomes second nature. The culture has driven a little rule into implementation in your head, and it is the same rule that all Italian-speakers have somehow or other implemented in their heads. Perhaps, to continue the analogy up to if not beyond the breaking point: native Italian speakers have the rule compiled in their heads, a much 'sleeker, more efficient machine' (as Smith says) than the interpreted version that still occasionally rises to the level of consciousness in my own operating system.

My point here is that nothing Smith has said about the non-compositionality of most executable programs casts doubt on the utility of such treatments. On the contrary, it helps mightily to clarify them, and to ward off likely misinterpretations.

6b — Reply to Dennett

Brian Cantwell Smith
University of Toronto

Let me start by thanking Dennett for two things.

First, I'm grateful for the effort he has put into understanding this project—a project, I admit, that can seem a little like a fire hydrant: the content comes out in lots of different sprays. I learned from his comments, and that's great.

Second, I want to thank him for mentioning the issue he identified as number 12: domestication of the 'middle ground' opened up by all sorts of traditionally dichotomous theoretical distinctions. That focus on the textured intermediate territory, rather than on limit cases, is very important to me. I think of it as the philosopher's analog of in vivo rather than in vitro analysis. In my experience, people who don't appreciate the importance of this kind of middle-ground stance find it hard to hang on to, especially at first. It is a well-entrenched intellectual habit (especially in analytic philosophy) to think that theoretical rigor demands 'clear and distinct' ideas, even clear and distinct cases. But just as there are dangers of drowning in complexity and detail, so too there are dangers of excessive (especially formal) abstraction, particularly for subject matters—of which I think epistemology and ontology are instances—whose stuff and substance only emerges in these often messy middle regions. Doing such phenomena justice requires a distinctive theoretical style. Although hard to get, initially, this middle-ground approach is also hard to lose once you've got it-particularly when you see its not being appreciated all over the place. So I thank Dennett for noting that right up front.

Needless to say, I can't respond to everything he has brought up. Instead of giving a point-by-point response, I want to make six general remarks bearing on the issues he has raised. In conjunction with his comments, I hope these will clarify what is going on.

1 Naturalizing Ontology

The first remark has to do with the project of naturalizing ontology. 'Why bother?' asks Dennett. The main reason, of course, is because I believe the subject matter demands it. What ends up as a methodological commitment is grounded in an empirical claim: that the theory of ontology and the theory of representation and intentionality are about intrinsically interconnected phenomena. To study one without studying the other would be like studying time without studying space. Time is not space, of course; no one thinks they are identical. But you would not get an adequate account of either space or time by studying it on its own. So too, I believe, for representation and ontology. How things are and how we take them to be, though by no means identical, are co-constituted in intricate ways.

I might say that I haven't always believed this.[6] During the 1980s I spent a long time trying to develop a theory of representation independent of ontology.[7] I was particularly interested in taxonomies of representational types (symbols, icons, descriptions, models, simulations, etc.)—a theory, I might say, in which isomorphisms figured.[8] Now I didn't have the smarts to invent targets to do the work that representations couldn't do. But my fundamental problem was that I couldn't hold the ontology fixed—couldn't stabilize it adequately—in order to develop satisfying accounts of the plethora of correspondences that held between them. I was unable to determine (except by fiat, which didn't satisfy me) which items were objects or basic elements, which were properties of those elements,[9] and which were relations among them. Small variations in how I registered the basic domains wreaked havoc with how I ended up classifying the representations defined over those domains. In the end I was forced to admit that the (ontological) question of whether something was an object could not be answered except with reference to the (epistemological) question of whether it was being objectified by a representing or cognizing subject. That is: my independence assumption did not work. So there is a lot of failure behind this claim that representation and ontology are parts of the same subject matter. That really is the bottom line.

So I started over, to reconstruct ontology and representation together. It is not just an exercise, at the end of which you end up with the same recognizable parts. The theory that comes out—the benefits it gives you—are different.

Perhaps the simplest benefit is that it gives you more resources to describe intermediate cases. The notion of feature placing, for example, turns out to be extremely broad and useful-and relevant, I think, to the issue Dennett raised about animals. The basic idea of a feature, which I take from Strawson (1959), is of something logically simpler than a property. Like properties, (concrete[10]) features are spatio-temporally instantiated, but, unlike properties, they do not involve a commitment to a discrete, individual, reidentifiable object, complete with unity or identity or individuation criteria, to serve as the exemplar or 'holder' of the property or feature or abstract type. Paradigmatic commonplace features are fog and other meteorological phenomena. The truth of an utterance of 'it is raining' requires only that there be raining going on 'around here, about now', as is sometimes said. There is no object to which the term 'it' refers.

Take another example. Suppose Dennett visits my California house, and on the second day remarks that the fog's come back. 'You are a philosopher,' I ask, 'has the same individual fog returned, or is it new fog, of the same type?' I don't know what Dennett's answer would be; but I know what it should be: 'Go away!' Similarly, suppose you and I go camping,[11] and you, getting up early and looking around, stick your head back in the tent and say 'It's amazing; we're camped right next to a whole ridge of mountains!' Again, suppose I pedantically inquire, 'You've used the plural "mountains"; just how many mountains are there?' There is no reason to expect that an answer is possible. The problem is not epistemological: that you don't know, that you can't count; that you can't see. Suppose the air is crisp, the view clear, and that we have all the time in the world. It doesn't help. The point is that there is no metaphysical warrant, at least no metaphysical warrant up there on the ridge, for one answer over another. Criteria for mountain individuation simply don't apply to such situations with anything like exact enough grip.[12] Similarly for a host of other examples. The point is simple: pre-theoretic philosophical intuition notwithstanding, much commonplace registration of the world does not require parsing it into discrete individuals.

Philosophers are a rarefied class; many of us, at least since our first course in model theory, have been persuaded that we do take the world in terms of discrete, reidentifiable objects exemplifying properties and standing in relations. Or anyway that that is the right idealization under which to pursue philosophical topics. I myself suffered under this misconception for many years. But I no longer believe it. (This is another of those things that are hard won. At first it is difficult to credit, but then, once you come to see that it is true, it is hard to imagine how you ever believed the traditional story: that it is a precondition for finding the world intelligible that you first parse it into discrete individuals.)

Feature placing is just a stepping-stone, of course. Adopting a richer ontological framework doesn't require the stronger thesis, that ontological facts are in part intentionally constituted. But examples of feature placing are useful because they suggest why that stronger claim is true. If pressed to supply answers to individuation questions in such cases (for example, to decide how many mountains there are), you will notice that the only way to do it is to make recourse not simply to the structure of the world (the details of the shape of the ridge), but also to the demands and contingencies of the projects you are engaged in. If we were committed to climbing all the mountains on the ridge, for example, that might affect our answers as to what distinguishes 'one mountain' from 'two.' If we were geologists, our answer might be different. Likewise, airplane pilots might arrive at judgment different again.

One common way to handle such variation in individuation practice is to claim that the word 'mountain' is ambiguous; that climbers use one sense, geologists another, pilots a third, and so on. But this strategy doesn't work. Senses multiply too profusely-varying per speaker, per occasion, per project. Eventually one is forced to admit that sense is indefinitely variable, and subject to factors anchored in the intentional projects of speakers. But this is an expensive admission: it reduces the 'multiple sense' proposal to no more than a relabeling of the original problem.[13]

In the long run, I believe, there is no credible alternative except to recognize that intentionality is implicated in individuation. Let me put it as succinctly as I can:

> *The identification and reidentification of objects involves an epistemic process of abstraction over the infinitely rich (and often surpassingly messy) ur-structure of the world. Among other things, the normative character of the intentional projects that agents are engaged in, when they commit these acts of abstraction, figures in the resulting 'clumping' of the world's effectively infinite detail. To be an object is to be a region or patch of the world that is successfully abstracted-where the issue of 'success' is tied into the normative conditions governing the dynamic project of which the act of abstraction is a constitutive part.14 The fundamental character of (what it is to be) an object is thus intrinsically hooked into the intentional life practices of the objectifying subject.*

One more point on this topic. As a way to muster support for simply availing ourselves of 'common-sense ontology', Dennett says 'Look, why not just assume sub-atomic particles and tables and mountains and galaxies, in the way that science does?' This leads me to mention a radical thesis that I hold, although I can't give it much defense here: namely, that science may not be committed to objects at all. Consider: an amoeba splits. Biology doesn't care about the individuals in the situation: whether one amoeba died and two new ones were born; or whether we now have a spatial distribution of unitary amoeba-ness; or whether one of the two emerging amoebae is the original one, and the other one is new; or any other possibility. Another example: in California I own an ancient redwood tree that has clumps of very substantial shoots (some as much as 50 feet high) sprouting around its base. How many redwood trees are there? Science doesn't know, and science doesn't care. Similar conclusions hold for fog, for the units of selection, for a myriad other examples. What this leads me to believe is that scientific laws (like animals) may in fact deal only in features;[15] and that the objects we think of as constitutive of science may merely be simplifying epistemic devices that allow humans to calculate.[16] Objects in science, that is, are in my view properly understood as part of the epistemic apparatus involved in the conduct of science as an intellectual activity (on a par with mathematical models); they are not ontological commitments of the theory as a whole. [[Discussion point

5.9]]

Put it this way: ontology and abstraction need naturalizing as much as meaning, semantics, and content. Assuming a 'standard ontological inventory' for purposes of giving a naturalistic account of intentionality, as Dennett suggests, is thus a doomed project: it is viciously circular. Think about how appalled we would be (or anyway naturalistically unsatisfied) if someone were to propose a theory of representation that dined out on intentional notions, as if they were freely available. The naturalistic challenge is to explain intentionality without viciously presuming intentionality. A similar moral holds for ontology, in my view. Because ontological categories are in part intentionally constituted, attempting to explain representation while dining out on ontology is, for analogous reasons, fatally circular.[17]

2 From E&M to M&E

Second, I wanted to make a remark about the role of physicality in the metaphysical project.

I remember talking to Fodor once,[18] trying to convey my amazement that reference could point outside a speaker's light cone. His response stunned me. 'Look,' he said; 'it doesn't matter what physics is like. Physics could be arbitrarily different, and it wouldn't have a shred of impact on the theory of intentionality.'[19] It is hard for me to say how deeply I dis agree with this sentiment. There is a sense in which I am something of a physicalist.[20] Not, mind you, a reductive physicalist-but someone who takes the character of the physical world to be essential in determining what intentionality is like. As a result, I take the consequences for a theory of intentionality of the structure of the concrete, material world to be enormous (as, I might add, must anyone who takes material embodiment seriously). The trick is to spell this out in a non-reductive way. Note that the issue is not merely one of engineering: that intentional subjects be physically implementable. The connection is much stronger than that. As I tried to show in OO, the structure of the physical world actually establishes the problem that intentionality solves (as well, fortunately, as supplying the wherewithal for its solution). [[Discussion point 5.10]]

For various pedagogical reasons, I take field-theoretic interpretations of physics especially seriously. I'm a complete amateur at physics (as my readers will know), but for purposes of understanding intentionality, field-theoretic interpretations have a decisive advantage. They make it evident that physics does not involve a metaphysical commitment to discrete fundamental individuals. [[Discussion point 5.11]]

To see this, assume a field-theoretic interpretation of classical, high-school (Newtonian-Maxwellian) physics: spatio-temporally extensive fields of force, mass, charge, etc., subject to various dynamical regularities. And consider what is involved when we talk about individual bodies, as for example we might if we were to ask about the gravitational force exerted by this cup on this pen. In calculating the answer, we might be tempted to characterize the problem as a mass of 200 grams and a mass of 30 grams standing one meter apart. My point is simply that, as everyone knows, this characterization involves some simplifying idealizations. It makes two acts of abstraction over the raw fields: one to collect up the region of space-time we call 'the cup' into a dimensionless unity; the other to collect up another region, which we call 'the pen', into a similar dimensionless unity. That is, we objectify both cup and pen: treat them as discrete, individual, infinitely dense space-time points.

Why do we do this? For a very good reason: the simplifications are necessary in order to yield a problem that is epistemically tractable. Staying true to the field-theoretic interpretation would require treating the cup as a full three-dimensional mass density manifold, the pen as another three-dimensional mass density manifold, and formulating the question as one about the gravitational attraction between two solid regions. Setting up the problem in this way (that is, without any abstracting simplifications) would require an infinite amount of information. And solving the resulting problem (a double triple integral of point-wise gravitational attraction between two regions) would require an infinite amount of work. Neither, in general, will be feasible.

In sum: working with solid 3d regions, which is all that physics is really ontologically committed to, yields epistemically intractable problems. So we simplify, for purposes of calculation. That is where individual objects enter.

Once the distraction of individuals has been set aside, one can see that the features of the physical world that most affect the nature of intentionality have to do with distance, coupling, and the locality of physical force (this has already come up in discussion, and I will say more about it in a moment). In particular, the 'point' of intentionality and reference, on my view, is to allow agents to be directed toward (ultimately, to care about) the world as a whole, beyond the (causal) limitations of that envelope with which, at any given moment, they are causally engaged. There are additional detailed connections as well. I mentioned an important one yesterday:[21] the differential character of physical regularities[22] engenders a kind of in-the-world deixis, which engenders an ineliminable indexicality in all representation and reference, which in turn underwrites the first-person qualitative character of phenomenological experience.

These are just a few examples of how I mean to take the physical world seriously.

3 Effectiveness

The third remark I want to make has to do with the relation of semantics to these issues of causation and local effectiveness. I subsume this under what I view as the problem of physical or material embodiment:

> How can small patches of the physical world (for example, us) exploit a small fraction of the sum total of ways of being that the world supports (namely, that fraction that is causally potent or causally effective), so as to allow them to register the whole world (not just the part they are in or constituted of) as exemplifying an almost limitless variety of properties?
>
> How, in other words, do we exploit a small fraction of the properties of a small part of the world to gain access to all properties of the whole world? This, I would argue, is the problem to which reference and semantics are the solution.

I hope this formulation clarifies my disagreement with Fodor. For if my characterization is right, then understanding the character of what I am calling the 'locally effective' (that is, those properties of local situations that can do causal, effective work) is essential to the project of understanding intentionality. And this for two reasons. First, the effective properties are what an agent gets to use; they are the 'material', as it were, from which an agent can construct its intentional solution. Understanding them is thus necessary in order to understand how intentionality 'works'.[23] Second, we need to understand what these properties are not-that is, the vastly larger fraction (99%) of the world's features and properties that aren't effective, or don't hold of the local situation, and hence that the agent can't be coupled to 'directly', by physical coupling-since that is what constitutes the 'rest of the world' toward which the agent is intentionally directed.

One question that inevitably comes up, when I put things this way, has to do with the relevance of quantum mechanics. If intentionality is intrinsically related to physicality, then is the character of the intentional affected by the fact that the physical world is not ultimately classical? The answer may be 'yes', though I confess I am not prepared to say very much about this yet. To date, I have constrained my study of the locally effective to phenomena that, as far as I can see, could supervene on a classical base.[24] I've done this in part because I have yet to see any compelling argument that the human brain reaches further.[25] Even though I take my subject matter to be intentionality full bore—that is, intentionality in any possible material manifestation, not just its human projection—nevertheless, the sheer magnitude of human accomplishment convinces me that basing such an account on classical physics is not too severe a constraint.[26]

In the long run, though, I admit that the study should probably expand to include quantum efficacy. But the nature of quantum influence may be quite subtle. For example, one place where quantum mechanics may bear on the nature of human experience, at least indirectly, is in issues of long-distance coordination, of the sort that violate traditional locality constraints (for example, as characterized in Bell's theorem). Note that the fact that we can register the world, see things, think, remember the location of Dennett's house in Blue Hill, and so forth, is because there is a tremendous amount of long-distance relatedness in the world. Maintaining a (moderately stable) conception of the (moderately stable) world depends on this (moderately stable) relational regularity. Is quantum non-locality a necessary precondition for such long-distance regularity? In informal conversations, some physicists have suggested that the answer may be 'yes'. If that is so, then that is surely one way in which the human condition may be crucially non-classical.

Other than speculative questions of this abstract sort, however, I doubt that quantum mechanics has much to say about our middle-scale intentional lives. So I lack sympathy for writers (such as Penrose) who feel that in order to penetrate the mysteries of consciousness we need to understand mind in quantum-mechanical terms. In 'Who's on Third?' (forthcoming a) I argue to almost exactly the opposite conclusion: that (again) using no more than a field-theoretic interpretation of high-school classical physics, one can see how the first-person, subjective, qualitative character of phenomenal consciousness must arise in any physically embodied agent that achieves an objective conception of the world around it. [[Discussion point 5.12]]

But return to the issue of simple effectiveness: how an agent can exploit what is effectively available to stand in for, care about, and otherwise direct it toward, that which is unavailable. This, in my view, is the best way to frame the question of intentionality. As you will predict, I ultimately locate the syntax/semantics distinction as a special case of this more general issue. But for pedagogical purposes, logic is not always the most illuminating place to start, in order to avoid being distracted with inessential aspects of the formalism.

In my undergraduate teaching, I start by studying clocks. I choose clocks for several reasons: (1) because they are familiar, (2) because they have a clear mechanism (clockworks), (3) because clock faces raise issues of interpretation and content, and (4) because clocks are so manifestly dynamic. Suppose we want tea at 4 o'clock. I assume that '4 o'clock' is a non-effective property exemplified by passing metaphysical moments (one every 12 or 24 hours). If '4 o'clock' were effective, it would be simple to build a tea-making device: you would construct a detector to respond to a moment's exemplification of that property, and connect it to a switch. When 4 o'clock arrived, the detector would respond, the switch would flip, the kettle would boil, and out would come tea (or whatever). But of course—to make a point so obvious that we typically don't realize how crucial it is—*you cannot get a meta-*

physical moment's exemplification of the property of being 4 o'clock to turn a switch. You can't do that because 'being 4 o'clock', as I keep saying, is a non-effective property. So what do you do instead? You construct a mechanism that uses properties that are effective, out of stuff you don't otherwise care about, and arrange it to be coordinated with the property that isn't effective that you do care about (a moment's being 4 o'clock). If the coordination is established properly, the former effective mechanism can stand in for the latter non-effective goal.

As Cummins just said [[in Discussion point 5.12]], one metaphorical way to understand this is to realize that intentional creatures have just a tiny keyhole through which to access the world. How far does their effectiveness reach? Because of proscriptions of locality, it reaches only to the surface of their skin. Strictly speaking, what is absolutely proximal—what impinges on your surface—is all you have to interact with. This is true of any conceivable physical agent: the infamous locality of physics restricts all engagement with the world to coupling with what is immediately present. Here we are, at this very moment, sitting in the living-room of an inn; our coupling to Dennett's farmhouse, even though it is only a few miles away, is at the moment very weak. So weak as to be 'undetectable'. When we want to go to Dennett's place for lunch, we cannot be driven by effective coupling to it (as a Gibsonian might imagine we are directed to an opening in a wall by effective coupling). So what do we do instead? We arrange the situation so that we can be driven by things that are effective, in the here and now (such as maps), that will enable us to get us to his house, there and then. The dance that this strategy engenders-of exploiting what is local and effective in order to be directed toward what is non-local or non-effective-this is the phenomenon I am talking about. [[Discussion point 5.13]]

4 Computation

I trust that it is clear how this effective/non-effective dance relates to issues of mind, reasoning, and logic. My fourth remark has to do with its relation to computing. To explain this, we need to look at the history of computer science.

For almost a century people have been developing a so-called 'mathematical theory of effective computability'—or, as it is often simply called, the 'theory of computation'. In spite of its name, however, I do not think it is a theory of computing, because it doesn't deal with computing's essentially intentional character. Nevertheless, I still consider it an amazing achievement. What it is, I believe, is a mathematical theory of causality-that is, a theory of exactly what I have been talking about: physical effectiveness. This theory will not capture everything that matters about our pre-theoretic intuitions about causality, such as how you can cause things to happen after you've died. But that's fine; scientific theories never exhaust the pre-theoretic intuitions on which they are founded.[27] What this theory does capture, in the long run, I believe, is what we will end up taking physical effectiveness to be.

So the 'theory of computing' supplies half the intentional story: the effective half—the part about what you can use, what you can do, what works, how hard it is to change one physical arrangement into another.

A brief historical caricature may help explain why things developed this way. At the end of the nineteenth and beginning of the twentieth century, following the impressive achievements of the Industrial Revolution, there was a tremendous sense of the power of machines. Some of these machines were bluntly physical (steam engines). Some were targeted at very specific material concerns (smelting iron ore). Clearly, however, some very useful mechanisms, such as clockworks, weren't so concretely specified. Although it was crucial that they be physically constructed, it didn't matter what specific materials they were made of.

People realized, from examples of this sort, that if you want to know 'what can be done', you can abstract from purely physical considerations-how big the mechanism is, how much energy it uses, etc. Paradigmatically, such mechanisms are used for detection, tracking, and other (at least inchoately) intentional tasks. Suppose you want to know when some particular train passes a spot on a railroad. You might put a sign or indicator on the train, and install a detector next to the track, to signal when the train passes. Sign and detector will obviously have to match, in physical characteristics, so that the latter can respond to the presence of the former. But beyond this, there are no requirements on what they should be made of, how big they have to be, etc. Because, of course, all we really need is to detect one bit of information: 'the train is here'.

Many such mechanisms can be imagined, of an essentially physical yet 'multiply realizable' sort, from simple detectors, to clocks, to what has seemed like the most powerful mechanism of all: one that could calculate, reason, do mathematics or logic. What people realized (to continue this glib story) is that, to get a theory of such devices, you have to let go of specifically concrete concerns. So what did they do? They went to the other extreme, and considered devices as completely abstract. Since what could be done (for example, by Turing machines) seemed not to have to do with specifics about particular materials, the theory took the opposite pole, and assumed that what could be done had nothing to do with materiality at all. This is why the theory of effective computability is framed as if computability were a purely abstract notion.

We are still living in the shadow of this history. The idea that the fundamental results of computability theory might be anything other than completely abstract is far from universally acknowledged. Many academics treat theoretical computer science as a branch of logic or mathematics.[28] And challenging this assumption bends some people out of shape.[29] Nevertheless, I believe, helped by people studying the powers and limitations of quantum computing, it is going to become increasingly apparent that computability limits are fundamentally material.

In terms of long-range intellectual trends, in other words, we moved from the completely concrete steam engines of the nineteenth century to the completely abstract inaccessible ordinals of the early twentieth century. Now, on the verge of the twenty-first century, we are settling somewhere in the middle. But through it all, the study has been a study of mechanism-of what can be done by concrete, material processes. As I say, I still don't think it is a theory of computing; real-world computation involves relationship, semantics, non-efficacy. But even if it doesn't explain our main subject matter, a theory of pure efficacy is a pheno menally important intellectual project, for which I have the greatest respect.

What is distracting—the reason this is not all universally realized—is that, because of its history, the theory of computability is still framed in semantical terms (computing functions, coming up with answers, representing numbers, etc.). Thus consider the standard practice of taking marks on Turing machine tapes to denote numbers. Though historically comprehensible, this practice, I argue, is actually wrong. The marks don't denote numbers, in spite of what everyone thinks. Rather, the numbers denote the marks. Computability and complexity theory, in my view, are mathematical models of complex configurations of marks. All the regularities captured in the theorems have to do with these marks and their arrangements, not with the numbers we associate with them. Why this is true, why you have to understand it this way-that is a story I can tell you over drinks.[30] The bottom line is that the role of the mathematics, in computability theory, is just like the role of mathematics (and of objects!) in physics: it's a classificatory, epistemic device, employed by theorists. Like all theories, the theory is semantic, but it is not semantical;[31] it is not about semantics. What the theory does is to use semantics (not just terms, equations, variables, etc., but also mathematical modeling relationships, like physics) to classify concrete, in-the-world, non-semantic regularities.

Thus consider the results of computability theory, such as the unsolvability of the halting problem, or the difficulty of factoring products of large primes. Both problems are framed semantically: that you can't decide whether an arbitrary machine will halt, on an arbitrary input, that you can't figure out what numbers are prime factors. As any good theoretician knows, however, if framed in purely non-representational terms, as issues of yes/no decisions, or of pure numbers, these problems can be solved, trivially, if you employ what are called 'non-standard encodings'. For example, if you represent numbers as lists of their prime factors, then factoring them takes no work at all! Given this vulnerability, which applies to all complexity results, my argument has three steps: (1) the only way to bar such non-standard encodings is by bringing into explicit view constraints on the representations (constraints on the marks), not just on what the marks denote; (2) once you bring in the minimal constraints on marks needed in order to preserve the theorems, you have brought in everything you need; there is no more work for the 'denoting numbers' aspects of marks to do; and (3) what happens, in the traditional practice, is that these entirely concrete constraints are implicitly modeled by numbers, the numbers that the marks are taken to denote.

But enough technicalities; this is not the place for details. The point is merely that what is today called the 'theory of effective computability', in spite of the way it is framed, has nothing to do with semantics. It is a mathematical theory of physical effectiveness, pure and simple, of exactly the sort that we need for half the intentional story. It is a mathematical theory of Cummins' keyhole. [[Discussion point 5.14]]

Before concluding this topic, I should admit one thing: how much work is opened up by the reconstruction I am proposing. If I am right that complexity theory is really about the capabilities of pure mechanisms, independent of semantic interpretation, I am committed to reformulating its results in non-intentional terms. The theorems cannot be framed in terms of decisions, or numbers, but as statements about how certain configurations of the world (that is, certain machines), if started off in given effective arrangements, will or won't or can't get into other effective arrangements, or about how, if you give a machine two different input marks, sufficiently complicated, these inputs will essentially 'drown' the machine, so that it won't be able to produce one kind of output mark from one, and another kind of output mark from the other. In other words, I am committed to reformulating all the theorems as claims about effective arrangements, simpliciter, without regard to anything those effective arrangements mean. [[Discussion point 5.15]]

5 Objects

Fifth, I want to say a few words about objects—the subject matter of the book (*On the Origin of Objects*, 1996).

If you take logic, or introductory philosophy, you might think that there are two fundamental kinds: (1) concrete, individual, particular objects (called tokens, if they are linguistic or semantically interpretable); and (2) abstract, perhaps Platonic, types, which the individuals instantiate. In the general case, a type will have multiple instances: there will be a one-to-many relationship between types and their instances. In this sense, the abstract type acts to 'bind together' what is similar across the (extensional) set of objects of a given kind.[32]

One of the things that a career in computing has given me an extraordinary appreciation of is the sheer complexity of real, in-the-world, material objects. Among other things, this has in turn led me to appreciate the profound inadequacy, as an account of reality, of this simple picture of types and their instances. There is nothing magic about computing, in this regard: librarians know it too, in their efforts to catalog copies, editions, translations, reproductions, templates, and so forth. But I came to the lesson through computing.

Here is just one example of the kinds of practical issue that drove me to the story I report there. I normally download my e-mail to the hard disk on my home computer, where I have several hundred megabytes of files, folders, pointers from files to folders, and so on. When I travel, I copy the whole mess onto my laptop. Suppose I come here to Maine, dial in to a local ISP, and get a message from Dan Dennett. Intending to file it, I follow a link I have set up to the folder reserved for messages from him. This would have worked fine, at home. But when I try it here, on my laptop, a system message is displayed, asking me to mount the hard disk I left in California. Why does it do this? Because I copied the pointer from my home machine onto my laptop, and on my home machine this pointer pointed to the copy of the Dennett folder stored on that disk. Of course my laptop has a copy not only of this pointer, but also of the folder in question. The problem is that the system wasn't smart enough to know that the pointer should be adjusted to point to the copy of the Dan Dennett folder that now resides on the laptop.

What I hope this tiny example shows, or at least evokes, is the messiness of real-world issues of concreteness, abstractness, and multiple versions of 'the same thing'. Pointers are normally taken to point to individual files, but my intent, for this pointer, was that it point to something slightly more abstract: the Dan Dennett folder, of which I have multiple copies. You could say that this abstract Dan Dennett folder is a type, of which the individual copies are instances; but other than dressing the situation up in formal guise, that move doesn't much help. The problem is that even an ordinary desktop contains an astounding proliferation of highly related objects, of various sorts, many of which stand to each other in analogous one-to-many or many-to-one relations.

Copies, virtual copies, pointers, caches, back-ups, editions, versions, replications, and so on—seemingly without limit.

Similar issues arise inside programs. Suppose you call a subprocedure with a matrix as an argument. And suppose the subprocedure changes the matrix. Was the original matrix changed, or did the subprocedure modify a copy? It depends on whether you passed it, as they say, 'by value' or 'by name'. Some other examples: one variable, multiple values; one IP address, multiple CPUs; one procedure, multiple call sites; one program, multiple copies, each of which can be run multiple times; one web page, multiple servers; one web page, multiple translations into different languages. And so it goes. Templates generate multiple copies, generators spawn new instances every time they are called, etc. And when the proliferating objects are interpretable, the situation gets even more complex. For example, there are problems of context-dependence: something that means one thing, in one context, can turn up in another context, or a copy of or pointer to it can turn up in another context, and mean something different (the Y2K problem is one especially famous example of context-dependence gone awry).[33]

Needless to say, if you work with these systems, you have to keep things straight. Some properties (the number of messages I have received from Dennett, say) hold of the abstract 'one' of which there are multiple instances or versions or copies. Some properties differ across each member of the group, in systematic ways (as we will see, this has to do with indexicality). Other properties (such as the location of a file on disk) may differ across the 'many' in no systematic way at all.

You might think that the way to avoid confusion would be to be extremely, even aggressively, clear—always knowing exactly which object type you are referring to. You might even want to have different names (for example, to distinguish the program, considered as an abstract object, from concrete copies of the program, from temporal runs of the copy, and so on.) Let me simply report that all attempts I have made at being extremely clear in this way have failed miserably; they drown in inscrutable complexity.

Humans apparently handle such situations in a very different way. They seem to have a feel for the sort of things different properties can hold of, and to infer the appropriate instance or entity or individual for any given property in question.[34] In a way, you can tell that the term 'program' refers somewhat differently, in different cases, because a kind of zeugmatic infelicity arises from combining different types of reference under a single conjunction: 'Is that program recursive and corrupted?' sounds 'off'.

For many years, behind the scenes, I have been trying to develop a calculus in terms of which to understand this kind of proliferating objectification practice. I call it a 'fan calculus'—a calculus of the 'one' and the 'many'—because so many of these situations involve one thing (what I think of as the point or root of the fan) that devolves or engenders or creates or spreads out into or is exemplified by or holds of multiple copies or versions or instances or tokens.

The classical type-instance distinction is a single fan, on this generalized scheme: what we call the 'type' is at the point of the fan; the (extensional) instances constitute its fringe. Even in the case of simple language, however, it is clear that a more complex classification is needed. Yesterday, in a discussion about indexical utterances, I made a three-way division, among type, token, and use. [[See Discussion point 4.2—ed.]] Very roughly, you can think of one fan connecting the type to each different token, and another fan connecting each token to the set of all its uses (if it is used more than once). To see the utility of this double-fan characterization, note that the mentalese word 'I' is indexical on the first fan, but not on the second. Your mentalese inscription of 'I' (if there is such a thing) and my mentalese inscription differ, systematically, in their referents. But unless I am deranged, all my different uses of my (single) mentalese inscription of 'I' refer to one and the same enduring individual.

In the case of types, tokens, and uses, we typically think of the types as abstract, the tokens as enduring and concrete, and the uses as concrete events. What is interesting about the proliferation of computational examples I cited, including files and copies and versions and editions and templates and copies of templates and generators and so on and so forth, is that much more complicated hierarchies seem to exist in which all the entities are apparently concrete. (Whether that is actually true, however, is not so easy to say. As the case of the Dan Dennett mailbox indicated, there may be slightly abstract unities in terms of which some regularities hold—whatever it is to be 'slightly abstract'.)

I mention all this only to say that considerations of this sort, involving complex relationships between 'one' and 'many', have influenced the account of objects (or perhaps I should say of objectification) presented in OO. One question that is of considerable interest, as I have already indicated, is how issues of concreteness and abstractness play out across these hierarchies. And as usual my answer is the predictable one: the most interesting cases, I believe, are somewhere in the middle.

If one is strict about laying out the hierarchies, moreover, intriguing patterns emerge. Even the simple case of a proper name, standardly described as 'one name, one object', involves two fans, as indicated in Fig. 5.1: one spreads out from the name qua type to its various utterances or uses; another fans back in from these different uses to a single person. Similarly, the fan structure of the mentalese indexical described above is given in Fig. 5.2: one fan from type to tokens, individual fans from each token to its uses; and finally another set of fans back in from those uses to the holder of the token.

With respect to the overarching project of naturalizing ontology, perhaps the most important observation is the following: there are similar patterns of complex fan-ins and fan-outs underneath or 'within' the notion of an object (as above objects, having to do with particulars and classes or types). Suppose we lay out any given concrete object in 4-space, as an extensive space-time worm. If I touch you, in a certain sense my hand will touch one part of you, say, your shoulder, at just one point in time. In saying that I have touched you, not just your shoulder, and not just now, I am saying that the touch, as it were, 'fans out' across space, to all of your body, and also across time, to make contact with you as an enduring individual. Suppose I touch you again, ten years from now, and for some reason ask 'Have I ever touched you before?' The positive answer that this question warrants can only be defended by noticing that the two spatial and temporal fan-outs end up being coincident, on one and the same enduring object envelope-an envelope, as I hope is obvious, that bears some resemblance to the result of the act of abstraction that we talked about earlier, in the case of the cup or the pen.

What I find intriguing is that so many cases, from the epistemic structures of reference to the ontological structures of individuals, involve various kinds of abstraction: 'gathering up' of a bunch of that which is in some ways different, and taking the result as a unity-as that which is one. Getting to the heart of this practice is an essential part of the story I want to tell. Moreover-and in a sense this is the heart of the meta physical story-my ultimate claim is that there is no technical way to deal with the stunning complexities of these interrelated fans except by a single, integrated account that makes simultaneous reference to the aboriginal structure of the world and to the normatively governed intentional projects of the objectifying agent. One can only make sense of these structures, that is, via a 'blended' epistemological-cum-ontological account.

Finally, let me say a little about the subject Clark has brought up (in conversation): the role of indexicality and deixis in all this, and how that ties into issues of effectiveness (as you know, I have a claim that the intrinsic indexicality of reference stems from the fundamental character of physical law). Consider a single person's multiple utterances of the word 'now'. And assume, again for simplicity, that each utterance is used to refer to instantaneous moments, so that a sequence of utterances ('now! now! now!') would be used to refer to a corresponding sequence of (very short) passing moments.

What is evident in this case is that the referential pattern involves something I call 'point-to-point correspondence'. One concrete use refers to one concrete moment; another concrete use refers to another concrete moment; and so on. Moreover, the temporal sequences line up. As regards the link between deixis and physical effectiveness, the fundamental insight is very simple: physical interactions have exactly the same point-to-point correspondence structure. What takes place now affects what is happening now; what took place then affects what was happening then; what will take place next time will affect what is happening next time. In effect, this point-to-point correspondence (both spatial and temporal) is intrinsic to the structure of physical law.[35]

What happens when we objectify is that we gather a region or patch of the world into a unity. To do that requires extending these patterns of correspondence from simple point-to-point relationships, of the sort that underwrite physical regularity and simple indexicals, to much more intricate and hierarchical fan-ins and fan-outs. Feature placing, of the sort I described earlier, involves more complex forms of correspondence than simple point-to-point, but simpler than what is characteristic of the exemplification of properties and relations by full-fledged objects. To get to these more sophisticated cases, one needs to start getting involved with types, instances, and so forth, which involve complex, cross-cutting fans.

For example, suppose that last week I thought about Clark, and made a mental note to ask him, when I saw him this week, how his wife Pepa is doing. As indicated in Fig. 5.3, a complex set of fan-ins and fan-outs governs this successful ability to refer to Clark as an enduring unity. Both Clark and my mental token are temporally extensive, for starters. However, unlike the case of 'now! now! now!', and (similarly) unlike physical connection, the two temporal sequences don't line up, point by point. On the contrary, it is essential to the logic of the situation that when I thought about Clark, last week, at time t_1, I didn't think only about the temporal slice that was him-at-time-t_1—that is, about him then. Rather, I thought about him, which is to say, I thought about him as a temporally extended person (the whole lower space-time worm). By the same token, when I met him here at the workshop, at time t_2, the person that I greeted and talked to was again not the temporal slice him-at-time-t_2, but the same complete temporally extended individual. So my mentalese token fans out into individual uses; each of which fans out to cover the whole singleton person. In order for the reidentification to work, they must be coincident in that temporal extendedness.

Reidentification, in sum, requires this kind of cross-cutting gathering up and spreading out. This is in sharp distinction to the vastly simpler point-to-point correspondence that is true of all physical interaction, and that is true of at least limit-case indexicals.

Needless to say, this is just a whiff of a picture. What I am really doing, I suppose, is diagramming the field-theoretic structure of simple reference to concrete individuals. My point is only that if we are serious about our naturalism, something like field-theory[36] is all we have to start with. Somehow or other, we objectifying creatures are able to do a sophisticated enough dance to parlay our simple, effective, local, point-to-point field-theoretic coupling with our immediate physical surround into these complex patterns of cross-cutting fan-in and fan-out that characterize objective reference to the world. I don't claim to understand more than one per cent of how it goes. But it is something that I want to figure out.

Enough about objects. I have just one more general remark to make. [[Discussion point 5.16]]

6 Evolution

Sixth and finally, let me say something about evolution. In brief: I don't want to accept evolution as a rock on which to build my church. But I will accept it as a flying buttress-as something that supports my church from the outside.

I agree with lots of things that Dennett says. For example, I think he is right to say that nothing but evolution could have gotten us here, could have done all this work. If evolutionary biology is right, which I presume it is, then sure enough, evolutionary adaptivity must be the means by which we learned to register the world. The causal history (of the emergence) of our registrational capacities, the causal history of our opening up into normativity, the causal history of how and why we take the world to be significant-all these causal histories undoubtedly unfolded along evolutionary paths, especially originally (for the last 10,000 years, social and cultural and political histories have presumably carried more of the developmental and explanatory weight).

Dennett is also right to suggest certain corrections of emphasis. Of course it is true that what we want to be intentionally directed toward is what matters to us, not what we used to be connected to. I certainly don't want to be sentimental about primordial or aboriginal union.[37] I also agree that processes of connection and disconnection, processes of registration of that to which one isn't (and maybe never was) connected, and so on and so forth-these constitute an Extremely Good Trick, which evolution discovered and exploited.

Finally, let me say that I am completely open to being instructed, as regards the details of evolutionary affairs. I am no expert, and look forward to knowing more. Moreover, I feel ready for the handshake. I believe that the constructive tenor of my account ('constructive' in an engineering sense), starting from very simple patches of the world and progressing up through mechanisms of simple non-effective tracking, to featural registration, to full-blown conceptual registration in terms of objects and properties, to a form of situated objectivity-this story of 'building up' registrational capacities seems to me very compatible with evolutionary history. And I look forward to understanding better how such capacities evolved.

If I agree with all these things, then why don't I talk about evolution?

The answer is essentially this: while I will admit that evolution is the means by which we learned to register, *I don't believe that registering is itself an intrinsically evolutionary process*. The explanation of how we came to do it may be evolutionary, that is, but the explanation of what it is we do is not.

To go back to Cummins' metaphor, I think of registration as a way of exploiting the 'keyhole' of what is effective, so as to end up being oriented toward what matters, including (largely) what is not effective-in order to be oriented to the world, to take the world as mattering. I see no reason to believe that this is an intrinsically evolutionary thing to do; that non-evolved creatures could not do the same thing. Moreover, I do not believe—and I suspect this may be a point on which Dennett and I substantively disagree—that the fundamental normativity on which registration rests is intrinsically evolutionary, either. Again, that we take the world to matter may have evolved. But that the world matters is not by itself an evolutionary claim. If creatures were magically placed here, or emerged via a different means (say, on another galaxy), the world could matter to them-just as much. [[Discussion point 5.17]]

So that's the claim: that this intricate, sly, surreptitious strategy for exploiting what's effective in order to stand in relation to what's not effective is a Phenomenally Good Trick. It is a trick on which evolution stumbled, and made out like a bat out of hell. It is even clear why it is so evolutionarily useful. So there must be an evolutionary story about how it was discovered, what kinds of registration it evolved first, how it worked, what advantages it conveyed, what was tried and failed, and so on and so forth. In order to make good on the claim that it is an Evolutionary Good Trick, however, the character of the trick cannot itself be defined in evolutionary terms. In order of explanatory priority, registration must be understood antecedently to evolution, if we are to understand how it is a trick-a trick that the universe supports, that evolution could pick up and make out with. [[Discussion point 5.18]]

To make this concrete, let me talk a little more about norms, because I think norms are the place where the issues become most urgent.

In RR, as Dennett pointed out, I made a distinction between 'statical' and 'dynamical' norms.[38] Think about normative notions in formal systems-the norms we apply to processes of inference in logic, for example, and analogs in such related fields as economics and game theory.[39] What we traditionally encounter are norms such as truth preservation, inference to the best explanation, utility maximization, and so on. What is interesting about these norms is that, while they apply to processes (and hence, in my terminology, are dynamical), they are defined in terms of explanatorily prior norms, such as truth, reference, explanation, utility and so on, that are defined on states (and hence are what I call statical).

My point in RR was that an extraordinarily important intellectual shift is taking place, across the intentional sciences: the explanatory order is changing. For a variety of reasons, ranging from evolutionary considerations, to the sheer difficulty of characterizing such age-old statical norms as truth and reference, to concrete experience building and maintaining and using computer systems, people have come to realize that the only way to define norms such as truth and reference in a useful and non-question-begging way is to base them on how things are used. In a sense, we are all becoming closet Wittgensteinians. It is not so much that statical norms are being discarded (truth, function, utility, etc.), as that they are being understood as derived from dynamical norms.

The reason why this shift in explanatory priority from statical to dynamical is so important is that it puts pressure on a source of dynamical normativity. If you are going to define your statical norms in terms of your dynamical norms, then you cannot define your dynamical ones in terms of the statical ones, on pain of circularity. You need something else as a source of dynamical normativity. What is that going to be? Especially for a naturalist, this is a very urgent question: what naturalistically palatable source of dynamical normativity is available, on top of which to construct the entire normative edifice necessary for full-blooded intentional characterization?

This is the role that evolution is playing, I believe, in many of the intentional sciences. Evolutionary advantage is an extraordinarily convenient candidate on which to rest an intentional story. Ruth Millikan and the general project of teleosemantics can be seen as one example of this trend. And I take it that Dennett is proposing something similar, when he says all normativity rides on Darwin's coat-tails. He is basically asserting that evolutionary advantage is the 'mother of all norms', the ur-dynamical norm.

Unfortunately, I don't believe evolution is a strong enough base on which to rest all of human normativity. For think about how much the general issue of dynamical norms includes. Ethics, for starters; and even more generally, how to live. I cannot muster arguments here, but I simply don't see how one could milk evolutionary survival for this full range of normativity, for all that has inspired individuals and cultures, led people to distinguish good from bad, and so on and so forth, over the ages. Remember: I am not denying that human normativity may have emerged evolutionarily, at least at first (that is, to the extent that there was substantive normativity prior to the development of culture and civilization-something on which I have my doubts). All I am denying, to repeat my standard refrain, is the claim that normativity is an intrinsically evolutionary notion. [[Discussion point 5.19]]

So I have said that evolution is not strong enough to be a basis for all substantive norms. Clark and Dennett ask [[Discussion point 5.19]]: 'Is it strong enough to give rise to them?' The answer depends on what you mean by 'give rise to'.

If by 'give rise to' you mean causally, historically, how did these norms emerge? What engine could have done all this work to get us here? Isn't this the only mechanism through which normativity could have emerged? My answer to that question is 'yes'. Evolution may have been the train on which norms arrived at our present station. And as I said before, you can see why it would have arisen, evolutionarily: the ability to care, to register their world, to take things as significant, confers a huge evolutionary advantage.

But if by 'give rise to' you mean (as I do) something more metaphysical, something more explanatorily substantive—and something more like the word 'origin' in the title of my book—then my answer is 'no'. Even if it is an a posteriori necessity that, given basic material facts about the universe, about the origins of life, etc., evolution is the only means through which normativity could have arisen in registering creatures, nevertheless, it is not constitutively intrinsic to our normativity, I claim, that we have an evolutionary history. Remember: I want to be able to explain what normativity and registration are, such that we can say of evolution that it stumbled on them. But the fact that evolution stumbled on them merely makes evolution the implementing mechanism whereby we came to be normative.[40]

In detail, I should admit, there is undoubtedly tons to be learned from our evolutionary history—about what kinds of normativity there are, what kinds we have evolved to be attuned to, etc. And (perhaps even more so) there is tons to be learned about what our registrational capacities are, what constraints they have evolved to satisfy, and so on. All those things are interesting and useful. I just don't want to allow the interest and urgency of asking those questions to seduce us into what I think of as a kind of non-reductive causal foundationalism: wherein we confuse what things are with the causal history of how they came to be. [[Discussion point 5.20]]

It may help, in understanding this, to think of registration first, and normativity second. Remember what I want to say about registration: that it is a mechanism whereby you arrange things so that you can track long distances, put together machinery and external signs and external scaffolding and so on and so forth, so as to stand in relationship to more and more and more-leading, ultimately (this is where consciousness and objectivity merge) to a profound and care-full orientation to the whole world. Do we need evolution to understand that? In a constitutive sense, I think not. In an historical sense, I think we do. It doesn't seem to me that evolution plays a constitutive story in understanding how the strategies of exploiting local effective structure can stand you in relationship toward that with which you are not physically coupled (though I admit: the hardest issue-the nub of the matter, as it were, and maybe for that reason an appropriate subject for drinks, later-is what 'being oriented to' really means, in all its respectful, concernful richness and depth). For now, let me just say that I believe the same thing about the norm-side: that an evolutionary account may tell us how we got here, and may tell us in detail what the costs and trade-offs are, in regards to honouring norms, but it won't give us a constitutive account of the structure of normativity itself.

It is time to stop; but I hope that this makes it at least a little bit clearer why I want to accept evolution as the train, but not confuse it for the goods that were transported on that train; to accept it, as I said at the beginning, as a flying buttress, but not the rock on which our registration of the world is founded.

6c — Dennett & Smith · Discussion

CUMMINS: This is what anti-realists do on purpose.

DENNETT: Commit this inscription error?

CUMMINS: Just, lots of them.

SMITH: Here's a simple case, to see what's going on. Suppose you want to know if your cat recognizes you as an individual as opposed to as just more 'Hughness'. And suppose people say, 'Yeah, look, because here's its neurophysiology and this cell lights up every time—same cell, so it must be the same person.' But how do you know it's the same cell? Maybe it's just reinstantiating some type in there. How do we decide that's a second use of a single token, as opposed to new instantiation of a single type? See, when you say 'one cell, so one individual', you're making one assumption about object identity in the head of the cat, and then assuming that you can use that identity to warrant a claim about object identity in the content of the cat, in what the cat's identifying. But I can redescribe the situation in the brain, and then get the presumptively competing suggestion about the content. All you're really doing, in other words, is piggybacking your analysis of content identity off brain identity. We should worry if our analysis of whether the cat recognizes Hugh as a type or an object depends on empirically equivalent ways of us theorists individuating its brain, especially since the cat itself doesn't individuate its brain at all.

I think these things happen very subtly, even in modest cases. Earlier today we were talking about whether we objectify things on a map. And it depended on how we registered the map in the first place. Take a line: do we call that a relationship between two points, where the points are the objects? Or are the lines the objects and the points just relationships between two lines-where they cross? Problem is, in one case you end up saying we're objectifying the line, and the other case you don't. It's that kind of thing. How we as theorists register the problem domain affects our analyses in ways that are stunningly consequential.

CLARK: So it's a worry about the baggage that comes along with the labels. Like all those worries people have about how you label the nodes of your semantic network.

DENNETT: So, good. You can go way, way back to Drew McDermott's old paper 'Artificial Intelligence Meets Natural Stupidity' [1976], and that was a sort of ur-anxiety about inscription errors which Brian has generalized.

SMITH: In a way, I think what I might say is: Look, inscription like this, it's something you have to do. It has enormous consequences; so you want to be tremendously modest and humble and cautious. It's not black and white; it's not as if you can say 'Here's an inscription error, here's not.' You always have to inscribe. The issue's just this: don't let the fact that you have to inscribe license you to project all sorts of ontological assumptions all over the subject matter without taking responsibility for them.

CLAPIN: OK, so that's Rob's point. That's what anti-realists maybe are doing, is that they are allowing a license for them. They say, well we've got to do this anyway, so there are no constraints.

CUMMINS: Yeah, other than internal coherence.

CLAPIN: So the way you're describing color, Brian's kind of saying is true for…

DENNETT: Objects.

SMITH: Everything. To think that taking human relativity seriously implies irrealism is only true on a Cartesian view that we're not part of the world. But if someone were to write a book called Being There, or something like that, and actually realize that we are here, then from that (correct!) point of view, human relativity shouldn't be metaphysically scary. It is profoundly consequential, but it's not skepticism or irrealism.

DENNETT: I think it's like this. I think we can talk about the colors of things on distant galaxies by helping ourselves, and knowing we're doing that, to human color vision, and using that as our standard. And Brian's saying we can talk about primordial objects by using human object vision and recognizing that that's what we're doing, as long as we keep track of the fact that we're using human object and property vision as our standard, as our perspective, we can sort of discount—sort of like discounting the illuminants. That's the idea, and I'm snowed by it, whether I should be or not. I sort of like it, so, we'll see.

HAUGELAND: Is this also an example of the same point—I can't remember what I've thought about this in the past—that, before the human race evolved, the moon was 240 thousand miles away, even though there weren't any miles then?

CLAPIN: In that kind of an example, is the thought to go sort of de re-because miles didn't exist, *de dicto* how many miles is just not a sensible way of talking?

HAUGELAND: Well the trouble with that way of putting it is that it supposes that de re just is as it always was.

DENNETT: But thank you for raising it, Hugh, because that helps with one of the reasons why I'm attracted to this: because it helps me fend off that awful de re/de dicto stuff.

SMITH: Also, it will have more bite when one realizes that taking the world in terms of objects and properties is underestimating the world. That's when the approach really starts to fight back-not in a way that Rob will swallow, perhaps, but in a way that is at least akin to something Rob would swallow.

HAUGELAND: 'If everything were rigidly blocked in the universe, there were no flex and slop and slippage, nothing could be out of touch with anything else'-that seems to me to be completely wrong. What is rather the case is that there couldn't be a distinction between being in touch with and not being in touch with. You couldn't be in touch with anything else specifically if moving this moved everything.

SMITH: Well I couldn't be some thing. I don't disagree. I think the very language we speak so presumes that there is a certain looseness in the world that our attempts to describe what it would be like if there weren't a certain looseness will all fail. But they don't fail completely. We can have this conversation which we're having, and it makes some sense. I actually think it's impressive that we can agree, 'Yeah, all these things fail, but, we actually do get a sense of what we mean.' We can actually reach a kind of consensus which, if we all said, 'OK, now we've got it; let's try to say it,' we wouldn't do any better than we just did.

CLARK: So it's not essential for there being objects that the universe not be connected like that?

SMITH: Yeah, it is, because in order for it to be an object it's got to have distance. And it's got to have shear.

HAUGELAND: It's got to have distance and it's got to have an internal life that is different from what's around it.

CUMMINS: This was Cartesian physics. Everything was locked together. It's just that minds were allowed to slop around a good deal-but the physics was all locked.

SMITH: Well dissipations of forces are tricky. The problem with Cartesian physics is that it isn't theoretically precluded that you can tell everything from just how one little thing is vibrating.

DENNETT: That's just what Newton saw. This is right. You're sort of recapitulating a Newtonian revolution here with your point about flex and slop.

SMITH: One thing that might help explain it is that I don't think space-time points are echt individuals. Other than particularity, I actually think they lack all of the characteristics that individual objects actually have to have to be objects. Think about the field theoretic interpretation of classical physics which is actually doing some work in this story. Imagine a rubbery manifold with forces going up and down, and all that kind of stuff. You can imagine that if everything were just spatio-temporally infinite manifolds, there could be lots of space-time points as it were, but there would be no clumping of them together into reidentifiable individuals that have heft, size, or separateness, and so on. I think that field theory, this rubber manifold stuff, is probably the best imaginative route in. By particularity I just mean roughly spatio-temporally concrete occurrence.

CLARK: So it's like saying particularity is not objecthood. They're not objects, are they, space-time points?

CUMMINS: Cartesian points don't move around, although they do have properties.

SMITH: That's right. They don't move around, because they don't do anything-that's the problem with them.

DENNETT: A nice way of thinking about it might be to think about the individual cells in Conway's life world-they are particular but not individuals. But a glider can be individual.[41]

SMITH: I think if I were to write the book over again I might have said either 'occurrent' or 'concrete' instead of 'particular'.

HAUGELAND: Well, I think what you want is concreteness. I'm not sure what you mean by occurrent, or do you just mean actual?

DENNETT: Concreteness is not individuality.

SMITH: Right; that is close to what I mean. Note, for example, that in the book I embrace a 'criterion of ultimate concreteness', so using 'concreteness' as a word would be relatively straightforward. I do believe that everything is concrete. But as a word, 'concrete' is misleading, too…

CLARK: And what are space-time points here, they are…?

SMITH: They are concrete, but not individuals. You wouldn't treat them as individuals.

CLARK: That seems kind of funny if space-time points again come out as concrete. They seem like paradigm cases of something that's not…

HAUGELAND: Well, look, what do you understand by concreteness? This is what I understand by concreteness. That in every respect in which it can have a feature, say, in some degree, that degree is fixed. Nothing is left free.

CLARK: OK, so it's well-definedness or something.

HAUGELAND: Well, no, it's more than that, it is in a way, the difference between kinds and particulars. It's a metaphysical thesis and I can't make up my mind whether it's analytic or not, that particulars are concrete. That is, you can have the picture of the man that doesn't indicate whether or not his fly is open, or whether or not he's got a bald spot on the back of his head, this is left open. A sentence likewise leaves things open. But the man is…everything is settled.

DENNETT: That's what I find appealing.

SMITH: Another thing that I want to say, which seems to me a simple point, but is hard to phrase using traditional terminology, is that objects-individuals, essentially-are also abstractions. And by being abstractions I mean that some of their concreteness has been…

CUMMINS: Lost.

SMITH: Well, in a way, but it's tricky. Consider a cup. On the one hand, the cup is fully concrete. Taking it as a cup, however-gathering and clumping this chunk or region of the concrete flux, and treating it as an individual unity-saying, 'Okay, this is a cup'-to do that is to ignore some of its concreteness.

DENNETT: That's the price you pay.

HAUGELAND: That's to say that the kind isn't concrete.

CLARK: So this is just the price of my kind of data compression.

SMITH: It's tricky. It's not just the kind that is abstract. I want to agree that the kind is abstract.[42] But the cup is not abstract in the way that the kind is abstract. In taking the particular cup to be a cup, to be an individual; that act of objectification is an act that ignores some of the concreteness.

HAUGELAND: That is there in that vicinity.

SMITH: Right, it's in that vicinity. Taking it as a cup ignores, it packages the thing together, takes this distributed part of the flux as a unity, a whole lot of things like that. Come at it epistemologically. It's really that objectifying is an act of abstraction. At least at first blush, it's not that the thing that's objectified is abstract, really, because it actually is as fully concrete as you think.

HAUGELAND: Right, the definition of concreteness is that everything that could be determinate about it, is fully determinate.

CLARK: I'm having trouble with the 'it' here.

DENNETT: That's where we get the inscription error you can't get out of.

SMITH: You are right that 'it' is the problem. You have to realize that there is more to the cup than figured in you're taking the cup as a cup.

HAUGELAND: Well, that's really true.

SMITH: Perhaps, for now, I should take that platitude that you, John, would agree with, about everything being determinate, and then just locate the individuality of the cup more in the act of taking it as an individual, and less in the concrete patch of the world there might be-

HAUGELAND: Here's a motto I would think you would be sympathetic with (even though it isn't actually using the words quite the way you want to), which is to say that the achievement of objectification is achieving an 'it' such that concreteness makes sense. So, to be objective just is to be, and, I think, a thing, an object. Actually, not just a thing, an object. You have to get it into a space of possible determinacies such that for talking about 'this one', full concreteness makes sense.

SMITH: Yep, but we still differ, for several reasons. One is that I want things to be concrete that aren't objects-

HAUGELAND: That's cool.

CUMMINS: That's all right.

SMITH: But I want the determinacy-no I'm not sure it's cool with what you just said, is it?

CUMMINS: Yeah, sure, that's OK.

HAUGELAND: Yeah, I said objectification, is-the mark of success is-that you've gotten the sense of the possible determinacy in place such that it can be fully concrete. But that doesn't limit where else might be concrete.

CUMMINS: OK and then so from the point of view of the structure, the data structure you do create, a lot of futures don't count as cup futures. So you get this distinction between qualitative and substantial change just built into the finitude of your representation.

CLARK: So here's what's getting balanced. In thinking about these things, you have to recognize them as objects. To recognize it as an object is precisely to think that it has all these features and that they are fixed. And on the other hand, what you really want to do is have a sense of what differences don't make a difference. So there can be all sorts of things that can change but you still ought to recognize it as the same object; the cup can get chipped, you still want it to be the cup. Hence even the Y2K thing, you might think that the problem here is that there's a difference that does make a difference that we never thought about.

SMITH: It also has to do with projects. The commitments that underwrite identity arise in part from one's commitments. It follows that the identity of an object doesn't inhere in the object itself-that is a very important theorem of this view. So that 'being an object is not an intrinsic property' would be a way to say it.

DENNETT: And that's why when the hyper-intelligent extra-terrestrials arrive and find that we're still stuck with objects and properties, this is the ultimate Y2K problem for us. Everything stops.

DENNETT: What holds constant when you turn the knobs-sort of tuning for the null, as we say, in radio direction finding.

CUMMINS: Yeah, or my idea of which invariant you need to track in order to understand the variance in the error signals.

DENNETT: Yeah, we're reaching convergence on this way of thinking of the idea.

SMITH: This idea is massively more applicable than just in the case of objects.

CLAPIN: The non-modularity of the mind as somebody said earlier in the week.

DENNETT: Well, the main thing is that Brian is saying, the overcoat is really thick, and that's where all the action is, or a great deal of the action is.

SMITH: I don't want us to infer wrongly from the fact that we theorists don't quite know what's going on, that there isn't something quite precise going on-even if what's going on is something that neither we nor anyone else can actually say. The lack of being able to say it doesn't mean that there isn't a fact of the matter.

DENNETT: Let me go back to my ur-example of indeterminacy of content. In 'The Ability of Men and Machines' [1978b] I describe a case where the engineers find this device on the beach and they study it and they agree completely about its physical constitution, and every atom of its being. They agree on exactly what trajectories it will follow under all circumstances. They disagree about what it is, what it's for. And it's only when we get to their content level, where they treat certain things as malfunctions-one of them treats certain events as malfunctions, the other one has a different gloss and says those aren't malfunctions, that's signal not noise—and I claim that it is not the case that there must be a fact of the matter about which is the right content gloss on this object.

CUMMINS: There's still a scope ambiguity, because the way you put it leaves out the possibility that there is a fact of the matter but they're both right. I get this all the time, because they say-'which isomorphism?' You know? All of them. They don't like that. Somehow something couldn't have two structures at once.

DENNETT: Right, but of course you can have them.

CUMMINS: And really, there's a real fact of the matter that they have all of them.

DENNETT: But there isn't a real fact of the matter about which one is privileged.

CUMMINS: That's right.

DENNETT: And that's the one point that I've always wanted to insist on.

CUMMINS: Privilege is always observer-relative as it were.

DENNETT: Thank you.

CUMMINS: Right, but it is a bad argument from the observer-relativity of privilege, and some premise that somehow builds in that it isn't there if it isn't unique to just rampant conventionality of all this. And you write sometimes in a way that suggests to me that you think that there's-

DENNETT: I issue a tentative mea culpa. I think I probably do write as if, I think you may have me there-

CLAPIN: But with the multiple isomorphisms in the same thing-usually the case is that only one of those structures is actually doing the effective work.

DENNETT: That's perspectival too.

CUMMINS: Yeah, that's perspectival too. There's any number. It's simple and clean to think of these couplings one at a time. But the fact of the matter is, this is just another one of these things. In Dan's case, you've got one engineer, as it were, coupled into one structure, and another engineer coupled into another structure. And since those two are different there's a temptation to infer that there's no fact of the matter of which structure is there. They're both there because after all, if they weren't, the two engineers couldn't be coupled to them.

DENNETT: I have a Quinian crossword puzzle. It's very simple as a crossword puzzle, but there's two solutions to it. I hand it out to my students and I say this is a simple little crossword puzzle, see if you can solve it. And they come up with the two solutions.

CLARK: If someone found a third solution, what do you say about that?

DENNETT: That's fine, too.

CUMMINS: They're all fine.

SMITH: Let me just say a bit about these two isomorphisms-two structures in one thing, right? I think what I would want to say is, there is one thing, of which two abstractions hold.

DENNETT: Yeah, that's fine.

SMITH: But it's not exactly as if there's indeterminacy in the concrete.

DENNETT: No, of course not. We agree.

CUMMINS: I think there's a metaphor that may be misleading you here, Brian. You tend to think of structures like shape. You say 'Well out in the world there's the cookie, and I've got a whole drawerful of cookie cutters,' right? And, in some sense, if all the cookie cutters are different shapes, then it just seems to follow that they couldn't all fit the cookie. But that's just because you've got a very limited notion of structure.

SMITH: No, look, that's not the point. I have no trouble thinking that seventeen different cookie cutters could fit the cookie. All equally well. What's indeterminate in that case is which type this token is an instance of.

CUMMINS: Why isn't it an instance of all of them?

SMITH: It is an instance of all of them. That's fine. What I'm saying is that one question one might ask is, 'Look, I'm not sure which of these cutters applies,' and the answer is 'All of them.' That's the question you're talking about; it has to do with the cutters. That's not what I am talking about. What I'm saying is determinate is the cookie, not the cutters.

CUMMINS: But you don't want to understand the determinacy of the cookie as somehow a matter of how many cutters fit.

SMITH: You absolutely don't. I agree with that. So I want to say, 'That's right: in the actual concrete thing there's no indeterminacy.'

CUMMINS: Why isn't the determinacy just the determinacy of fit? It determinately fits this cutter, it determinately fits that cutter, and it determinately fits that other cutter, and that's all there is to it. You exhaust those facts you're just done.

DENNETT: Let me read the passage I stubbed my toe on, all right? On page 68 of OO, Brian says. 'Somehow or other—and this I take to be the most important and difficult task facing the cognitive sciences—it must be possible to have determinate representational content, i.e., for there to be a fact of the matter as to how the world is represented.' Brian goes on to say, 'it will have to be an answer that does not depend on how anyone registers or individuates those mechanisms-again, for the simple reason that it happens in people, for example, without anyone doing that.' Right. It doesn't depend on how any observer registers or individuates the mechanisms, but there may still be many different ways of interpreting those mechanisms. And no one of those ways is privileged.

CUMMINS: And, moreover, all of them might be either a little or a lot wrong.

SMITH: Sure, but the point is, if I'm looking at the rug, registering the rug, and you're a theorist and you're looking at my mechanisms, you've got all kinds of ways of doing it. That's fine. But no amount of slippage or indeterminacy or multiple categorization or simultaneous truth or anything in your interpretation of my content has any consequence as to how I take the rug.

CLARK: It doesn't follow from that that you take the rug just one way though, does it? The fact that how you take it isn't determined by how someone, as it were, takes you to be taking it, doesn't imply that you take it just one way.

DENNETT: This really is Brian's problem of the indeterminacy of radical translation.

SMITH: It may be that the way I take the rug is in fact to take it simultaneously as instantiating three different types or something. I'm not saying that I can't multiply categorize the rug. It's not the plurality there that is worrying me. Nor, given my love of smeariness, do I have any problem with saying that I take it to be an indeterminate category.

CLARK: But you do think that there is something about your state that absolutely fixes, as it were, whether it's a plurality of three or four or-…

SMITH: Think about this fact: that while there may be questions about, as it were, the classification of the cookie, it doesn't make sense that the cookie is indeterminate. I guess what I'm saying is that when I take the rug to be a certain way, there's a concrete situation here and a concrete situation there and a concrete relationship, about which there are then questions of how to categorize. And something that's going on here is that I may be categorizing the rug in some way, or not, or something like that.

CLAPIN: Don't you have to be, to be taken as an object?

SMITH: That's right, but there are lots of ways to have a take on it, but not take it as an object. I guess I'm saying that this situation here that is happening is itself fully concrete.

DENNETT: Of course it is. But now look-

HAUGELAND: Fully concrete under some descriptions.

SMITH: No, no, that's where we go back about an hour.

DENNETT: It's fully concrete.

SMITH: Concreteness is not a property of characterizations of things.

HAUGELAND: I'm not sure that you can just say that.

SMITH: Well, I certainly can't just say it in a peremptory sense!

HAUGELAND: But the thesis of the indeterminacy of translation, carefully named, is that there are some characterizations of things which cannot be fully determinate. There are just systematically different ways to characterize them which are all equal, there's no choosing among them except sort of randomly or on convenience or something.

DENNETT: They all tie for first.

HAUGELAND: They all tie for first. And there isn't a right answer, there just isn't a right answer.

SMITH: The problem is that I object to something in that thesis.

DENNETT: I think you do too. That's where there's a problem.

CUMMINS: Now I don't think you need to object to that. I think you can allow that-and still hold on to your concreteness.

CLAPIN: I thought concreteness was open to all possibilities-you know, it's kind of, before it's been conceptualized, before it's been categorized, so it's open to all of those categorizations, all of those interpretations. There's a sense in which the concrete things are precisely indeterminate, that all categorizations tie for first because concreteness is before categorization.

SMITH: That's right, that concreteness comes before categorization. That's why it is a metaphysical position, not an argument. That the world, as it were, comes completely concrete. Subsequent to that, there are issues of categorizing.

CUMMINS: Yeah, sure. I think I'm with you on this, because I think that when, as it were, the world gets targeted, when it becomes the case that the thing that is the target is fully concrete, nothing in the structure of my intenders will allow me to read off all that concreteness or anything. The intender is a pretty sloppy instrument, but fortunately, in this case, the world is there to saturate, to fill in all the holes in the cheese, in a way. And of course equally my representations will just sort of skim the surface in various kinds of ways. But the-whatever it is-

SMITH: The stuff.

CUMMINS: Yeah—it is fully concrete. I don't have any problem with that. But I think you [Smith] ought to have a problem with that.

CLARK: It seems like what you're appealing to is just the fact that there's something absolutely determinate going on. But of course there is. It's like, OK, so something absolutely determinate is going on when the thing crosses the electric eye, and that event triggers something else. But that doesn't make it, as it were, determinate whether or not the electric eye is taking it as a bee-bee rather than a fly, or whatever.

SMITH: I appreciate that. I appreciate that if I'm looking at the rug, there's something absolutely determinate about the rug. And there's something absolutely determinate about my state. And there's something absolutely determinate about the relation I bear to the state I am in. So there's something absolutely determinate going on around here (in the vicinity of my head). And the question is, one of the things that's going on over here is a taking, right?

DENNETT: Not one. That's the point.

SMITH: OK, there's some taking going on around here. (In the end, of course, what one has to say is that part of what is absolutely determinately going on around here can be taken as a taking-or taken as some taking.) The question, though, with respect to that absolutely determinate taking, is whether it has an absolutely determinate content?

DENNETT: Yes, that's the question.

CLARK: OK, so 'taking' there is irrespective of content. I'm having trouble keeping them apart.

DENNETT: Here's the question as I understand it. There's Brian looking at the rug, and three neuroprousts are scoping out the situation. As neuroprousts, they know everything about what's going on in his brain and about the light impinging and so forth. So they agree on the absolute determinate situation vis-à-vis Brian and the rug, right down to the finest details. So that's the absolutely determinate thing, and they all go to write their neuroproustian accounts of Brian's taking. And they come out different because it's like my Quinian crossword puzzle. It just turns out-extraordinarily implausibly, but you want to make this just an actuarial point. That they come out with three different contents for Brian's taking, and now the question is whether one of those is privileged.

SMITH: OK, so, go on, suppose they come out with three.

DENNETT: Now, there is this Cartesian intuition—it has been Quine's job and my job and a number of other people's jobs to beat it up at every opportunity—which is the museum myth of meanings, which insists that at least two of those neuroprousts are wrong, and you, from the inside, know what the truth is. And that is the fundamental intuition that Quine is setting out to destroy, and I think he's right.

SMITH: Right, but notice something. My claim that there is an absolutely determinate content—that my content is absolutely determinate-doesn't imply that there is any way to settle the question of which of these three others is right. It doesn't imply that there is any way, in this world, to grant one of them priority.

HAUGELAND: Does it imply that at most one of them is right?

SMITH: No. Nor does it imply that the subject of the taking knows, as it were, what is and what isn't right. The reason it implies none of those solutions is because all of these things that the neuroprousts are doing, and what the-what does Descartes call it?-infallible introspection is presumed to do, the transparency of the…

HAUGELAND: The natural light.

DENNETT: The light of reason, yes.

SMITH: All of those things are more registrations of the taking. They are registrations of my original registration.[43] Both the neuroproust's registration of my original registration, and my own meta-registrations of my original registration will approximate and categorize and lose detail and so on and so forth. That is part of my picture: that absolutely every story massively misses what it registers. So part of what I'm saying is that this picture of registration and the location of ontology is in fact a kind of negotiation between the epistemic act and that to which it's directed. Such a picture makes room for all the Quinian kinds of points, and the sorts of points you're making. Just as you, Rob, were saying a moment ago, it is because our ways of getting at the world are approximate and sloppy; there are all kinds of room for error. But the picture also preserves, I think, a durable intuition, which I think is right: not only are there absolutely determinate phenomena, but also that there may be absolutely determinate content, even if that content doesn't totally tie down the part of the world it registers.[44] Note that the fact that the content is absolutely determinate in my story doesn't mean that content captures all of what's absolutely determinate about the rug-as usual, it massively misses; that's why this is a story of loss, as I keep saying in the book. My only point, here, is that I think that in all of the cases that you are bringing up to show that my position is wrong, there is an extra layer of reference or registration or description, between what I am claiming is determinate and what you are claiming is not determinate. And I'm saying that in one sense you are right, but that that is why-

DENNETT: All right, I see where that's going. As long as you clarify all these things that don't follow from your position.

CUMMINS: Can I just ask one really quick question? Inscription isn't always error?

SMITH: No—or rather: yes, you're right. I actually meant to say that.

CUMMINS: So it's open for Dan to say, OK, when I tell my evolutionary story I'm inscribing like mad, but I'm getting it right.

SMITH: In a way I'm trying to be really clear about this in the irreduction chapter. I say look, what I require is not that you don't inscribe. What I want you to do is to take responsibility for the fact that you are inscribing (as, of course, you must).

CUMMINS: Well if inscriptions are errors there must be some mismatch between them and something.

DENNETT: They must miss their targets.

SMITH: It's a better or worse kind of story.

HAUGELAND: Dan, why didn't you reply to Brian just now, when he made the response to you, that he's pointing out that any registration of any phenomenon is bound to fall short of the full determinacy of the phenomenon-and so there's obviously various ways in which different registrations can fall short-that the issue isn't the certainly undeniable fact that any registration must fall short, but rather that, in some cases, any registration which is sufficient to capture a certain kind of richness in what's there, must inevitably overcapture it, and so there's more than one way to do that? And there's no choice between those.

DENNETT: I like that because it nicely conveys a point which people have been making in different ways, and that is, there's a real benefit in carving the world one way or another. The cost is-it's presumptive, and you always get some leverage that won't work as it will turn out, that goes beyond what you've been given.

CUMMINS: There's no free lunch. Getting A right inevitably means you compromise B.

SMITH: By saying 'overcapture' do we mean not only that it doesn't, as it were, represent things that are the case, but does represent things that aren't the case?

HAUGELAND: No, it will render some things in a determinate way. It cannot but render them in a determinate way, to capture as much as it does capture, when there would be other ways of rendering it in a determinate way, distinct determinate ways, which are equally good—I mean, indeed capture exactly the same part of what they capture from the original structure.

SMITH: But anyway my answer to that is going to be the same: from the fact that no registration is preferable, maybe even in principle (which is sort of being assumed here) the indeterminacy of the thing registered doesn't follow for me.

CLARK: But nor does its determinacy. Isn't that kind of the point?

SMITH: No, that is right. That is why the stuff about determinacy is only on page 52, instead of page 252. It is because it is a metaphysical kind of determinacy, not a-

CLARK: But what makes you think that's a determinacy of content? The fact that there's more to the content than any story will capture doesn't imply that the content's determinate.

SMITH: No it certainly doesn't imply that. However, neither does it imply the falsehood. The fact that there's more to something than a story will capture doesn't imply that the story's content is either determinate or indeterminate.

CLARK: No, that's right.

CUMMINS: When you're taking scientific laws here, you're thinking of dynamical laws? You're not thinking of laws that, for example, just tell me what's in the kit? I mean, I think it's a scientific law that there are electrons.

SMITH: Right.

CUMMINS: Rather than tell me what's going to happen next, or anything like that. It's not dynamical. It doesn't-

SMITH: Right. I'm saying that it's not obvious to me that physics is committed to there being electrons as opposed to the (weaker) claim that the electron feature is spatio-temporally instantiated in various ways. That, whether it's one electron or seventeen, and so on and so forth—physics doesn't care: it makes no commitment to reidentifiable individuals.

CUMMINS: Well, in my view of things, a lot of science is about mechan isms and how things are built and put together out of stuff, and-it all sounds less plausible as a story about those things than it does-

SMITH: It may be less plausible about engineering-

CUMMINS: As a story about dynamical physics.

CUMMINS: And you mean this in some very strong sense, I mean Dan believes this too, because if the laws of physics were arbitrarily different then nothing would ever replicate and so, as it were-

SMITH: Right. I mean it very strongly.

CLARK: You don't think of them as inscription errors, field theories?

SMITH: Like all registration, they are somewhat pre-emptive. But since they don't register in terms of objects, they give us a leg up on what it is for subjects to register the world in terms of objects.

CUMMINS: I want to make sure I understand the project. The image I get is sort of, OK, because, for example the way we're built, there's going to be this sort of keyhole effect, that you can't, your actual coupling with the world is pretty limited. So the question is, how can you see so much through such a little hole?

SMITH: That's right.

HAUGELAND: Unlike homing in on a magnet if you're an iron filing?

SMITH: Yes, exactly: unlike homing in on a magnet. There (if you are made of iron) you can be driven by the magnetic field. The problem is, it is hard for me to be driven directly by most of the things I care about—such as by Andy's philosophical views.

DENNETT: So, in the past I've talked about making something that can detect whether something has once been on my desk. It's extraordinarily hard, unless, of course, my desk was made of uranium or something and it imparted some Geiger-countable property-then, you could use that as a proxy. But in fact we are able to detect all sorts of properties for which there are no natural cheat proxies. How the hell do we do it? We have this elaborate technology for tracking things so that we can, with really very little effort, register, think about, all these weird properties.

CLARK: There does seem to be a sense in which it's our practices of timing things that brings 4 o'clock into being. It's not exactly as if it's sort of out there and we just have trouble tracking it.

DENNETT: You're not accusing him [Smith] of making an inscription error are you!?

SMITH: Andy, you are right; o'clock properties are pure human constructs. You might think that they are so stunningly non-effective in part because they were created, but that can't be quite right. What really matters about the o'clock properties, for the point of the example, is not that they are constructed, but that they are purely formal, in a certain (not so simple) sense.

CLAPIN: So it's kind of like adding to physics. Physics has numbers, has maths, it's just adding a bit of sort of logic and what we now think of as implementation theories as a bit of extra formal apparatus for physics-there's a bit more maths.

DENNETT: It's a different maths.

CLAPIN: It's a slightly different maths that's being used.

SMITH: Yep, it is new math. But what is important about (so-called) computability theory is not the math per se, but that it makes new claims-claims that get at concrete regularities that seem to hold, in the world, that involve issues of stabilization and digitization and so forth, claims that seem to be level-independent, that appear to hold across different substrates and at different scales.

CLAPIN: Like numbers do. Like cardinality does.

SMITH: Well I'm a little reluctant to cosy up too close to the numbers, because many people think that they are a genuinely abstract phenomenon, whereas what I am talking about here are concrete phenomena, at different levels of abstraction.

DENNETT: So it's sort of dependencies that are scale-independent and substrate-independent.

SMITH: Right! Or rather, originally it was exactly that-genuinely independent. Increasingly, though, it is morphing into a theory of dependencies understood in relationship to (arbitrary) scale, in relation to different kinds of substrate.

HAUGELAND: Well, substrate-independent and scale-independent don't mean substrate-less and scale-less; rather that you can have different and in some sense perhaps arbitrarily different scales and substrates and see the same phenomenon.

SMITH: Right. Exactly. All I am saying is that I think the theory of how they relate to different substrates may actually end up being part of the new theory, the direction the theory is taking.

DENNETT: Oh, well, certain features, like resistance to decay independently of the process you're considering, or constancies…

SMITH: And compensations for stability, you know, and for tunneling, cosmic rays, things like that-what kind of circuits will be stable in a nanometer scale, what kinds of stability will in fact hold over periods of weeks, years, etc.

DENNETT: Insulation properties, in fact.

HAUGELAND: Well, you can't make a Turing machine tape with frogs either.

SMITH: Right! There was a person, I remember, when [Digital Equipment Corporation's] DEC20 was made, whose responsibility it was to track the radius of curvature of the lines etched into silicon, because the bits tended to fly off etched into silicon, because the bits tended to fly off the tracks, if they went around sharp corners too fast—like errant Ferraris. The pulses would just radiate, you know; not make it around the corner. I don't know exactly how medium-independent or non-independent that is.

HAUGELAND: Well, medium-independence never meant that you could do it in any medium whatever. It never meant that the medium is irrelevant. It just meant that you could have the very same thing in quite different media.

CLAPIN: Two media would be enough.

HAUGELAND: No, the thing has to be sort of open-ended.

SMITH: Suppose I say, 'Look-this table leg is medium-independent because I can take away the wood and put in aluminum…'

DENNETT: Try putting in water.

SMITH: Right, you can't make a leg out of water.

HAUGELAND: You can't make your computer out of water, either.

SMITH: Right. It is all a little gray. Absolute medium-independence won't work, as if it didn't make any difference what you build it out of. We all agree with that. And pure medium-dependence doesn't work either, as if it had to be that of this specific set of electrical components, or that specific piece of protein. What we need is an appropriate 'middling' level of dependence and independence.

This gets to a point you've made, John, about the importance of engineering. I think our engineering practices have very refined intuitions about what kinds of properties materials need to have, in order for what we are building to work. Anything that has those properties will serve.[45]

If you want to calculate p, then a wide variety of materials will work. If you want something that will run at gigahertz, the range is smaller. If you want something to hold up a table, the range is (perhaps) more constrained yet. My prediction is that this whole terrain will eventually be mapped. And as it is mapped—this is really my claim—the maps will tie together physics as we know it today, which is ultimately concrete, and the (alleged) 'theory of computability' as we know it today, which looks very abstract.

CLAPIN: Why isn't that logic? Why isn't that properly logic?

SMITH: Because they are merely constraints on physics.

CLAPIN: But one way to think of logic is precisely as a description of how to just set up the physics the right way. This was the insight of computation: to turn syntax into something physical. So it seems to me that there is this match between logic and physics and the kind of representational redescription you're talking about, when you use a new representational code…

SMITH: What matters about logic is that there are physically-realizable syntactic configurations that you can interpret, in such a way that the effective transitions end up being semantics-preserving. It was that honoring of the semantic that I take to be the fundamental insight of logic. What happened, historically, I believe, is that computer science borrowed all the theoretic apparatus of logic-including terminology that had been developed in order to talk about semantics-honoring transitions, but then, in a deep way, forgot about the issue of honoring semantics. They took vocabulary that comes from a tradition that was interested in things like proof. But they used that vocabulary to study issues that are really about pure (uninterpreted) mechanism. For example, think of what is called 'denotational semantics' in computer science. Obviously, the word 'semantics' occurs in that label; you might think it would have to do with meaning. But what I believe this phrase really refers to, in computer science, is the relationship, given some machine, between effective arrangements that can be given to that machine as an input (called the 'program'), and a mathematical model of the behavior that results, when the machine is started up on that input. That's not semantics! Note, in particular, that you can construct such a mathematical model for any piece of machinery whatsoever. I can construct a denotational semantics of a can-opener, for example, or for arbitrary mechanisms built out of Meccano.

DENNETT: So it's logic gates…

SMITH: Yes, except the 'logic' part is gone! See, I think the theory of effective computability has a wonderful first name. It is a theory of the effective. C'est tout! It is the second part of the name that is problematic: the computing stuff, which I think has to do with issues of interpretation. Real world computing, I firmly believe, is drenched in genuinely semantic issues. But the theory of computability, the body of work we teach in computer science departments, fails to deal with these semantic issues-issues that actually constitute the practice.

HAUGELAND: You say that in the history of computer science and the portion that has to do with the 'computability issues' and so on, took over the apparatus that went with the problem of the formalization of logic and issues of proof and so on. But in fact that wasn't really part of their topic, right?

CLAPIN: Part of whose topic?

HAUGELAND: The computer scientists'. That, actually, semantic issues weren't the real issues that were being solved with the computability theory. You're not dismayed by the fact that the semantics fell away. If you're dismayed by something it's that they didn't realize that the semantics fell away and kept using the words.

SMITH: That's right. Because it's hellishly hard to tell them that semantics matters, when they are already using the word for something else! This all gets back to the points made in the '100 billion lines of C++' paper. What computer science uses the word 'semantics' for is the relationship between a static program and the dynamic process it engenders.[46]

HAUGELAND: As in Scott semantics?

SMITH: Yes, as in denotational semantics, more generally, of which Scott semantics is a type.[47] In fact denotational semantics and operational semantics are two characterizations of the same relation (one abstract, one more concrete-the relation between a program and the behavior it engenders.

DENNETT: This is Allan Newell and what I call my Julie Christie problem.[48]

SMITH: Yes, absolutely. So you've got to find a use for it, and all this apparatus and stuff. Why are they interested in Martin-Löf [1984] and the intuitionists and so on and so forth? The fact is computer scientists aren't radical hyper-intuitionists like Yessenin-Volpin [1970]. All that's going on is that they are studying an intrinsically effective subject matter. The relation between a program and the behavior it produces must be effective. Of course, if you want to study that relationship, and use mathematics to do so, you will be interested in mathematics that concentrates on what is effective. With just a bit of detachment, in other words, and a bit of historical perspective, you can understand why the mathematics went in the direction it did.[49]

CUMMINS: A footnote about pointers. Pointers are, in a way, the kind of parade case of embedded intenders, or nested intenders. Because the example you had, in a sense, the low-level intender got fooled because, as it were, the world switched targets on it. The problem was that there's a higher level intender in which it was nested to get something else.

SMITH: It's more complicated than that because it was a copy of the pointer that got fooled.

CUMMINS: Oh yeah, it is more complicated, but do you see what I mean about the nesting?

SMITH: That's right. The interaction between that and the object identity…and I'm just saying we're assuming object identity. It blows fuses in your brain. It's amazing stuff.

HAUGELAND: The world, or taking the world as mattering, is ultimately what matters?

SMITH: The former: the world is what matters.

HAUGELAND: Either could be an intelligible claim.

SMITH: That's right, but I think the latter is derivative from the former. It is the world as a whole that matters. Understanding that the world matters—taking the world to matter—that stance matters, too. But it is a subsidiary normative condition, a condition on what it is to be human (or perhaps 'humane'). So taking the world to matter is what ultimately matters about you. That's the sort of humility built into this Brentano-esque form of being oriented. The thing that matters most about you is that you recognize that the world matters more than you do.

CUMMINS: One of the disputes between myself and Millikan on functions all along has been that I wanted to say she can't understand how evolution works unless she can first identify the functions, and she thinks it's the other way around. I think this is a similar kind of thing.

SMITH: Right; I think it's a similar point.

CLARK: Strong enough to be the norms, or to give rise to them?

DENNETT: Give rise to them, yes.

SMITH: Actually I am just about to talk about that.

CLARK: But in between is the question, what makes sense of them-what makes sense of the norms. And there you really want to say that evolution doesn't make sense of the norms. On the other hand it's part of what it takes to make sense of them.

HAUGELAND: 'Make sense of them' means 'that in terms of which we can understand them'?

CLARK: Yeah. It's part of that in terms of which we can understand them.

CUMMINS: I'm with Dan on this, in that I think that just like species, what evolution stumbled on, the 'it' that it stumbled on is something that is constituted by the fact that it's the endpoint of that branch on the tree, and nothing else. And so you said we want to understand registration such that we can figure out why it was a good thing to stumble on, and I think actually that's a really kind of misleading way to put it. Evolution didn't stumble on registration. It built it out of smaller things, and there wasn't any such thing until it got it, because to be a registration is just to be the endpoint of that branch, developmental branch. I don't really believe this, but I've got to the point where I know how to say it in a way that sounds so plausible that I have a hard time resisting it.

HAUGELAND: I'm with Brian.

DENNETT: When you described this wonderful growth of registration and all those wonderful things, you presupposed a 'we' that wanted to do this. The we for whom this was the obvious product. And one wants to know, OK, help yourself, there's an agent. There's an agent with goals, there's an agent with purposes, there's an agent that is trying to find out more about the world. Where'd that come from? Now-

SMITH: No, look. Let me try to explain. You may be right that I often speak as if I were presupposing the existence of an agent, but you're misinterpreting me. And, remember, I do think that evolution is what did the work-I wish that I had that statement chiseled right here on this table, so that you're not tempted to think that I don't believe it. But given that, here is what I want to say. Suppose your 12-year-old grandchild and I are flying along in our spacecraft, and we notice a planet, which has plants on it. I say, 'Wow! Evolution is happening, all over again!' He says, 'Hey, can we speed it up?' I say, 'Yes, you know, there is something-a Really Neat Trick-that evolution on this planet doesn't seem to have been stumbled on yet. This trick involves organisms playing games with their internal structure so that they can track stuff that they are not physically coupled to. It might take another 100 million years for that to happen. Why don't we just drop down and do a little genetic engineering, give evolution a shove?'

In principle, this is something that could be done. Registration is a way of being that works. I don't think it is just the endpoint of some evolutionary branch. The world is such that registration is a way of being that is powerful.

And it would be powerful, whether evolution found it or not.

DENNETT: What's fascinating to me about that is that now you're playing the card which I intend to play, usually, and get hammered on by people like Dick Lewontin who says don't, don't, don't think of evolution as these problems that are posed and then are solved by evolution-these sort of Platonic problems that are solved. And I'm very happy to be this sort of minimal Platonist. You know, there really are these problems independent of history that are posed, or could be posed. We can conceive of them being posed again and again and again across the cosmos, and solve them in the same way. That's what a Good Trick is. And I think that's fine and I think that's quite consistent with evolutionary theory. It is of course an idealization. There's danger of inscription errors. But it's a deep way to think about evolution. And when people like Gould and Lewontin chastise themselves and their fellow theorists for doing it, it is at the cost-and this is sometimes glaring and even to the point of being comic—that they can't even talk about convergent evolution. That's why convergent evolution is a sort of non-topic for Gould—it's bizarre—and for Lewontin. The reason it's a non-topic for them is because they can't let themselves talk about the same solution to the same problem they reinvented. But I think it's a deep part of evolutionary theory to be able to separate—just as you say, to separate the discovery from the process.

CLARK: But I think that one thing that you don't want to do-and this kind of fits in with something Joan [Wellman] was saying last night—is to get forced into a discussion about where normativity comes from when all that really matters, for most of John's projects, is what it's like when it gets there. Yet understanding where you are actually isn't, I suggest, independent of understanding how you got there. When you want to understand where you are, you want to understand where you're likely to go. And the kind of processes that got us there are still, one way or another, operative.

SMITH: Remember, I didn't banish Dan to the other side of the town; I took him as a flying buttress; that's right. There's something extremely important about understanding how we got here.

WELLMAN: Brian, that's the second thing you said in your list of three last night to me. You said it matters—the way that something is implemented matters to the way it can be.

SMITH: Yes, I think that's right. I'm all in favor of understanding history and implementation. At the risk of agreeing with you all so much that I sound wimpy, there is even a reading of 'rides on the coat-tails' that I can agree with: that evolution is the train it took-a train made of coat-tails. Our normativity, our registrational prowess; they all came via that route.[50]

One way to understand this is in terms of the design space. That's one thing that being a computer scientist teaches you: to be interested in the entire space, the whole fitness landscape, not just in the structure of a particular solution. Maybe on the other side of some great canyon in the fitness landscape there lies an enormously powerful and possible way of solving some of the problems you face-but evolution will never find it, because it is too bloody expensive to go in that direction. No creature could survive an attempt to cross that canyon.

So just as I think how you build something is really important, but not necessarily constitutive, I also think that understanding what's constitutive, and what the space of possibilities is and so on, has got to be helpful in terms of how the evolutionary story went. So I'm all in favor of this handshake with evolutionary theorists. I just don't think that norms and mattering rest constitutively on evolution.

DENNETT: Then I think we agree.

CLARK: I don't see any reason to deny that.

CUMMINS: I'm not sure I do.

SMITH: I don't believe there's complete agreement, but I do think there's a kind of…

DENNETT: I think there's still some tension here…

HAUGELAND: There's a question which I've been sitting here trying to formulate, and I'm not sure how it bears, but I have this inkling that it does. And that is, we were talking about Good Tricks and having a phrase like, 'it's a Good Trick in one's kit'. But there's some question as to how these are individuated, what makes them a trick and what makes them good. And then how they could be brought about; whether evolution is the only way. For instance, why isn't it a Good Trick to produce planets composed of heavy elements; to get a whole lot of hydrogen to attract itself to the point where it then fuses into helium and then collapses into a dwarf where the helium fuses-then it explodes and those pull together and then you've got a planet. That's a pretty tricky process, right?

DENNETT: Read Lee Smolin [1997] and you'll see somebody who says, add it right in there, another Good Trick. That's what evolutionary cosmology is about.

HAUGELAND: There's no selection there.

DENNETT: Oh, there is for Smolin. I'm agnostic about Smolin's cosmology, but I think it is a not provably incoherent cosmology that simply embraces what you're trying to do as a reductio, and says no, look, we actually do have an evolution with selection of whole universes, that is, whole ways of having the basic so-called constants of physics. And some of them produce…

HAUGELAND: Yeah, but you've changed the subject. I'm talking about this universe with our constants of physics-

DENNETT: It is a Good Trick and can be seen to be a Good Trick, but, as usual, you never see it against the background of the failed universes that don't happen.

SMITH: What about saying 'Look, why don't I just inhale some dioxin and decompose…'

CLAPIN: Why is that a bad trick?

SMITH: Yeah, what's the 'good'-

CUMMINS: Well, you have to have replication in the picture. Replication and selection wasn't in the picture that John told. It is in the story that Smolin told.

DENNETT: It is in Smolin's story.

CUMMINS: I don't believe it for a minute. I don't disbelieve it, either. I just don't get it.

DENNETT: But my point is, you've asked the right question, but don't presume that there isn't an answer to it.

HAUGELAND: Well, the question is not to Smolin, whom I've never heard of, but rather to you, who don't believe in the alternative universes, let alone them procreating and competing. Just to you, believing, like we Weinbergians or whatever, that it started back there in big bang and it's been buzzing along ever since. Well there it's been evolution in the sense of the trajectories but not evolution in the sense of selective pressures. And yet there are things which it's not obvious why you wouldn't call them Good Tricks. Unless you build selectiveness into the notion of good thing, and that's a suspicious move.

DENNETT: Well, I've got a long answer to it, but I don't have a short answer to it, and it's time for lunch. But you can read the long answer because it's in my book *Darwin's Dangerous Idea*.

— Were this page blank, that would have been unintentional —

C • Representation

— Were this page blank, that would have been unintentional —

7 — Rehabilitating Representation

1 Introduction

No concept has played a more important role in cognitive science than that of representation. The classic model of mind on which the field was founded was representationalist to the core, due to the second of its two founding assumptions. First, as reflected in its name,[1] cognitive science took intelligence to be epitomised by individual, rational, deliberative thought (in roughly Cartesian spirit, even though dualism was soundly rejected). Second, cognition so conceived was taken to consist in the formal manipulation of explicit, composite, language-like *representational* structures—rather as in logic. Although that founding view is often called 'computational,' for a variety of reasons I believe that that name is misleading,[2] so I will instead refer to it as **logicist**.

After being celebrated for many years, representation has recently suffered quite a drubbing. Challenges have been mounted on all sides—philosophical, neurophysiological, anthropological, and dynamicist. In its place, a spate of new views have been proposed, ranging from low-level neuronal models of brain function through autonomously navigating vehicular robots to high-level vaguely Heideggerian accounts of practice and sociality.[3] Though different in style and substance, these counterproposals are alike in one critical respect: they all recommend that the classic model be rejected in favour of a variety of more dynamic, embodied alternatives. Because of this common rejection of the classical

[1] I.e., as opposed to having been called 'the study of intelligence,' or some other moniker giving cognition less centrality.

[2] Smith, Brian Cantwell, 'One Hundred Billion Lines of C++', «ref».

[3] «Refs»

model, these otherwise rather disparate alternatives are often loosely grouped together—originally under the label 'situated cognition,'[4] but more recently, perhaps because it more clearly incorporates neuroscience along with the other suggestions, under the label I will use here, of **embodied cognition**.[5]

The embodied alternatives tend to subject both founding assumptions to critique. First, instead of accepting as the premise that intelligence paradigmatically consists of cognition conceived as individual ratiocination, these views tend to privilege improvisational response and real-world (and, to varying extents, social) interaction—rather on the model of navigation. Second, there have been tendencies for all camps, each in its own way, to argue that the implementing mechanisms of this improvisational behaviour must be *nonrepresentational*.

As regards its status as socio-intellectual history—i.e., in terms of dominant rhetoric, prevailing assumption, and overall disciplinary profile—the shift from detached abstract reasoning to engaged material participation has largely been won. No one any longer denies the importance of context-dependence, of real-world interaction, of concrete embodiment. In fact contemporary students are likely to view favoring a logicist or 'formal symbol manipulation' view of mind to be as retrograde as holding a positive attitude towards pure introspectionism or Skinnerian behaviourism.

Independent of the merits of adopting a situated or embodied approach, however, it is not clear whether the wholesale embrace of **antirepresentationalism** may not ultimately prove as much of a straight-jacket as the original overly-zealous embrace of (especially 'formal') representationalism. If uncritically embraced as stand-alone directives, after all, even the most salutary correctives may lead down paths that miss their target as much as the views they were originally introduced to modify. Thus suppose Δ in figure 1 is introduced with the intention of shifting the target of mainstream inquiry (M) away from a and closer to b. If Δ is pas-

---

[4]The session at the workshop during which this paper was first presented was entitled "*Intentionality and Situated Cognition*."

[5]Smith, Brian Cantwell, 'Situatedness/Embeddedness', Wilson & Keil (ed), *MIT Encyclopedia of the Cognitive Sciences (MITECS)*, Cambridge: MIT Press, 2001, pp. ■■–■■.

sionately embraced as a research path in its own right, instead of being recognized as an adjustment to M, it is likely to lead to a′—as far or farther from the desired b as the original a. And then if another correction Δ′ to Δ is introduced in turn, the whole process may repeat, causing inquiry to proceed in a haphazard way. (Fundamentalism on the left is as untenable as fundamentalism on the right.)

Something more straightforward is needed.

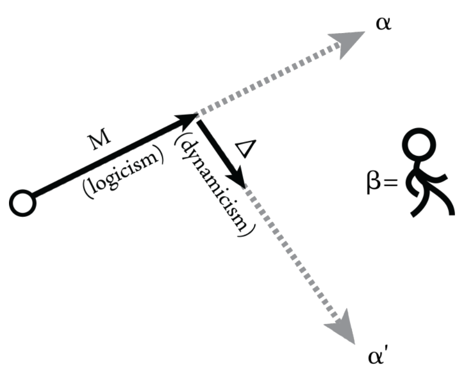


Figure 1 — Ideology in Cognitive Science

Perhaps the weakest suggestion is for a hybrid or **amalgamated** view: use non-representation wherever and whenever it works (empirically, pragmatically, theoretically), and then add in representation wherever it is appropriate or needed to handle "more complex" cases. Something of this sort is suggested in Clark's *Being There*,[6] and is advocated by as staunch an anti-classicist as Rod Brooks.[7] But no matter how commendably balanced, on its own that strategy is a bit vapid. Sure enough, as Braitenberg, neo-Gibsonians, and others have emphasized, non-representational mechanisms are capable of producing vastly more complex and subtle behaviours than classicists ever imagined.[8] But a simple amalgamation strategy doesn't answer any of the constitutive questions: when or why representation might be needed, what contributions it may (uniquely?) be capable of supplying, when it is *not* required or advisable, etc.—to say nothing of what the powers and limits might be of pure mechanism or pure embodied behaviour.

---

[6]Clark, Andy, *Being There*, Cambridge: MIT Press, 1998.
[7]Rodney A Brooks, 'Intelligence Without Representation,' John Haugeland, ed., *Mind Design II*, Cambridge: MIT Press, 1997, pp. 395–420.
[8]The point was made as early as in Herb Simon's *The Architecture of Complexity*. Cambridge, MA: MIT Press, 1969. See also Braitenberg, Valentino, *Vehicles: Experiments in Synthetic Psychology*, Cambridge: MIT Press, 1986; other «refs».

Moreover, to assume that the two traditions can be glued together without alteration—as if in an assembly—is a bit of a dream. The suggestion also fails to illuminate the question of what kind of representation would best suit a combinatorial approach (abstract, formal, logical, imagistic, etc.); nor does it say anything about what should play the role of anti- or nonrepresentational complement (physical dynamics, existential thrownness, etc.).

We need to cut deeper.

A more powerful idea is suggested in figure 2: what I will call a **generalisation** strategy. Rather than assume that logic encapsulates the essence of what it is to be representational, the suggestion is to recognise representation as an (at least potentially) richer and more encompassing notion in its own right, and then to identify (and perhaps criticize) the logicist variety as just one particular species. Among other merits, this approach has the virtue of not "giving away" the notion of representation to the predecessor view, as if logicians somehow understood representation's be-all and end-all.

From an intellectual point of view, the generalisation strategy requires dissecting the traditional conception of representation into two parts: (i) what is universal about representation in general, applicable to all species; and (ii) what is specific to the particular form of representation embodied in the classical view. In practice the strategy has rarely been approached so theoretically. It has instead proceeded in a more "bottom-up" way, through numerous attempts to identify other (allegedly non-logical) species of representation. Of many suggestions, perhaps three are most famous: (i) *imagistic, iconic, pictorial* or *visual* representation—a perspective from which logicist representation is viewed as fundamentally *linguistic* or *propositional*;[9] (ii) *procedural* representation—in contrast to the

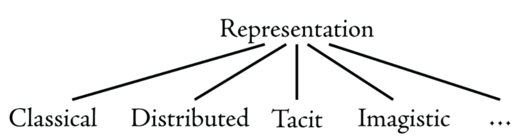


Figure 2 — Species of Representation

[9]«Ref Kosslyn, Shepherd, and others»

presumptively *declarative* character of representation in logic;[10] and especially since the rise in popularity of connectionist and other network models, (iii) *distributed* representation—as opposed to what was most often in these debates simply called *classical*.[11]

Since they were largely operating within a representational context, the defenders of these alternatives tended to concentrate on bringing particularities of specific cases into focus (logicist and other), rather than addressing the overarching issue of representation in general. The result was to leave us without much of an understanding of how representation might or should be combine (in an intelligent agent) with more direct forms of dynamics or embodiment. But there were a number of significant exceptions, perhaps especially including John Haugeland, who not only attempted to compare and contrast what he calls *logical*, *iconic*, and *distributed* "genera", but who also made some remarks about the general case.[12] Strikingly, he introduced his paper as "even more than usually tentative and exploratory"; called its results "at best preliminary and incomplete, perhaps much worse";[13] and took up his discussion of representation-in-general with yet an additional caveat:[14]

> "An explicit account of *representation as such* will not be necessary; that is, we can get along without a prior definition of the 'family' within which the genera are to be distinguished. A few sketchy and dogmatic remarks, however, may provide some useful orientation, as well as places to hang some terminological stipulations."

[10]'«Ref Winograd and others»
[11]«Ref the PDP volumes, the Smolensky/Fodor debates, etc.»
[12]Haugeland, John, "Representational Genera," in W. Ramsey, S. Stich & D. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Hillsdale: Lawrence Erlbaum, 1991, pp. 61–89. Reprinted as ch. 8 in Haugeland, John, *Having Thought*, Cambridge, MA: Harvard Univ. Press, 1998, pp. 171–206.
[13]op. cit.; p. 172.
[14]op. cit.; loc. cit.; emphasis added.

Yet in spite of his cautionary remarks, the three paragraphs that Haugeland devotes to the topic not only contain substantial insight, but are also so widely cited as cognitive science's best characterisation of representation that I have taken the liberty of reproducing them here, to serve as something of a starting point (sidebar, p. ■■).

In order to give the generalization strategy as much play as possible, the characteristics of logical representation identified by advocates of alternative species—i.e., its being *linguistic* or *propositional*, *declarative*, *systematic* and *productive*, etc.—can be added to

Haugeland on Representation†

"A sophisticated system (organism) designed (evolved) to maximize some end (such as survival) must in general adjust its behavior to specific features, structures, or configurations of its environment in ways that could not have been fully prearranged in its design. If the relevant features are reliably present and manifest to the system (via some signal) whenever the adjustments must be made, then they need not be represented. Thus, plants that track the sun with their leaves needn't represent it or its position, because the tracking can be guided directly by the sun itself. But if the relevant features are not always present (manifest), then they can, at least in some cases, be represented; that is, something else can stand in for them, with the power to guide behavior in their stead. That which stands in for something else in this way is a *representation*; that which it stands in for is its *content*;[a] and its standing in for that content is *representing* it.

"As so far described, 'standing in for' could be quite inflexible and ad hoc; for instance, triggered gastric juices might keep a primitive predator on the prowl, even when it momentarily loses a scent—thus standing in for the scent. Here, however, we will reserve the term 'representation' for those stand-ins that function in virtue of a general *representational scheme* such that: (i) a variety of possible contents can be represented by a corresponding variety of possible representations; (ii) what any given representation (item, pattern, state, event, …) represents is determined in some consistent or systematic way by the scheme;[b] and (iii) there are proper (and improper) ways of producing, maintaining, modifying, and/or using the various representations under various environmental and other conditions. (This characterization is intended to be neutral not only among genera, but

†Haugeland, John, "Representational Genera," pp, 172–73. Emphases and notes in the original.

the logicist characteristics most often criticized by the anti-representationalists: the fact that logic is allegedly *explicit*, *formal*, *context-independent*, *static*, and *abstract*; the fact that it emphasizes *reference, rationality*, and *truth* over other semantic properties and norms deemed by some to be more appropriate to pragmatic intelligent conduct; etc. Unless any of these properties can be defended as constitutive of representation itself, the generalist would want (i) to forge a notion of representation that is not committed to them, and then (ii), if they are not only inessential to representation as a whole but also inappropriate for the full

also between internal and external representations, and between natural and artificial schemes.)

"Since the content of a given representation is determined by its scheme (and since the point of the facility is to be able to represent what isn't present or currently accessible), it is possible for representations to misrepresent. What this amounts to will vary with the specific scheme, and even more with its genus; but it must hark back eventually to the possibility of the system(s) using it being *misguided* in their attempted adjustments to the features of the world. But misrepresentation should not be confused with improper deployment on the part of the using system, nor bad luck in the results. These can diverge in virtue of the fundamental holism underlying what can count as a representation at all: the scheme must be such that, properly produced and used, its representations will, under normal conditions, guide the system successfully, on the whole. In case conditions are, in one way or another, not normal, however, then a representing system can misrepresent without in any way malfunctioning."

a. This use of the term 'content' is not altogether standard. Most contemporary authors (and I, in the other essays in this volume) mean by the "content" of a representation something distinct from the object it represents, and which determines that object (as sense determines referent, for instance). Here, however, I mean by 'content' that which the representation represents—the "object" itself—but as it is represented to be (whether it is that way or not). Thus, it is a possible object—which may in fact be actual, or similar to something actual, or neither. [Note added 1997.]

b. For instance, if (or to the extent that) particular representations are tokens of well-defined types, the scheme will determine the content of any given token as a function of its type—or, at least, these will determine how that content is determined. Thus, if any extra-schematic factors (such as situation or context) co-determine contents, then which factors these are and how they work are themselves determined by the scheme and type.

range of cognitive behaviours, to identify other species that, while still genuinely representational, do not exhibit those specific characteristics of logicism.

Though not explicitly described in these terms, support for such a generalising approach can be found in a flurry of recent discussions of representation in the philosophy of mind.[15] The strategy has also had the benefit of leading cognitive scientists to read in areas of philosophy beyond logic and the (relatively narrow) classical "Language of Thought" school of philosophy of mind—e.g., to look to Ryle, Merleau-Ponty, James, Heidegger, Dewey, Langer, etc., for inspiration.[16]

Note too that generalisation can easily be added to amalgamation in a combined hybrid strategy. There is no need to insist that a representation, even appropriately generalised, must apply to *all* aspects of human cognition. The point is just to make room for the possibility that some (or perhaps even many) aspects of intelligent behaviour may require some notion of representation for their proper explanation—i.e., to recognise that there may be aspects of cognitive behaviour that cannot be accounted for by (for example) a purely dynamical approach, even if they do not fit into the classical "logicist" framework.

Read this way, the generalisation strategy has much to recommend it, and in many ways I will adopt it here. But it, too, especially by itself, does not cut deep enough.

In this paper I will argue for a third approach—something I will call a **reconstructionist** strategy. The (admittedly *ex post facto*) argument for reconstruction runs roughly as follows:

1. It is true that the classical model is too specific (narrower) than is required or appropriate for many of cognitive science's purposes.
2. It may also be true that some (even constitutive) aspects of a person's overall cognitive processes may be nonrepresentational—as suggested in the amalgamationist strategy.

[15] «Ref Cummins, Chemero, Clark and Grush, etc.»
[16] «Refs»

3. Independent of the merits of (2), however, it is also true that the logicist model of representation is narrower than representation per se requires, and so the logicist approach to representation should be generalised, and new non-logicist species of representation identified and explored—as recommended in the generalisation strategy.
4. But something stands in the way of our doing this generalisation.
5. Although the classical model was based on some very deep insights into the nature of representation,
6. Those insights were expressed in ways that were not only too narrow, but in addition outright misleading—i.e., not just false of representation in general, because restricted to the circumstances (and expressed in the language) of the specific view, but inadequately understood *even in that restricted (classical) case*.
7. What we need, therefore, is not just to generalise, but to reconstruct, the classical view: reframe and rephrase it, re-understand its essential features.

For a simpler but different example of reconstruction, to see the strategy in action, consider a case I will talk more about below: the constraints of "computational effectiveness" that lie at the very basis of logic and computer science. There is no more important conceptual ingredient in the classical view than this notion of what can be algorithmically or mechanically done.[17] For various reasons, as we will see, these effectiveness constraints, even in syntactic guise, have classically been viewed as *mathematical* and *abstract*. What I will argue is that even in classical settings, and in spite of the character of classical analysis, they are not, in point of fact, abstract after all, but instead are direct (if implicit) consequences of the material character of the underlying computational substrate. They have been *understood* as abstract, but classical understanding is wrong. In point of fact they are *concrete*.

This is an example of reconstruction, not generalization, because I am not claiming: (i) that effectiveness can be legitimately

[17]The term 'effective' is inscribed in the foundations of computer science: its core theory is called a theory of *effective computability*.

understood as abstract in the classical case—i.e., in situations where formal logical explicit representation or inference is mandated; but (ii) must be understood as concrete (material, physical) in more general situations—e.g., those involving non-classical forms of computational and/or representational activity. Rather, I am making the stronger claim that even *in paradigmatic cases of first-order logical inference*, the operative constraints on "what can be done" (what can be proved, what can be inferred, what can be mechanized, what can be computed) are and always have been ultimately physical, even if they have not classically been understood in that way. In others words, the reigning theoretical presumption that effectiveness and computability are appropriately understood abstractly or syntactically isn't too *narrow*. It is *false*.

In what follows I will to varying degrees adopt all three strategies—amalgamation, generalization, reconstruction—but in reverse order:

8. First we need to reconstruct the classical view, which among other things will allow us to see, in some depth, what was particular about the classical view, and what circumstances if any recommend its use;
9. Then we will be able to generalise the notion of representation appropriately, formulating a more powerful, encompassing replacement;
10. Then—and only then, with a generalised notion in hand—we can address the question underlying the first amalgamationist strategy: of which aspects of cognition do, and which aspects do not, need to be understood in representational terms;
11. Once that is in hand as well, we will have arrived at a possible *substrate* for a comprehensive account of mind.[18]

[18]It is no theory of mind; that would be something vastly more ambitious. I call it a possible substrate only in the hope that, by diagnosing the relation between physicality and concrete representation, it may supply conceptual terms in terms in terms of which a successful theory of mind might be formulated.

Three final preparatory comments.

First, it is ironic that representation has been misunderstood on *both* sides of cognitive science's pro- and antirepresentationalist debate, blocking substantive progress. But although it helps to point this out, my aims are not ultimately critical. Rather, what I want to figure out is how to be *positive about both sides at once*—i.e., how to do justice to the intuitions underlying each. The issue is not merely rhetorical or motivational—or even socio-intellectual, where (as mentioned) the issues are largely settled. Like most modern writers, I approach cognition sympathetic to a renewed emphasis on embodiment, activity, and practical "being in the world," of the sort that motivates the embodied cognition movement. At the same time, however, I am concerned that many of the profound insights that underwrite the classical model (particularly, as we will see, semantical insights) are being lost, in the rush to embrace "in the world" concrete embodiment.

More pointedly—and in a sense this is the real aim of the paper—I worry that, in eschewing abstract formality in favour of concrete materiality, a spate of embodied cognition theories, from cognitive neuroscience to cultural theory, even if dressed in impeccable scientific credentials or urbane French garb, are unwittingly falling prey to a kind of *causal reductionism* or *causal fundamentalism* incapable of understanding what is ultimately distinctive about minds and mentality—having critically to do with semantic directedness. Put it this way: the most urgent challenge for embodied cognition, in my view, is to

> *Preserve—perhaps even rescue—semantics through a (beneficial) shift from abstractness to concreteness.*

It will take some work to see what this comes to. I will start with two critical reconstructions, followed by a dozen or so targeted generalisations. Once we have those in place, we will be able to start combining what matters about each side of this overly dichotomised debate into a unified and durable successor.[19]

[19]Interestingly, this recombinant reconstruction is necessary in order to achieve another goal of great importance: that we unify the understanding of representation that serves in technical fields (such as logic, computer science, linguistics, cognitive science, etc.) with understandings of representation in literature, the arts, and humanities.

Second, it is important to understand that the pro and antirepresentational debate in cognitive science falls on the **sub-personal** side of the *personal/sub-personal* distinction.[20] Assume that by 'subpersonal' I will refer to the mechanisms or ingredients out of which intelligent creatures are made, and by 'personal' will refer to the full-blooded intentional agents thereby physically constituted. It is the full person, that is, who is the subject of consciousness, the bearer of rights, the participant in social norms, the member of community. It is the subpersonal mechanisms that implement or realize persons with which cognitive science is primarily concerned. I would thus take as falling within the scope of the questions being addressed here debates about the representational character of the retinotopic map in areas V1 through V5 of the visual cortex, and debates about whether, in empathy, we *represent* the emotional lives of others, or do something more akin to *taking them on*. But I would not, by itself—at least not without comment and consideration—take a positive answer to either question to be evidence that, *as people* (i.e., at the personal level) we "represent" our environment or our friends in the course of our daily lives.

By making a personal/sub-personal distinction—by foreswearing an *identification* of people with the meronomic components of their bodies—I absolutely do not want to suggest that the two are *independent*. Neither do I want to endorse claims, such as those of McDowell, that attribution to ingredient mechanisms of such intentional characteristics as "being representational" is merely "as if".[21] On the contrary, I believe that the relation between the representational, semantic, intentional and/or normative character of "that of which we are made" and the representational, semantic, intentional, and normative character of "we who are thereby made" is extraordinarily vexed. In other contexts I will argue that co-constituting ties bind the authenticity of the full-blooded intentional involvement of persons in the world and the genuineness of the representational character of the material ingredients of the world in virtue of which are they are such full-blooded participants.

[20]«Refs; including McDowell's response to Dennett»
[21]«Ref»

But this is not the place to pursue such questions. Here I want simply to introduce a term that I have developed more fully elsewhere, which will help us to mark the personal/sub-personal distinction and to stay out of the notorious conceptual confusion that stems from ignoring it. In particular, *unless explicit comments are made suggesting otherwise*, I will say that, at the personal level, we **register** the world in terms of the objects, properties, situations, states of affairs, features, etc., that we thereby take it (the world, that is) to consist in. Thus as I write these sentences, as it happens, I *register a building across the street*, *register a lake on the horizon*, *register a thorny academic situation of which I have just learned as petty and unfortunate*, etc.

A few comments about the notion:

1. By 'registration' I intend to index the fact, shared by representation, that human thought, perception and understanding of the world is ineliminably 'as'.
2. I take the term to be neutral as to any distinction between or among *sense*, *perception*, *thought*, *judgment*, etc.
3. Unlike 'conceive' or 'cognize' (and in this respect more like 'see' and 'perceive'), I take 'register' to be a "success" verb. If, in ordinary circumstances, a person registers—i.e., takes there to be—a tree, then it is fair to assume that there was a tree there to be so taken, and that the person did so take it, in the full semantically and normatively appropriate way.[22]
4. The term 'register' is usefully neutral on the division of responsibility between person and world for the resulting ontological "take"—i.e., is neutral as between naïve realism (I successfully register a table *as* a table because it *is* a table), strong forms of constructivism (I successfully register it as a table because of the contingent and historical forces constituting the social community of which I am a member, or even due to the particular exigencies of my own in-

[22]This is a rather realist characterisation of "success"; it should be replaced as appropriate for other metaphysical views. The point is simply that "a registered b" should be true just in case something roughly of the form "There is b and a took it to be b" is true.

dividual history), idealism, solipsism, and many other epistemic, ontological, and metaphysical proposals.

5. As should be evident from the above (including occasional use of the qualifier 'successful'), I take registration to be normatively laden, in the philosophical sense of serving as the subject of such issues as truth, objectivity, worth, etc.
6. By following the verb's direct object with 'as …', the construction facilitates at least a first step towards distinguishing how we, as theoreticians (cognitive scientists, epistemologists, etc.), register situations or phenomena in the world, and how we take them to be registered by other subjects or people or agents, of whom we may be speaking. Thus I might say, of an infant, "She registers her mother's coming to the door not as the re-appearance of a recognized individual object, but more as "re" or "repeating" placement (in Strawson's sense) of the feature *Mama*."[23]
7. If *not* used with an explicit 'as' construction—i.e., notwithstanding (6)—I will assume that the direct object of 'register' to include the aspectual nature of the way in which that phenomenon or entity is registered by the (individual designated by) the sentence's subject. In this way 'register' differs from at least common uses of such perceptual verbs as 'see.' Thus while some would claim that it is possible for the sentence "Randy saw the Northern Lights" to be true even if Randy did not recognize them *as* the Northern Lights, I consider it an implication of the sentence "Randy registered the Northern Lights" (without any following 'as' clause) that Randy did so take them.[24]

In these terms, I would characterize the **representational theory of mind** as a theory that claims that *human cognition is underwritten by processes involving the manipulation or use of representational*

---

[23]It is only a first step, because of the evident but fraught issue of *how we register the subject's registration* (e.g., in the 'g' part of the sentence 'a registers b as g')—including whether we can, and if so how much, and in what respects.

«For the infant case, ref Jun's Duke dissertation.»

[24]I.e., the direct object position of the verb 'register' is thus not assumed to be referentially transparent.

ingredients. *Per se*, that is, the computational theory of mind does not mention registration; if it needs to explain registration (as I believe it must), then it must do so as a consequence of the substantive claim that registration is something that people do. I thus consider it to be a substantive question how much of human existence and/or participation in the world rests on registering it—as opposed, say, simply to bumping into it, or responding as a purely physical or mechanical device. Similarly for two questions to which in the long run the present investigation is likely to be primarily relevant: (i) how much human registrational capacity is underwritten by representations—either internal, external (as suggested in discussions of scaffolding etc.), communicative, etc.;[25] and (ii) conversely, whether representations are implicated as ingredient mechanisms to underwrite human capacities that, at the personal level, do not involve registration.

For now, it is enough to say that when speaking at the personal level, of whole human beings, I will speak of *registration*, unless (in which case it will be explicitly marked) it is genuinely personal-level representation that is at issue, as for example might arise in a discussion of parliamentary democracy. Except in such marked cases, however, uses of 'representation' will refer to entities that are constitutive, realizing, ancillary, external, supportive, communicative, or otherwise implicated in the world that we as persons inhabit.

Third and finally, although the paper is entitled "Rehabilitating Representation," it is the *notion* of representation I aim to renew and refurbish, not the representational theory of mind.[26] Indeed, it is no part of my purpose here to argue for or against such a representational theory. My concern is only that representation is a more powerful notion than recent treatments would lead one to suspect. The question of whether the mind is representational strikes me as both substantial and open—a question to which we are as yet far from knowing the answer.

---

[25]«Ref Clark and others»

[26]If renovating concepts ruffles your ontological or epistemological feathers, take this as elliptical for refurbishing discourse that makes substantial use of the concept.

2 Logic

The concretization of effectiveness described in §1 is just the first step in our reconstruction of the classic logicist view. Others steps have to do with semantics, formality, and the structure of norms. To understand any of them, we need a clear grasp of the conceptual (though not technical) structure of logic—the aim of this section. I will assume a modest working familiarity with basic logical notions, of the sort presumed throughout cognitive science; my aim here is simply to clarify some of logic's underlying conceptual framing.

In particular, I will proceed in two steps: (i) describing an a-temporal or static logical or representational basis, and (ii) a computational increment, introducing the notion of process.

2a Logical basis

As diagrammed in figure 3, the classic logical picture consists of five ingredients, grouped into three kinds:

1. Two realms: one syntactic (*S*) and one semantic (*D*);
2. Two relations, one on each realm: a "proof-theoretic" derivability relation *P* on *S*,[27] and a real-world or domain-theoretic entailment or dependence relation *R* on *D*;[28] and
3. A semantic interpretation function (*I*) from *S* to *D*.

The syntactic realm *S* consists of the representations themselves—

[27]Often written as an infix '$\vdash$', as in '$S_1, S_2 \ldots S_i \vdash S_k$'

[28]Entailment ('$\models$') is usually understood as a relation on *S*, as in '$S_1 \ldots S_i \models S_k$', or as a relation among elements (or sets of elements) of *D* and a sentence S_i, as in '$D_i \models S_k$' (in for example a case where D_i was a possible world in which S_k is true). What I mean by saying that entailment (*R*) is defined on *D* is that, however it is formally defined, entailment ultimately rests on a relation *R* defined among elements of *D*, to which it relates, in any sense in which sentences are involved (except self-reference) through *I*. It is *R*, the relation among elements of *D*, that is, that "wears the trousers" as regards entailment. For example, suppose one says that that S_1 (in *S*) entails S_2 (in *S*)—i.e., that $S_1 \models S_2$. That would be true just in case the interpretation $I(S_1)$ *bears R* to the interpretation $I(S_2)$.

In a standard extensional model of first-order logic, *R* would be something like inclusion, where *I* maps sentences onto sets of models in which *S* is *true* (i.e., so that $S_1 \models S_2$ just in case $I(S_1) \subset I(S_2)$).

structures it is in general convenient to call **representational vehicles**: typically, expressions in a formal language. Paradigmatically, elements of *S* are formed from a finitely-specified set of atomic elements or *simplexes* (variables, constants, predicate and relation symbols, quantifiers, various sentential operators, and so on), assembled in a variety of ways into well-formed *complexes* (sentences, quantified terms, etc.) via inductively-specified syntactic composition rules, in the familiar way. Classical knowledge representation schemes in artificial intelligence (which I am including in the general "logicist" camp) introduced a bevy of structural and aesthetic variations. Various properties specific to *written* languages, for example, having to do with one-dimensional lexical syntax (including for example the idea of *named* variables) were set aside in favour of a more abstract conception of representational structure, leading to proposals that more closely resembled graphs, or even abstract structures from non-well-founded set theory.

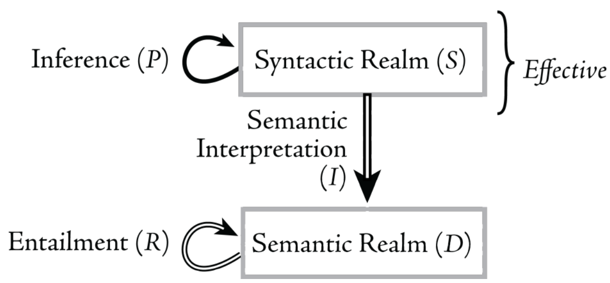


Figure 3 — The Conceptual Structure of Logic

Details do not matter here, though, since our aims are conceptual rather than technical. It is enough to note that in classical models, syntactic structures (representational vehicles), both simple and complex, are assumed both to be definable and identifiable on their own—i.e., to have determinate, autonomous identity conditions, independent of and explanatorily prior either to the semantic realm, or to the two relations (derivability or interpretation).

The semantic realm *D*, on the classical view, is normally treated model theoretically, and in that guise taken to be *abstract*. Again, elements of the semantic realm are also (usually) taken to be discrete and determinate, with ontologically and explanatorily autonomous identity conditions. In a typical case, the semantic realm *D* would be assumed to consist of a (possibly set-theoretic) domain or structure of objects, properties, relations, functions, etc.—again, in a wholly familiar way.

The proof theoretic or inferential relation *P* is defined over the syntactic realm *S*. *Legitimate* inference relations *P* (out of the space of all possible *P*s) are identified in virtue of various *semantic* constraints on *P*, defined in terms of *I* and *R*, as we will see. It is they, ultimately, that give *P* its main substance. But before semantics is allowed to get a a toe-hold, *P* must satisfy three critical conceptual well-formedness conditions.

1. *P* must be definable over the *formal* or *syntactic* properties of the representational vehicles (i.e., the elements of *S*).

Stunningly, what it is to be a formal or syntactic property isn't entirely clear;[29] what it is to be a syntactic property is rarely theorized. Nevertheless, as made famous in cognitive science circles through Fodor's formulation of his **formality condition**, *form* or *syntax* is generally taken to have both a positive and a negative aspect (sidebar). Positively, it has to do with the grammatical or syntactic structure—namely, those properties, including the identity conditions, of the elements of *S* in terms of which *S* is defined; negatively—and this is critical—syntax is assumed *not to involve or make reference to any semantic properties*. Thus it would be malformed, because not formal, to define an inference relation that applied only to *those expressions that Jerry Fodor currently favours*, or to *those expressions that are true*.

2. In a computational or cognitive context, the derivability relation *P* must also be *effective*, in the sense of being able to carried out, or at least checked, "mechanically."

The rule "From expression s_1 derive the constants '*T*' or '*F*' depending, respectively, on whether, a hundred years from now, S_1 will or will not have appeared more often in published logic textbooks" is adequately formal by the first criterion, but fails to be effective by the second.

[29]Since logics are usually introduced individually, by ostension, the syntactic properties of a particular system are usually simply pointed out, and accepted, by-passing the requirement for a general account. But see below.

Fodor's Formality Condition†

"What makes syntactic operations a species of formal operations is that being syntactic is a way of *not* being semantic. Formal operations are the ones that are specified without reference to such semantic properties of representations as, for example, truth, reference, and meaning. Since we don't know how to complete this list (since, that is, we don't know what semantic properties there are), I see no responsible way of saying what, in general, formality amounts to. The notion of formality will thus have to remain intuitive and metaphoric, at least for present purposes: formal operations apply in terms of the, as it were, shapes of the objects in their domains."

†Fodor, Jerry, "Methodological Solipsism," «Ref»

3. In rather anti-Wittgensteinian spirit, syntactic properties, as well as having to be effective and non-semantical, are required, in logical settings, to be both syntactically and semantically defined *without regards to their use*.

Thus no room is made for defining a relation Q that is transitive so long as it is not used more than three times in a derivation.[30]

Given well-formed syntax and appropriate compositional rules, the interpretation function I is typically defined inductively in the following compositional sense: given a complex (syntactic expression) S^* of S, consisting of parts S_1, S_2, S_3, etc. the interpretation $I(S^*)$ is assumed to be defined in terms of the interpretations $I(S_1)$, $I(S_2)$, $I(S_3)$, etc., by an inductively-specified process of formation that is purely a function of S^*'s *formal* (in the positive sense—i.e., *grammatical*) structure. The inductive structure of the syntactic formation rules, plus this so-called **semantic compositionality** (essentially: an isomorphism or homomorphism between the grammatical structure of S and the "formation" of in-

[30]I.e., so as to license the inference $Q(x,y) \,\&\, Q(y,z) \Rightarrow Q(x,z)$ up to three times per derivation, but no more—as one might be tempted to suggest for a relation such as *Near*. Of course this constraint (any many others) can be "worked around" by coding the number of applications in varieties of the predicate itself, but the rule stands that, per se, use is not a legitimate ground for syntactic definition.

terpretations D[31]) ensures that the language is *systematic* and *productive*—capable of expressing untold new things, in a regular way. It is normally of great importance that all the basic vehicular ingredients—the stock of elements comprising S, the grammar or syntactic formation rules by which they are assembled into complexes, and the interpretation function I (that is: everything except the semantic domain D) be *finitely specifiable*. "Infinite expressive power via finite means" is something of a mantra in logicist quarters. It is all a little a fantastic Meccano or Erector set, with an unlimited supply of perfect, infinitely strong, weightless parts, in a world without friction, rust, or decay.

2b Computational component

Needless to say, for even the most cursory account of logic a vast amount more needs to be said—about truth, for example, and soundness and completeness. I'll make a few such remarks in a moment. More important for our purposes, however, is how much has not, *and does not need to be*, said: (i) anything about the nature of the representational vehicles, for example—whether they are linguistic, pictorial, distributed, etc.; (ii) anything about the nature of the domain D—such as whether it is composed of fields, features, objects, properties, dreams, ideas, or such. Logicism's specific assumptions in this regard will come up later, in the generalization phase; one of the points of framing the conceptual structure as we have done is to prepare it for a much wider than normal set of possibilities

Instead, consider what is involved in turning the foregoing logical picture into an active, computational system—of the sort that cognitive science classically imagined to be an appropriate or at least possible model of intelligence.

If not actually static, representational systems of the sort just described are at least *a-temporal*; the proof or derivability relation (P) is just that: a formally specified abstract *relation*. But even in logical guise there is almost always a residual bias towards think-

[31]Note: nothing requires that D itself have any structure whatsoever. So for example, the interpretation of the (inductively-defined) expression ((2+3)*(4/5)) is the atomic number four; not anything with a structure corresponding in any way to the grammatical form (_,_)·(_,_)).

ing of *P* in a forward direction, as bearing some relation to active patterns of rational thought. In the hands of cognitive science, given its interest in how people actually work, background bias becomes foreground concern: it is necessary to convert *P* into a *temporal process*. This "temporal mechanisation" of *P* can be viewed as the "computational turn" on the logical framework.

Needless to say, mechanising inference is far from trivial. Some raw materials are already in place: the relation *P*, and the formal properties of the expressions of *S* in terms of which it is defined, are all already constrained to be formal/syntactic, and so *P* (it is understood) is thereby amenable to direct computational implementation. The major problem is that, for any plausible representational system of the indicated sort, the proof relation *P* will be wildly branching. Given a set of expressions S_1, $S_2 \ldots S_k$, it becomes a major issue to determine which particular S_i to have the system produce, out of the (typically vast) set licensed by *P*. Many knowledge representation, planning, theorem proving, "search strategies," and other cognitive science projects can be seen as attempts to solve this problem, under the general rubric of "controlling inference".

The standard way to attack this problem was the following: one implements the representational vehicles—the representational vehicles, the elements of *S*—as data structures in a computer system, and then writes a program (we'll call it PROG) whose function is to make transitions from initial elements of *S* to final elements of *S* in some interesting or plausible way. The idea, that is, is that program PROG specifies the (potentially quite complex) behavior of the inference process *over the specified domain of representations S*. In computational jargon, this can be characterised as saying that the program specifies the behavior of a process over the data structures.

A note in passing. Though there is nothing especially problematic about proceeding in this way, one fact about this situation has proved remarkably distracting. It has to do with a conflation of, and subsequent theoretical confusion between, two distinct languages. The standard way to construct representational vehicles, as mentioned above, is to construct them in terms of a representational system which I will call the **representation language**

(L_S—i.e., the language in which syntactic formulae *S* are defined). The programs for controlling inference, at least in their so-called "source" versions, typically consist of a set of expressions in *another* formal language, which we will call the **programming language** (L_{PROG}). (For example, suppose one were to implement an inference system to work with expressions in the first-order quantificational calculus: the programming language might be C++, whereas the representation language would be the an encoding of first-order quantificational calculus expressions in C++ data structures.

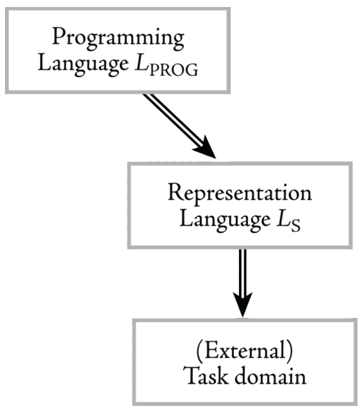


Figure 4 — Programming

As indicated in figure 4, what matters is that L_S and L_{PROG} are *different languages*. They refer to different domains, are subject to different constraints, and have radically different interpretations. One simple way to characterise their relation is to note that the programming language L_{PROG} stands at one level of semantic ascent *above* the representation language L_S: what expressions in L_{PROG} denote (i.e., are "about") is *formal transitions over (syntactic) elements of* L_S. The programming language L_{PROG} is a meta-language, that is; representation language is object language. It is only because the two languages have been confused, I believe, that cognitive science has called the classic representational view of mind *computational*.[32])

Sometimes, the process that makes the moves from starting expression to final expressions (perhaps via a long series of intermediate steps) is reified into a separate conceptual component of the overall system—leading people to say that the representations are "read" and "written".[33] However identifying the specified process as a sub-process of the overall behavioural system is both unnecessary and generally confusing[34]—and anyway there is no

[32]Smith, Brian Cantwell, *One Hundred Billion Lines of* C++, «ref».
[33]E.g., cf. Haugeland.
[34]In part because there is a tendency to confuse the process over representations (i.e., the process that manipulates the sub-personal representational structures *S*) with the overall process of which it is a part (which, if

reason why the overall system has to be implemented in that way. More generally, therefore, and as far as possible to avoid confusion, I will avoid any talk about programs PROG and the programming language L_{PROG} in which they are written, and simply talk about one process (the overall, inclusive, one, of which the representational vehicles are a part), whose behavior comes about as the result of **effective transitions** from elements of *S* to new elements of *S*.

2c Semantics

Given this overall picture of logic, four points need to be highlighted, to prepare us for subsequent reconstruction.

The first has to do with the question of just which parts of figure 3[35] are, and which parts are not, subject to effectiveness (computability) constraints. The answer is straightforward: effectiveness concerns only the upper third of figure 3; the formal (in the positive sense) properties of the representational vehicles *S*, and the proof-theoretic or inferential relations *P* between and among them—i.e., just 2 of the 5 ingredients of the overall logicist picture.

Crucially, not only is there no requirement that the interpretation function *I* be effective; there is no reason to believe that thinking that *I* must be (or even is) effective is even conceptually coherent.[36] For "effective" means something like "mechanically implementable," or "is a kind of operation that a Turing machine can do"—i.e., a temporally extensive operation that starts and ends with concrete entities or at least arrangements—a type of constraint that doesn't make any sense when talking about the non-temporal relation between, say, a numeral and a presumably abstract number. Moreover, what is *formal*, in what we are calling the (positive) grammatical sense, is again just the representational vehicles. The semantic domain *D* is sometimes characterised as formal, but whatever that means, it must be in a different sense from either of the positive or negative readings we have given to

the cognitive model or AI project were successful, would be a person). This is just one of the mistakes that Searle makes, for example, in his famous Chinese Room thought experiment «refs».

[35] P. [redacted].

[36] Cite AOS, appropriate volume.

that term (for example: it might mean *mathematical*).

More generally, whatever is the nature of the interpretation relation *I* between representational vehicles and the entities they designate or denote, it is not something that *happens*. The numeral '2' designates the number two, or so at least it is normally presumed; but that "designation" is not a *process*, not something that *happens*, not something that takes *energy* or *time*. Semantics—at least in the small—is something that "obtains."

Thus to take a human example, suppose one has a thought about the Brooks Range, or about Cheops, or about the great day on which the United States elects its first female President. The relation between one's thought (be it a state of mind, an active brain process, or whatever) and what it is about—the "directed arrow of reference" that starts in one's head and leaps out across time and space to the north slopes of Alaska, to an Egyptian ruler in the third millennium B.C.E., to a day in the (with luck) not too distant future—that referential arrow is not part of the energetics of the world. Referential entities, even referential *activities*, do not bathe their referents in any flux of discriminable energy. As I have said in another context, not even the NSA[37] could build a meter, to be worn in one's pocket, that could detect whether the bearer was the subject of an intentional act. The problem is not that referential signals are too faint, or that our physics is not sufficiently advanced, or that some form of quantum mechanical wizardry is at work. Rather, the reason is that semantic properties—being referred to, being true, being consistent, etc.—*are not effective*. As we will see, that is one of their enormous virtues—something that causal reductionists ignore at their peril. (And of course it is a good thing that reference, for example, is not effective; it is exactly the fact that reference is *not* effective that allows us to refer to the past, or to refer to the future, without therein violating physical proscriptions on forward or backwards causality).

2d Naturalisation

So that's the first point: syntax and inference (proof) are subject to effectiveness conditions; semantics and interpretation are not. The second remark, which is related, has to do with science and

---

[37]The U.S. National Security Agency.

naturalisation. In cognitive science and computer science, if not in logic or philosophy per se, it is common to think that the logicist tradition, in virtue of its commitments to formality and to the (at least potential) mechanisation of inference, is thereby rendered naturalistically palatable, in the following strict sense: that logical systems, *in virtue of being formal*, thereby somehow secure an at least potentially causal explanation.

Nothing could be further from the truth. Yes, as just suggested, there are some reasons to suppose that the upper half of figure 3—having to do with proof, syntax, and inference—may be amenable to causal explanation (though even showing that is going to take contentious reconstruction). But nothing in the picture laid out above provides any reason to suppose that the semantic interpretation relation *I*, or the semantic domain *D*—or indeed any interesting semantic property—need necessarily succumb to causal account. Logicians, in my experience (as opposed to logically-oriented computationalists), are *mathematicians*, not *naturalists*.

Indeed, some of the most prominent results in logic, such as the incompleteness theorems, could not even be formulated in purely naturalistic language. But we don't need any such radical conclusion here. For our purposes all that matters is that questions of what secures the interpretation function *I*, what sort of account semantics will ultimately be explained by, either in a particular case (such as arithmetic) or in general, is not required, by anything in the logicist framework, to be naturalised or even naturalisable. Perhaps semantics can be naturalised; perhaps cognitive science will show us how to naturalise semantics.[38] But *logic* doesn't show us how.

This negative observation will figure centrally in the upcoming reconstruction. As intimated above, various prominent counterproposals to the logicism—from theories of self-organising systems to proposals to understand cognition as a dynamical system to literary philosophies of the body—are, in this sense, more conservative than the logical tradition from which they aim to free

[38]Exactly this is the aim of such projects in philosophy of mind as Fodor's asymmetrical dependency theory, Dretske's informational account of semantics as counter-factual-supporting correlation, Millikan's theory of semantics as grounded in a biological notion of proper function, etc. «refs»

themselves, in virtue of being more committed (in advance) than logic to a scientifically traditional form of causal explanation.[39]

2e Norms

The third remark about the logicist framework summarized in figure 3 is very important. Logical systems are **normatively constrained**: strong evaluative metrics govern the ways in which the four ingredients (syntax, semantics, operations, semantic interpretation) are tied together. The importance of these norms cannot be underestimated; they are utterly critical to the stuff and substance of representational schemes. Without them, the whole apparatus of logic (and representation more generally) would collapse.

The primary norms embraced in the logical tradition are **soundness** and **completeness**: two versions of a rough requirement that what is derived (by *P*) correspond to what is true (in *D*). Though 'soundness' and 'completeness' are not very symmetric *terms*, the norms have a clear symmetry. Systems are sound just in everything that can be derived (formally, effectively) in the syntactic realm *S*, from a starting set of premises, is true or valid,[40] in the semantic domain *D*; systems are completeness just in case the converse is true: everything that is true or valid in the semantic domain *D* can be derived in the syntactic realm *S*.[41] Normally, one simply proves or demonstrates the soundness of a system, and the shows its completeness (if things work out well[42]). But to prepare us for a more general account, we can think of this as a two stage process. First, soundness (truth-preservation) and completeness are specified to be the governing norms. Soundness and completeness, that is, should in the first instance be understood as *regulative*; then the proofs that the given system is sound (and perhaps complete) should be viewed as *demonstrations that the system in question has met its regulative constraints*.[43]

[39] «Highlight this irony»
[40] I am not distinguishing truth and validity here …
[41] A more general reading of soundness and completeness is given in §■■, below.
[42] «Put in a note about completeness, in model-theoretic guise, often being a sham»
[43] «Insert a sidebar on the division of labour between *truth* & *soundness*—

When logical system are presented, traditionally, the syntax, grammar, proof regimen, interpretation function, etc., are all usually simply laid out in ostension—as if they had arrived, full-blown, as "facts" for theoretic consideration. But lurking underneath this symmetric presentation is a critically decisive asymmetry: whereas, as we have already seen, it is the formal or inferential facts (i.e., issues having to do with the upper half of figure 3) that are subject to constraints of effectiveness, it is the semantic facts—interpretation, truth, validity, etc. (i.e., issues having to do with the lower half of the diagram) that, from a normative point of view, are in the driver's seat. That is: at the most general level, the normative constraints on a logical system take the following form:

> *The (upper-level) effective transitions are normatively regulated to honour the (lower-level) semantic facts.*

This general pattern—of the effective mandated to honour the semantic—is as deep a fact about logic as there is. It will stand with us throughout our upcoming reconstructions (indeed, it survives even much more radical reconstructions than we are able to assay here[44]). If all one wanted were a *causal construction kit*, one would be crazy to choose logic. What logic gives us is something radically more substantial: *normatively-governed* construction kits.[45] Without norms, logic would be an empty vessel, devoid of substance—uninterpreted mechanism flapping aimlessly in the breeze.

2f Independence

The fourth and final comment about logicism has to do with the relation between the syntactic and semantic realms implicit in the

---

i.e., between what parts of the "worth" of a logical system are supplied by the language and inference and interpretation rules, and what parts by the axiomatization of the task domain ⟦cf. the division of labour between the calculus and the laws of motion in physics⟧.»

[44]Smith, Brian Cantwell, *On the Origin of Objects*, Cambridge: MIT Press, 1996.

[45]That's not quite fair, of course; what logic gives you—as I hope to make clear before the end of the paper—is a radically specific form of *semantically governed* construction kit.

picture we have been working with (again, see figure 3). In particular, the overall picture is constituted against three independence and one dependence claim.

The first independence claim concerns the two basic realms in terms of which logic's conceptual framework is articulated. Paradigmatically, the realms are established—or, as one typically but curiously says, "specified"—independently: one delineates them in separate stages, giving each its own autonomous ingredients, identity conditions, etc. Since modern logic was developed to deal with issues of mathematical inference, it may be that the ontological character of the realms was assumed to follow from whatever ontological conditions warrant the metaphysical existence of general mathematical entities. But whatever the reason, the two realms are assumed to be autonomously specifiable.

The second independence claim (again separating the two realms) is implicate in the so-called **formality condition**, mentioned earlier—the universally-accepted requirement that the operations or transitions on *S* constitutive of *P* be defined purely in virtue of the (positive) form or syntax of the elements of *S*, *independent of those expressions' semantic interpretation I(S)* [*D*. The formality condition, that is, is s second way in which the realms *s* and *d* are separated.

Third, the interpretation function *i* is critically assumed to exist independent of (and again explanatorily prior to) the operations or transitions constitutive of *p*. This is essential in order for the governing norms to take the form that they do. It would be ill-formed (i.e., would violate this third explanatory autonomy) to take a sentence *S* to mean something like "This very sentence has not yet been derived"—since in such a case *S* could be true, *but could not be (soundly) derived*. In general, the conceptual structure of soundness and completeness requires that the interpretation be established (or exist) *prior to and independent of* the operations, in order for the normative constraints on the operations to honour it. If A's job is to honour B, then B had better be defined independent of A, or else one runs the danger of setting up a vicious cyclicity.[46]

---

[46]This way of putting things exaggerates necessity, though not the accepted structure of formal logics. As we will see, it may be possible to defined

But then, once the realms are all separated out in this way, and these strong independence standards are in place—i.e., once the autonomy and separateness of the two realms is firmly established in all the requisite ways—then the norms operate *exactly by tying the two realms back together again*. It is this **reconciling tug**, as I've said, that gives logical systems "bite." In a way, the underlying conceptual structure is almost ironic: first one ensures that everything is cleanly and totally and utterly kept apart (logically, conceptually, ontologically, whatever) so that, once things are separated, they can be *regulatively brought back together again*.

It is going to be of the utmost importance to determine what the initial separateness, and what the subsequent tying back together again, come to in an adequate representational reconstruction that is suitable for embodied cognition. For now, it is enough to see that the very raison-d'être of a logical system derives from this never entirely reconciled but nevertheless reconciling tug between the two realms. Minus semantic interpretation and governing norms—i.e., as a pure structural construction kit—logical and representational systems are wimpy. For purposes of sheer assembly, abstract Erector sets, hydraulics, or C++ would be vastly better—or even, for that matter, carbon-based molecules or DNA.

(semantic) interpretation *partially* in terms of operations, without rendering the resultant norm vacuous. But constitutive interdependence of this sort is one of the radical generalisations we will take up in §■■; it is never, so far as I know, employed in a logical system.

3 Reconstruction I • Computation

It is difficult to say exactly what it is about the classical picture that troubles proponents of an embodied approach. But at least eight properties have drawn comment from various writers. They are listed in figure 5, with the presumptive character of logicist models indicated on the left, and the properties recommended for a new, embodied or situated conception of cognition on the right. I make no claim that this list is complete, that the issues it enumerates are independent, that it does justice to all anticlassicist sentiments, or that it is correct in its characterization of logic (in fact I will presently argue that at least one entry in the left column is false). But it will serve for our purposes.

Crucially, the list doesn't mention *representation*. It is critical to our project, however, to recognize that one of the arguments frequently heard in the "embodied cognition" camp is that it is exactly in virtue of being representational that logic exemplifies the properties identified on the left—and therefore that, in order to manifest the properties listed on the right, a system must abandon representation.

For an advocate of generalisation, therefore, who resists (especially in *a priori* form) this strong antirepresentationalist stance, the tabulation raises two challenges: (i) to understand which of the characteristics in the left list are true of only a particular (logicist) species of representation, rather than of representation in general; and (ii) concomitantly, to the extent that any of the characteristics listed in the left-hand column turn out in fact to be species-specific, to understand how a generalised conception of representation can deal with the corresponding property identified on the right.

Given our concern with reconstruction, however, we first need to analyse the generalist's starting assumption: whether the characteristics listed on the left hand side of the table really do hold of representation on a logicist conception. That is: to what extent is the left-hand column *correct*?

	Logicist		Issue		Embodied
1	*Abstract, disembodied*	·	Materiality	·	*Concrete, embodied*
2	*Explicit, linguistic*	·	Vehicles	·	*Tacit, non-representational*
3	*Disconnected*	·	Environment	·	*Fully engaged*
4	*Separate, independent*	·	Realms	·	*Not separated*
5	*Static, atemporal*	·	Temporality	·	*Dynamic*
6	*Digital, discrete*	·	Character	·	*Continuous*
7	*Context-independent*	·	Interpretation	·	*Context-dependent*
8	*Ratiocination, thought*	·	Activity	·	*Improvisation, navigation*

Figure 5 — Dimensions of Differentiation

Start with the first point of alleged difference between logicist and embodied views: the claim that classical logic treats its subject matter abstractly—and thereby fails as a model of human cognition, because of its consequent inability to deal with important facts about humans' material embodiment.

At least in its first half, regarding the abstract treatment of formal logic, the claim seems true on the face of it. Not only does model theory almost universally analyse semantic realms in terms of purely abstract set-theoretic domains,[47] but even syntactic realms, while somehow vaguely concrete (for example in the sense of sustaining an idea of syntactic *tokens*, and being subject to mechanical realisability) are still not treated in reigning theories *as* concrete—as bluntly physical or material in any important (e.g., energetic) sense.

As already intimated, however, I believe that although this abstract view is socio-intellectually or epistemically correct about how logicians treat or analyse logic, it is *ontologically* misleading. It

[47]By 'purely' abstract I mean a set all of whose members (recursively) are abstract. In contrast, suppose A is a two-element set containing elements B and C, where B and C are also sets—B a set of camels and C a set of zebras. In such a case all three sets, *qua* sets, may arguably be considered abstract—but since the inner two are made up of concrete elements, I would not consider any one of the three *purely* abstract.

conveys the idea that logic is constituted abstractly, without impact or constraint deriving from physical reality. Surface appearances notwithstanding, and pace the protestations of practicing logicians, I will argue that the entire substance of the traditional logicist view rests on very real constraints that derive directly from a logical system's concrete materiality.

To see this, though, we need to step back from logic for a moment, and approach the subject matter of representation and computation from a far more general perspective than usual. That will be the task of this section; we will return to logic and the logicist model in §4.

3a Meaning and mechanism

The most fundamental issue underwriting representational and computational systems—and, more specifically, the issue that underwrites the classic logicist tradition in cognitive science in particular—is the interplay between **meaning** and **mechanism**. So important is this issue, this contrast, this generative tension, that in other writings I have dubbed it the **primary dialectic** of the intentional sciences. What it comes to depends on what one takes 'meaning' and 'mechanism' to mean; but at a very rough level, the question is something like the following:

> How can things that are entirely concrete—no magic, spirits, divine intervention, etc.—*without violating that inexorable underlying materiality*, nevertheless, in the appropriate sense, "transcend" that materiality, so as to think, dream, mean, wonder, refer, be right and be wrong?

I have called this a dialectic, but that does not mean it is an outright opposition. Cartesian predilection notwithstanding, few believe that meaning and matter are opposites or distinct substances, in the sense that the world consists of those two kinds of things, glued together with God's own epoxy of set theory. Rather, at least for materialists or physicalists, the question is how ordinary bodies or mechanisms, which in one sense *are* merely physical, in another sense are *not* merely physical, but must instead authentically and legitimately be understood (perhaps even constituted) in intentional terms?

I believe it is impossible to understand the whole edifice of syntax, semantics, formality, truth, soundness, etc., as adumbrated in the previous section and refined over more a century of academic scholarship, except as an attempt to instantiate a plausible answer to this daunting metaphysical question, albeit in an extraordinarily restricted setting.

3b Effectiveness

It would be natural to assume, of this dialectical pair of meaning and mechanism, that the meaning or semantics side (truth, meaning, representation, content, etc.) would be the troublesome element. It would be natural to assume, that is, given 300 years of spectacularly successful natural science, that we would have an adequate and even good grasp on the material or mechanism side.

It would be natural—but it would be wrong. It turns out that coming to grips with the "mechanism" half of the dialectic has proved almost as difficult as understanding meaning and truth.

The notion that has been in primary focus, in the quest to tame the mechanical, has been that of **effectiveness**—as betrayed in the fact (already mentioned) that the reigning mathematical foundational theory taught in computer science departments is called the *theory of effective computability*. As it happens, I have grave doubts as to whether this vaunted theory merits its ubiquitous name "theory of *computation*," but that it focuses on effectiveness is surely right. The aim behind this body of work has been to formulate, in as clear and theoretically profound a way as possible, what can be done, by a concrete physical mechanism—both absolutely (i.e., without restriction on time, space, or other finite resource), and relatively (in the sense of with relation to more or realistic constraints on allowable resources bounds).

These issues have been explored in what seem to be relatively abstract systems, under the guise of syntax, proof theory, and numerical computability. Theoretical results are by and large framed mathematically (e.g., in the difficulty of solving this or that mathematical problem, or the complexity of, for example, factoring products of large primes). It is this rampant mathematicization that, I believe, though not problematic on its own, has within the larger scheme of things proved radically misleading. In another place I argue at length that all computability results—both absolute, as in Gödel's incompleteness results, the unsolvability of the halting problem, etc., and relative, as in the results of complexity theory, the difficulty of deciding classes of formulae, etc.—derive directly from physical, material constraints on underlying mechanisms. Sure enough, in the theory as we know it the results are *framed* mathematically, but so are (at least many) results in physics and chemistry. Present theoretic practice notwithstanding, the subject matter of theoretical computer science is by my lights entirely concrete.

Let me admit straight up that this is a contentious claim; I have yet to meet a logician who believes it. Informally, though, in my experience, most working computer scientists not only *believe* it, but so thoroughly *assume* it that it takes work to show them that it is not in fact something that logicians presuppose.

What makes the issues subtle—but at the same time interesting—is that it is clear that major computability results are not specific to any *particular* material substrate. Factoring primes is approximately equally difficult, whether one uses vacuum tubes, silicon transistors, or even tinker toys. This betrays what I dub the **secondary dialectic** underlying computing: between the *abstract* and the *concrete*. My claim is that, whereas physics (and perhaps material science) has focused on the completely concrete, and mathematics (presumably) on the completely abstract, the "natural home" of computability results lies somewhere in between—*but much closer to the concrete end than is normally (especially theoretically) realized*.

Historically, the reasons why the formal "theory of computability" has framed its results mathematically are sure many, including (but not limited to) the following:

4. It is a perfectly evident observation that computation, at the level at which theories have dealt with it, can be, as it is said, *multiply realised* on a wide (perhaps even limitless) variety of different substrates;
5. The theoretical aim has been to identify very general results, rather than specific material concerns (for example, it is only recently that computer science has begun to deal with real-time results);

6. Historically, the tradition developed out of concerns with metamathematics, making an abstract perspective more natural;
7. Scientific results are almost always expressed mathematically; since the computability results were not framed in terms of any readily-identifiable units (kilograms, ergs, etc.), the equations appear to traffic in purely numeric quantities;
8. Since the problems for which computability results were developed had primarily to do with mathematical subject matters, the languages used to represent them were by and large context-independent, which turns out to imply that one could frame results purely in terms of types, without regard to concrete specific facts about individual tokens (the way one needs to do when treating indexical expressions, for example).

Of these five, the first (multiple realisability) is indubitably most famous, but the last, having to do with the relation between types and tokens, may cut the deepest. Since the subjects matters taken up in the theoretical context have (contingently) been primarily abstract, it has proved convenient to deal with them abstractly. But what it is that is abstract, in my view, has been misinterpreted. In particular, I argue:

> Reigning theory of logic and computing treat computational entities (states, marks, etc.) as *abstract individuals*, whereas in fact they are more properly understood as *concrete type*—i.e., as types of concrete things.

The reason why the difference matters is because the constraints on the notions (what it is to be a state, what it is to be a mark, where the properly-vaunted computability come from, metaphysically) derive from the concrete, physical world—the world of which they are types, rather than from the abstract, logical, or mathematical world (where types presumptively "live").

Arguments supporting the changer in perspective rest on such facts as that, if one changes the physics of the realizing substrate, one can change complexity results at will. Intimations of this were recognised as early as in the 1930s by Robin Gandy,[48] who showed that the absolute computability results depended in immediate and subtle ways on the character of the physical mechanisms on which they were assumed to be implemented.

In the end, though, the proof of any theoretical claim rests heavily on its theoretical utility. Some of the arguments I advance are negative: that not recognising and understanding the physical nature of effectiveness leads directly to various negative entailments: one can solve the halting problem, one cannot explain the ubiquitous notion of a "reasonable encoding," etc. The lion's share of the argument, however, rests on positive results: that if one does recognise the concrete nature of effectiveness, one can (among other things) achieve the following sorts of results:

9. Explain the notion of a reasonable encoding (both what the constraints on being a reasonable encoding are, and also why the notion of a reasonable encoding has received so little theoretical attention);
10. Make sense of the rise of Girard's linear logic, computer science's interest in intuitionistic type theory and constructive mathematics, etc.[49]
11. Predict the proposed fusion of foundational theories of quantum mechanics and computer science-based theories of information;
12. Make sense of why physicists are interested in super-Turing computability, continuous models of computation, quantum computing, etc.; and
13. Resolve otherwise unexplicated tensions between what is real and what is virtual (e.g., in popular conceptions of computational technology).

In spite of these benefits, the proposed adjustment in our understanding is not without cost. For example, it is a inescapable consequence of reconstructing the current (so-called) theory of effec-

---

[48]Gandy, R. (1978), 'Church's Thesis and principles for mechanisms', in K. J. Barwise, H. J. Keisler, and K. Kunen, eds., *The Kleene Symposium*, Vol. 101 of *Studies in Logic and Foundations of Mathematics*, New York: North-Holland, pp. 123–148.

[49]«Ref Girard, Martin-Löf, etc.»

tive computability as a theory of *effectiveness* that it emerges from that reconstruction as no longer being a theory of *computing*, because it deals with only the first (mechanism) arm of the primary dialectic, not with the second (meaning, semantics, reference, truth, etc.). When conjoined with the present point, that the underlying constraints that give substance to the theory are direct consequences of the concrete, physical nature of the underlying medium, one is forced to conclude that what is universally known as the theory of effective computability is, in point of fact, (and presumably will eventually be historically recognised as) a **mathematical theory of causality**—namely, a theory of what can be done, in what time and with what resources, by what sorts of arrangements of concrete, physical stuff. That such a theory should be framed at some level of abstractness, away from very specific concerns having to do with particular materials, is entirely to be expected. It is for this reason that I have dubbed the properties that the theory traffics in *effective* properties, rather than *physical* properties; they are properties that systems (or states) can *do consequential work in virtue of possessing*.[50]

Two final remarks.

First, a proponent of embodied cognition might argue that even if we do reconstruct computability theory as a theory of causality, it will still be too abstract for cognitive science: that in order to understand cognition "in-the-wild,"[51] one needs to understand not only relatively abstract causal properties of the system, but quite concrete properties (such as heft and materials)—e.g., in order to understand rhythm and dynamic movement. That may be, but there is every indication in theoretical computer science that the theory in question is rapidly being refined so as to deal with more and more direct physical parameters (in order, among other reasons, to treat issues of three-dimensional packag-

[50]It is also unclear exactly what it is to be a physical property. Being a million light-years from Alpha Centauri is presumably a physical property, but not an effective one; it would be impossible, at least in any remotely practical sense, to build a device that could "detect" the exemplification of this property.

[51]The term is from Edwin Hutchins, *Cognition in the Wild*, Cambridge: MIT press, 1996.

ing and real-time computing). Moreover, the embodied cognition movement has to be interested in bodies and materials at *some* level of abstraction. Suppose one were to replace the control circuit for the muscles of an animal with an electronic souped-up version; what matters, presumably, even to the most materially-oriented theorist, is that the signals match, that power be supplied, that the right function be computed in real-time, etc. There are questions of whether such implants could work—and how much of our cognitive facilities could be upgraded in this way. But virtually no one thinks that a brain implant would literally have to be made of DNA-based neurons, in order to function in a "materially" appropriate way. Put it this way: neurophysiology and the theory of effective computability are climbing up the same mountain, even if from different sides.

Second, let me reiterate what I hope is clear: that reformulating our understanding of (what is known as) computability in concrete, material terms, in the recommended fashion, is an enormous as-yet open intellectual task. My guess is that it will take decades for the transformation to take place. For example, all absolute and relative computability results—that whether a Turing machine will halt on an arbitrary input cannot in general be algorithmically decided, that factoring primes is hard, etc.—will have to be reformulated as issues about mechanisms, not issues about *numbers* or *decisions*.

Nevertheless, this first reconstruction of computation—recognising the physical character of the notion of effectiveness that constitutes half of the primary dialectic on which computing rests, and that serves a lynchpin in our understanding of logicism—is a necessary prerequisite, I believe, of understanding the essential character of representation.

4 Reconstruction II • Semantics

Turn then to the second arm of computing's primary dialectic: meaning and semantics. There is a major reconstructive move to be made on this side, as well, again have to do with physicality. This time, however, the issue is not with the relation between the abstract and the concrete—what I called the secondary dialectic. Indeed, in order to get at it, we first have to set a potentially distracting of abstractness aside.

4a Models

Consider semantic domain *D*.[52] As we saw, in classical logicism this realm is usually treated abstractly: as a set-theoretic construct of objects, properties, and relations, perhaps extended with functions, situations, states of affairs, facts, propositions—and sometimes possible worlds. It is not that the atomic objects on which this construction is based need necessarily be abstract—i.e., it is not that *D* need necessarily be purely abstract[53]— but rather that the composite structure into which the objects and properties and such are assembled, for semantical purposes, is (again, typically) more of a mathematical structure than it is, say, the full disheveled situation out the window.

One self-evident generalising step already presents itself, therefore: if the embodied cognition movement aims to deal with material creatures interacting with their environments, we will have to adjust our conception of semantics so that semantic domains don't just include concrete individual objects "at the bottom," as it were, but are themselves *full concrete environments*, such as train station platforms in modern Tokyo, or the messy situation where the Amazon pours out of Brazil into the Atlantic Ocean.

There is a methodological subtlety here. The reason that semantic domains are paradigmatically mathematical or abstract is that, in the classic tradition, semantics is usually studied **model-theoretically**. The semantic interpretations of representational vehicles are analysed in terms of abstract (set-theoretic) models or "stand-ins" for what I will call the genuine target domain, rather

[52]Figure 3, p. ■■
[53]See fn. ■■ on page ■■.

than directly, in terms of the real target domains that the vehicles are authentically about. The situation is depicted in figure 6. Suppose we construct a logical axiomatisation of the patterns of car movements on the expressways surrounding New York. *D*, in the figure, would be the actual, metaphysically occurrent concrete situation on the roads around town; *M* would be a *mathematical model* of those freeways, in terms of which the formal semantics would be formulated.

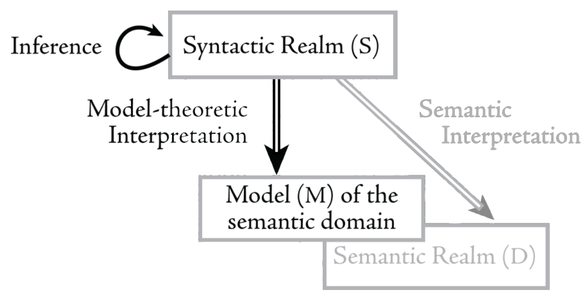


Figure 6 — Model-Theoretic Semantics

The rhetorical situation in this situation is complicated by the (often-noted) fact that the term 'model' is used technically, within the logical tradition, in a non-standard way. It in technical or theory-internal contexts, logicians speak of an element or structure of *M*'s being a *model of a sentence S* (or of some other syntactic or representational entity). But that way of putting things splits from lay parlance, which would more likely call *M* a model of the target domain *D*. Thus imagine some aeronautical engineers producing a blueprint of a new kind of airplane wing, designed (say) to avoid turbulence at high lift. Suppose, to test the design, they build a plastic model to try out in a wind tunnel. Normal parlance would call the plastic device a *model of the wing*, not a *model of the blueprint* (i.e., would call *M* a model of *D*, not a model of *S*). In logic, however, the entity that is analogous to *M* is said to be a *model of the sentences*—i.e., a model of the entity that is analogous to the blueprint.

On the face of it, this is just a terminological ambiguity, so confusion need not reign so long as we keep usage clear. (In this paper, since my audience is cognitive scientists, not logicians, I will side with lay practice, and talk about *M*'s being a *model of the target domain*, not of the representational vehicles.) But more serious issues arise if, forgetting that *M* is a model, we mistakenly take (what I will call) **insignificant** properties of *M*—i.e., properties of *M* that are *not* intended to model anything in *D*—to be part of the interpretation of *S*.

It is a truism, after all, that not all properties of a model *M* can be intended to represent or model properties of that which it models. Thus in our example, the model of the airplane wing was made of plastic, but presumably with no implication that the aircraft wing was to be fabricated of plastic. Similarly, the cost of the model presumably bore no modelling relation to the cost—or indeed to any other property—of the thereby-modeled wing. Of course one can construct examples in which these things are false: one could construct a wing in which materiality (plastic, wood, whatever) of the model corresponded, directly or indirectly, with some property (perhaps the materiality) of the thereby-modelled wing. The point is only the following: only some (usually a finite number) of the (infinite) properties of a model are ever intended to correspond to only some of the (infinite) properties of what is modelled.[54]

For our present purposes, what matters about the model-theoretic approach to the semantics of logic has to do with its abstractness. In particular: *From the abstractness of set-theoretic models, nothing* (necessarily) *follows about the concreteness or abstractness of genuine (target) semantic realms*. Methodological abstractness, that is, need not vitiate subject matter concreteness. So discussions of the issue of abstractness vs. concreteness—item number 1 in figure 5 (page ■■)—should not be influenced by the fact that logicists do semantics model-theoretically. Doing so is compatible with arbitrarily concrete commitments about the nature of the semantic domain.

At the same time, we mustn't assume that it is intrinsic to embodied cognition that the relevant semantic or task domains must

[54]What is really going on here is that the relation between *M* and *D*—what I am calling a "modelling" relation—is itself a semantical or representation relation, and should be studied as such. In part because of this use of semantical relations to study semantical relations, and also because of the point raised in the text—that it is tempting to forget which properties of the model are genuinely significant, and then inadvertently to take insignificant (non-modeling) properties of *M* to be significant—my own preference is to avoid model-theoretic semantics entirely, in favour of what I would call **direct semantics**.

be entirely concrete (i.e., talked about in terms of bare materials). Suppose an agent is designed to proceed in the face of a single obstacle on its path, confident that its navigational skills will allow it to negotiate its way around one thing, in real-time. When it encounters a group of more than one obstacle at once, however, it is designed to stop and plan a deliberate route around them. Are "groups of obstacles," or "routes," concrete or abstract? Who knows? And no matter how concrete the domain into which an embodied agent wanders, it will always be true (minimally, because of finite resource bounds) that such agents will need to deal with those domains at some level of abstraction.

Put it this way:

> *For cognitive science to deal reasonably with embodied and embedded cognitive agents, the secondary dialectic adumbrated above, having to do with the relation between the abstract and the concrete, will be as applicable to environments and task domains as it is to creatures and cognition itself.*

4b Formality

Setting issues of abstractness provisionally aside, then, turn to the second critical reconstruction, this time having to do with semantics and formality.

It is a prominent and profound fact about logicism that logical systems are considered *formal* systems; that logic is the product of the *formal* tradition, that to construct a logical model of something is often identified with *formalising* it. Just what 'formal' means, however, is one of the most diabolically complex issues in this entire subject matter.[55]

Overall, there are two rough sense of formal that need to be distinguished. The first, which I will call *methodological*, has to do with what it is for logic (and perhaps computing) to be a formal discipline, what it is to "formalise," and the like. I will not deal with these concerns here except to say that they seem intimately tied up with expectations and assumptions mentioned earlier: about naturalisation, about the possibility of giving explanatory scientific accounts, about the possibility of mathematical analysis, and the like.

[55] «Ref AOS»

What I do want to focus on is another set of formality intuitions, this time more *ontological* in character, having to do with how such systems are as a matter of fact structured, with how they work. Some of these intuitions were mentioned earlier, in §2. In particular, it is taken to be a criterial condition that inference (proof, operations) work or proceed *formally*. This requirement, which, as mentioned in the previous section, Fodor has dubbed the **formality condition**,[56] is viewed as an absolute mainstay of the classical representational tradition. It is thought to bring to logic and computation, and thereby to cognitive science, its strongest weapon in the struggle to resolve the primary dialectic, and thereby to finally defeat the threat of the mind/body problem. Indeed, there are those who would say that formality is the very foundation on which the material success of the classic tradition relies.

It was also mentioned earlier that the formality condition has two different senses. The *positive* aspect of formality has to do with shape, syntax, grammar, or "form"; it militates that inference operations be definable (and work, causally) in virtue of the syntax or form of the constituent expressions (representational vehicles). It is the *negative* aspect of formality, however, that concerns us here: the ubiquitous assumption that both the syntactic properties and identity conditions on the expressions or representational vehicles (elements of *S*), and the operations or effective transitions defined over them, must be defined *independently of semantics*.

It may be that one of the *consequences* of the negative reading of formality has to do with naturalisation: that some of the overall logicist story (at least the upper half of figure 3) will be amenable to causal account. The original *motivation* underlying the negative reading, however, stems from a very basic insight about representation in general. And that is the insight we are after. For one the most serious Achilles' heels in the embodied cognition stance, as suggested in the introduction, is that, in distancing itself from the formal logicist tradition, it runs the risk of missing this insight, that underwrites formality: an insight that not only implicitly undergirds the classical tradition, but that cuts to the very heart of what representation is like—indeed, to what repre-

[56]See the sidebar on p. ■■.

what representation is like—indeed, to what representation is for.

The idea is this. Genuine semantic properties—being true, referring to Cheops, etc.—are not of the right sort to figure in how symbolic or representational systems actually work. *Semantical properties*, to put this in terms we have already used, at least in general, are *not effective*.

Intuitively, one reason semantical properties aren't effective is that they are often (perhaps always) *relational*. The truth of the sentence "dinosaurs were warm-blooded" seemingly depends on facts that obtained hundreds of millions of years ago—facts that, in a rough and ready sense, are simply *too far away* to do any work in affecting the right-here, right-now microdynamics of how an inference system works (human or machine). More generally, representations often bear semantic relations to situations or states of affairs that are *distal*, and distal things, because of the locality requirements of physics, simply cannot get into the act in affecting the here-and-now.

There is a discrepancy, that is, between:

14. The local, effective, immediate structure of a representational system, in terms of which it (causally) works; and
15. The paradigmatically distal, non-effective, semantic structure of the system, in terms of which it is normatively characterised.

The dialectic is mortal. Nothing will matter more to the story to come than the interplay between these two kinds of property. Indeed, it was already evident in the normative structure of the classical model we started with that what representational norms *govern* is the syntactic or proof-theoretic or effective local workings of the system, whereas what the norms are based on or are designed to *honour* are the system's semantic contents. We will get back to norms presently; what we can do here is to state, very simply, what I will take to be our second reconstruction—a reconstruction of logicism's commitment to a (negative, ontological) reading of formality:

Semantic properties aren't effective

Semantic properties, that is—the "orienting towards the world" properties in virtue of which representational systems are normatively governed— cannot in general be assumed to be effective, in exactly the sense of "effective" we talked about in the first reconstruction, above. They are not properties in virtue of the possession of which systems can make concrete things happen.

Three comments.

First, there is a scope ambiguity in the foregoing statement: whether it is being claimed that *no* semantic properties are effective, or only that *not all* semantic properties are effective (i.e., that one cannot conclude, in virtue of a property's being a semantic property, that it is thereby guaranteed to be an effective property). Call these the **strong** and **weak** readings of the reconstruction of formality, respectively. As will emerge later on, I believe that the strong reading is true; but for now we can make do with either.

Second, on the (negative) ontological reading of formality within logic itself, the claim was made that formal systems operate *independent of* the semantics of their ingredient states. "Independence" turns out to be a notion not unlike modality; it comes in strengths: logical independence, ontological independence, metaphysical independence, etc. It is no aim of mine here to say which notion of independence the formal tradition is committed to. What we can see, however, is that an independence claim is stronger than the "semantics is not effective" version just formulated (thus we are already starting our second strategy, of generalising).

In particular, we are in a position to begin to see what is wrong, or anyway too restrictive, about classical formality. Formal logic essentially infers, from the (manifest) *non-effectiveness* of the semantic, that the workings of the system must be *independent* of semantics. Sure enough, if semantics is not effective, then how a system works cannot depend on semantics in any very full (at least in any causal) way. But—and this is a critical generalising point—*there is a world of difference between non-dependence and independence*. That this is true is made obvious by reflecting on human affairs: someone can take your views into consideration, in forming their opinion, without adopting either extreme: of being slavishly dependent on what you think, or being so independent of what you think as to be wholly autonomous and uncaring. In human affairs, both limits are recognised forms of pathology. Between the two lies an entire realm of *partial dependence*—or perhaps better described, *partial interdependence*. To foreshadow a bit, partial interdependence, of this rough sort, will eventually emerge as the constitutive relation between (reconstructed) syntax and semantics—that is, between concrete, "make-it-happen" effectiveness, on the one hand, and non-causal directed-to-the-world normative governance, on the other.

But we are getting ahead of our ourselves. The point is that from an ontologically point of view, formality is wrong, because too extreme. But it rests on a profound insight, about the non-effectiveness of the (normatively-governing) semantic. Preserving this insight, and understanding its import in cases of embodied, embedded, engaged cognition, is the key to the challenge identified in the opening sections: understanding how to retain semantics through a transition from the abstract to the concrete.

5 Towards a participatory account

Before we turn in full force to the second, generalisation phase, it will help to summarise how far we have come. For already the outlines of a more powerful conception of representation can be discerned.

5a Logic, formality, and concreteness

What we are driving towards is a profound dialectical interplay between the effective and the non-effective. At the deepest level, I claim, this dialectic (albeit in a restricted form) has underwritten—has always been what matters most—about the logicist program. All sorts of familiar (and essential) features of the logical conception can be reconstructed in its terms. If we take "**rehabilitation**" to mean "reconstruction plus generalisation," then:

16. The "effective" structure of a representational system is the rehabilitation of *syntax* or *form*;
17. What the system does, mechanically, is the rehabilitation of *proof theory or inference*;
18. The situations or states of affairs in the world towards which the system is (normatively) oriented is the rehabilitation of *semantic interpretation*; and
19. The fact that the system is not (in general) effectively coupled with those situations towards which it is normatively oriented is the rehabilitation of the claim that inference operates *independent of semantics*.

The last of these claims of course has to do formality. Throughout the discussion so far, I have identified two different (ontological) readings of formality: a positive reading, having to do with syntax or "shape" or "form," and a negative one, meaning "independent of semantics." As should by now be evident, our two reconstructions dealt with the positive and negative readings, respectively:

20. The positive reading was reconstructed in terms of the fact that systems work in virtue of the presence of effective (concrete, causal) properties;
21. The negative reading was reconstructed in terms of the absence of effective coupling with the semantic domain.

The Representational Mandate

1. Conditions
   a. A representational system must work, physically, in virtue of its concrete material embodiment (the role of effectiveness).
   b. But it is normatively directed or oriented towards what is non-effective—paradigmatically including what is physically distal.
   c. Being neither oracle nor angel, it has no magic (non-causal, divine) access to those non-effective situations; just caring about them is not enough (physical limitations bite hard!);
2. So what does the system do?
3. It
   a. Exploits local, effective properties that it can use, but doesn't (intrinsically) care about—i.e., inner states of its body and physical make-up, in interaction with the accessible (effective) physical aspects of its environment
   b. To "stand in for" or "serve in place of" effective connection with states that it is not (and cannot be) effectively coupled to
   c. So as to lead it to behave appropriately towards those remote or distal or other non-effective situations that it does care about, but cannot use.

Figure 7 — Representational Mandate

That is, the two reconstructions can be viewed as "**concretisations**" of formality—as reformulations in concrete, physical terms of something that classic logic has dealt with in an (unfortunately) abstract way. That concretisation will stand us in good stead with respect to the goal identified in §1: of preserving an understanding of semantics through a shift from the abstract to the concrete.

5b The representational mandate

In a sense, the moral so far is a recognition that concerns of concrete materiality and have lain submerged, just below the surface, in the traditional logicist conception—out of explicit theoretical view, but nevertheless playing a critical role. What differentiates the new view is that those concerns are being brought into clear and unambiguous focus. In fact they almost define the character of the new view. For notice how thoroughly issues of concrete materiality permeate the emerging conception.

Representational or intentional systems, as we have seen, (at least typically) stand in semantic relations to distal and other non-effective situations. Such systems are normatively governed by those relations that they bear to those situations or states of affairs. But in a mechanical sense (on pain of violating physicalism), such systems cannot *work*, causally, in virtue of those relations—exactly because they are not effective. They can't work that way because the (semantic) properties tying them to those situations, and the properties of the situations that they are thereby tied to, are in general relational. So what do such systems do? They are constituted or arranged in such a way that they can use the (local) effective properties of their local, immediate structure—i.e., they use what is available to them: the effective properties of their causal ingredients, in conjunction with the effective (causal) properties of the environments in which they are deployed—*so as to behave, appropriately with respect to those distal and other non-effective situations.*

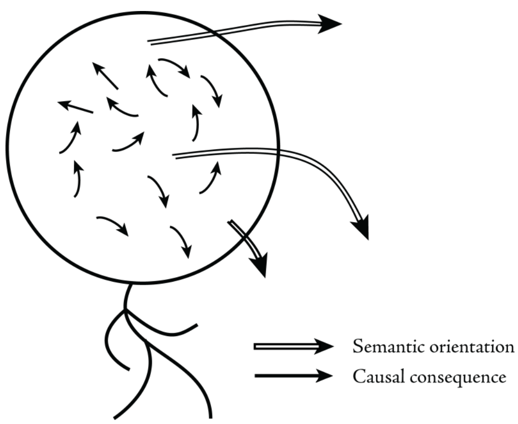


Figure 8 — Participation, First Pass

That is, a representational system:

> Exploits the effective properties of its inner states—properties that it can use, but doesn't intrinsically care about—to "stand in for," or "serve in place of" effective connection with states that it is not effectively coupled to, so as to lead it to behave appropriately towards those remote or distal situations —situations that it *does* care about, but that it *can't* use.

Or more simply yet, representational systems:

22. Exploit what is local and effective
23. So as to behave appropriately with respect to (to satisfy governing semantic norms regarding) what is distal and non-effective.

We still have to a considerable amount of work to do in order to see what this characterisation comes to in detail. But it will stand us in sufficiently good stead, over the long haul, to be worth a name. As indicated in figure 7 on page ■■, I will call it the **representational mandate**.

5c Coordination Conditions

A caricature of the view we are closing in on is given in figure 8. The system is constituted of a variety of states, and embedded in an (also causal) environment. In general, those states will exhibit two kinds of property:

24. **Causal consequences**, due to their effective properties, including the role they play in the overall machinery of the system (depicted as single-tailed arrows: '→'); and
25. **Semantic relations**, towards the states of affairs in the world to which the system is normatively oriented (depicted as double-tailed arrows: '⇒').

The stuff and substance of the system derives from the interplay between and among these two kinds of relation.

But figure 8 is too simplistic. It immediately needs to be generalised. First, it is not just the agent that is made up out of dynamic, efficacious states; the same is (in

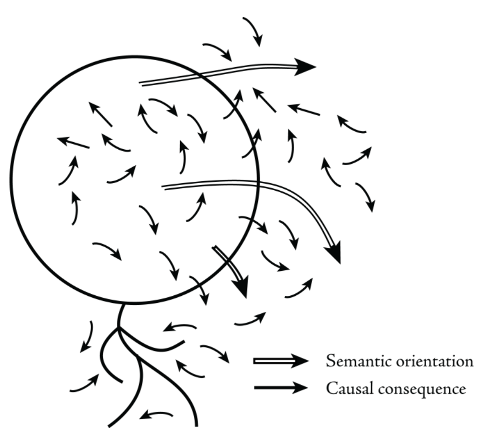


Figure 9 — Participation, Second Pass

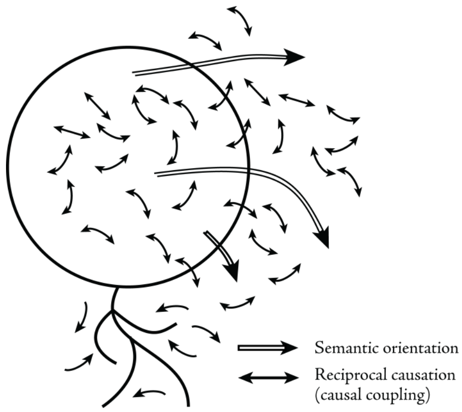


Figure 10 — Participation, Third Pass

general) true of its environment. So causal arrows need to be introduced into the environment. Moreover—in order to make room for the critical causal or effective engagement of the agent with the environment—arrows must also be added that cross the boundary (in both directions) between agent and environment. This much is shown in Figure 9. In addition, given that we are aiming at a general account (and with a nod to Newton's first law of motion), it is more general to change the (single-tailed) causal arrows to bidirectional ones, so as to license reciprocal causation, as indicated in figure 10. Finally, as shown in figure 11, it helps to indicate that semantic relations ('⇒') have vastly greater reach than causal arrows. They are not limited to states of affairs to which the system has effective access, but can leap across gaps in time, space, and even possibility, in dizzying array.

That is not to say that we have explained how arrows of semantic directedness are established, or even (metaphysically) what they are. Given a background physicalist metaphysics, they are going to depend on large-scale (distal and social) relational patterns, rather than on immediate patterns of local, effective coupling. But what is crucial to recognise here is that, once the two have parted company (for whatever ontogenetic reason), *it is the gap between them that al-*

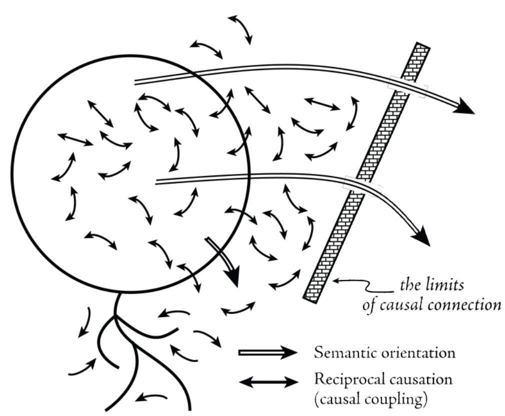


Figure 11 — Participation, Fourth Pass

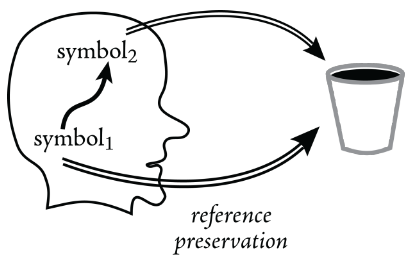


Figure 12 — Reference Preservation

lows normativity to establish a governing foothold. In fact such norms will eventually be identified as topological constraints on the relations between and among these two kinds of relations.

Simplistic caricatures of some familiar norms are shown in (structural coupling) the next set of figures. Truth- or reference-preservation—the traditional norm on sentential inference, and on term rewriting—is schematized in figure 12. Figure 13 depicts a basic constraint on perception: that a system, upon encounter with a situation f, end up in (or construct) an inner state that represents f. Similarly, a baseline condition on effectors is diagrammed in figure 14: that they cause to come into existence that situation that is represented by a state that triggers them.

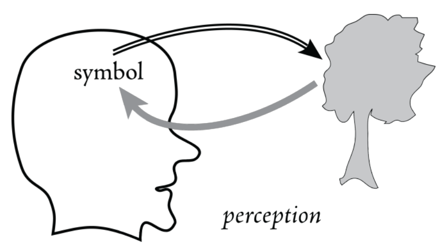


Figure 13 — Perception

Although it is reassuring to retain traditional norms, as soon as one applies this general framework to real-world systems,[57] it becomes evident that these are just three of a large number of potential normative constraints, some radically complex, and some of considerable interest to cognitive science.

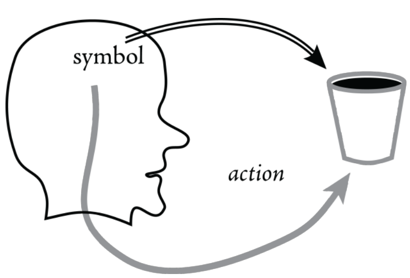


Figure 14 — Action

In the 1980s, when first working on these issues, I proposed a general framework in terms of which to analyse such norms, called **coordination conditions on content and causal connection** ("**C5**"). But this was more desiderata than theory, since I did not have enough ma-

[57] For example, to commercial software.

chinery to spell out any additional norms in much intellectual depth. What was needed is what we will address in the second stage of our project: generalisation. That is, we need to consider in detail what sorts of technical generalisations and decisions are required in order to develop this general picture into anything approaching a workable, comprehensive account.

6 Generalisation

No unique set of generalisations is required in order to do justice to participatory systems. Too many consequential subtleties branch out in too many entangled directions to permit accurate cataloguing. Moreover, to do real justice to embodied cognition ultimately requires starting over—building the entire account from the ground up, based on new metaphysics. Still, laying out some of the adjustments and alterations to the traditional conception of representation is a rhetorically and pragmatically instructive exercise. Among other things, it goes some way to illustrate the sorts of issue that a more radical reconstruction will have to face.

In this section, in this spirit, I'll mention a dozen or so such generalisations (see figure 15), grouped into three rough classes:

26. **Participatory:** having to do with the fact that systems "occupy," "inhabit," or are "situated in" their worlds;
27. **Ontological:** having to do with the nature of those worlds that systems inhabit (and the material they are made of); and
28. **Normative:** having to do with the nature of the governing norms to which intentional systems are held accountable.

In a sense, all three depend on a prior, more thorough-going generalisation that permeates everything we have done. I will call it **embodiment**: a recognition that representational systems, and the worlds they inhabit, are constructed from concrete, physical stuff. As we have seen, this fundamental embodiment establishes the powers and limitations of cognitive systems, and undergirds the constituting dialectic between what is and what is not effective. I call it a generalisation of logicism, not just a reconstruction, for reasons that emerged in the last section. Once the foundational conception of traditional logic—especially its bivalent emphasis on formality—is understood concretely, as suggested in the last section, a radically more general picture of representation is unleashed than is traditionally imagined.

But embodiment alone is not enough.

The first "participatory" group of additional generalisations includes several features (besides embodiment) that have been

Generalisations

A · Participatory	
1. *Interaction*	· Engagement between system and its environment
2. *Embeddedness*	· Syntactic and semantic domains overlap (limit: fuse)
3. *Context dependence*	· (Weak) Interpretation dependent on context of use
4. *Involvement*	· Orthogonal inside/out & symbol/referent boundaries
B · Ontological	
1. *Entanglement*	· Representation and ontology inexorably interrelated
2. *Nonconceptual*	· World doesn't come "pre-parsed" into objects, properties, relations, and other "formal" categories.
3. *Abstraction*	· Commonsense ("natural") ontology—objects, properties, etc.—require abstraction over world's basic, messy, "non-conceptual" structure
4. *Deixis*	· Local, incremental, differential character of physical law implies that content is deictic or indexical
5. *Context dependence*	· (Strong): Meaning dependent on context of use
6. *Features*	· Temporally-indexed features ("it's raining") as a simple form of abstraction (« predicate-object)
7. *Non-discreteness*	· World neither first- nor higher-order discrete
C · Normative	
1. *Ends (telos)*	· Generalisation of logic's traditional pair: *Soundness*: Wanting what you get *Completeness*: Getting what you want
2. *Dynamics*	· Interdependence between statical (truth, reference, etc.) and dynamical norms (what to do, how to live).
3. *Objectivity*	· Commitment to the existence of the world

Figure 15 — Generalisations to Logicist Representation

touted as characteristic of "situated cognition." Embracing them will thus give us a handle on many of the traits listed in §■■ as distinctive of an embodied view. The second and third groups, having to do with ontology and normativity, implicate issues that have not received nearly as much explicit attention, at least to date. But these concerns are beginning to make their presence felt, and (as I hope to show) in some ways they cut deeper into the fabric of an embodied perspective than the merely participatory. Dealing seriously with them adequately requires a more extensive treatment than I can afford here; I will be able to give them just some very introductory remarks.

6a Participation

As we saw, logicism distinguishes two realms: a syntactic realm (*S*), of representational vehicles (such as expressions), and a semantic realm (*D*) or task domain, containing the objects or entities that the representational vehicles are about. Moreover, the governing architectonic took all causal (effective) transitions ('→') to be inferential, understood as an (inferential) relation on the syntactic realm.

INTERACTION

Perhaps the most widely touted characteristic of embodied cognitive systems, taken to distinguish them from logical inference schemes, is the fact that they interact with their environments. So a natural first way to generalise the logicist framework is to license causal connections ('↔') across the *S-D* boundary. This move captures an extremely common intuition, that underlies the very notion of perception, and is implicit in such ubiquitous ideas as (i) the standard conception of sensors and effectors; (ii) the "robot reply" to Searle's Chinese room, (iii) the virtual platitude that our senses "connect us to the world," (iv) Harnad's proposal for a generalised "Total Turing Test" to assess intelligence, (v) the imaginative force of a "brain in a vat," thought to be disconnected from any possible semantic realm—and so on and so forth. I will dub it **interaction**: a proposal that effective operations not be limited to system-internal transitions, but include causal coupling across the boundary between systems or agents and the environments they inhabit.

EMBEDDEDNESS

But though it extends pure embodiment, INTERACTION is still too weak. In fact the formulation just given—essentially, of an "inner world" of symbol or thoughts, and an "outer world" that the symbols or thoughts represent, to which it is connected by sensors and effectors—is a great example of the limitations of a purely amalgamationist approach. For in the act of valorising causal traffic *between realms*, the proposal shares with logic the presupposition that the realms are *distinct*—that the world or task domain that the agent is reasoning is wholly "exterior": outside or beyond the internal realm of mental activity.

The error stems from sundering agent and world. Once the two are conceptually separated, no amount of mere causal coupling is strong enough to glue them back together again.

An example of the difficulties that causal coupling with the environment does not repair arise up in what I have (in another context) called "non-effective tracking": the maintenance, in time, of a dynamic representational state that represents an external on-going process to which an agent is not coupled. This is the sort of thing a creature might do in "mentally tracking" a moving object while it is occluded from visual sight—or that we ourselves often do (badly), after someone has called from the airport and said that they will be home in half an hour, as we imagine them getting into the car, turning onto the freeway, getting to the right exit, etc., so as to be able to predict their arrival. What is striking about such cases is that they involve *non-causal* (i.e., non-effective) *coordination between realms*. In particular, the governing normative conditions on non-effective tracking exploit the fact that the passage of time for an agent, and the passage of time in the agent's task domain, are *one and the same*. They aren't merely "in synch"—in the sense of being two things kept in step by causal coupling. They were never separated in the first place, in any way that would require their being brought back into synchronisation.[58]

Agents are not just *embodied*, in other words, in the sense of being made of concrete physical stuff. They are also *embedded*:

[58]This point must not be confused with the question of the relation between *representing time* and *represented time*. For any dynamic representation of a dynamic phenomenon, those two will be (at least logically) separable. In cases of both effective (standard) and non-effective tracking, the two are as a matter of fact (approximately) coincident: that is what makes them cases of *tracking*. But this issue—a special case of the relation between sign and signified—is orthogonal the relation between agent and world. (See the discussion of involvement, below.)

they live in, are made of, and dwell among the things that constitute their environment. We therefore need a second generalisation, which I will call **embeddedness**: a recognition that the syntactic or effective domain (the stuff of which the system is made, and the agent's "inner life"), and the semantic or task domain (the world the agent represents, the things that it cares about, etc.) will at a minimum overlap, and in the limit be the same.

EMBEDDEDNESS provides for various forms of *coordination* between the realms of representational activity and realms that that representational activity is about. Metaphysically, the point is that not all coordination involves causal or effective coupling.

A striking but familiar example of non-effective coordination is provided by clocks. Clocks are clearly representational: the arrangements of hands on their faces[59] represent what we might call *o'clock properties*: 4:01 p.m., 4:02 p.m., etc.—i.e., properties exemplified by passing metaphysical moments. Clocks were hard to build for exactly the reasons identified in the representational mandate: o'clock properties are indisputably non-effective..[60] It follows that concrete systems can only orient towards them by representing them—by exploiting something else that is effective, that is coordinated with what is not. The task for a clock (or clockmaker) is to exploit the effective properties of the inner workings (clockworks) in order to establish an appropriate relationship between those aspects of the hands that are effectively controllable (the position around the dial) and the non-effective temporal property thereby represented.

The normative conditions on clocks are given in the sidebar on p. ■■ (in brief, clocks are right when the property represented by the position of the hands holds of the metaphysical moment that it is). Needless to say, the temporal conditions on full-scale temporal reasoning and temporal consciousness will be radically more complex. For example, they will involve Husserlian issues regarding the intricate relations between the temporality of perceptual processes and the temporality of dynamic activity thereby per-

[59] I am considering analogue clocks here, though nothing hinges on that simplification.

[60] If "being 4:00" were effective, one could build an automatic kettle that put up water for tea by detecting that the passing moment was 4:00. But of course no such mechanism is possible.

Norms on Clocks

The norm governing the position of the hands is relatively straightforward: at any given moment t, the configuration of hands $\square_t$ should represent the o'clock property □ that is true of t. In a sense, a clock has to *track* the passage of time. But it has to track a non-effective property of the passing time; that is what makes the situation representational. In a sense, one can think of the task facing a clock as the dual of traditional inference: whereas inference, at least as traditionally construed, involves moving from one representation of a presumptively stable environment to another, clocks must do the opposite: maintain a stable (at the level of "meaning") relation to a changing environment. Taking the analog (continuous) case as an example, this leads to the following two "correctness conditions" for clocks

(1) correct-speed: $\dfrac{\partial \llbracket ◷ \rrbracket}{\partial t} = 1$ (2) correct-time: $\llbracket ◷ \rrbracket(t)$

ceived—intricacies that are necessary in order for systems to authentically perceive the world as on-going and dynamic. The now, the point is only that EMBEDDEDNESS will in general implicate complex forms of coordination and (potentially non-effective) relationality between the effective and the non-effective dimensions of the overlapped (or even unified) "syntactic" and "semantic " realms.

CONTEXT-DEPENDENCE · I

EMBEDDEDNESS opens up the possibility of understanding another of the prominent intuitions underwriting the "situated" movement in cognitive science: a recognition that the representational states of real-world systems are *context-dependent*. Context-dependence is not so much a fact of embodiment per se as it is a semantic consequence of this kind of embeddedness: the fact that material systems are often *located* in their worlds— situated in specific circumstances, in ways that have consequences for their semantic interpretation.

I won't say much about simple context-dependence here, of the sort that characterises indexical expressions (*I*, *you*, *we*, *here*, *now*, etc.), because it has been so extensively studied. If we make a distinction between a symbol or term's *meaning* and its *interpretation*—where meaning is taken to be approximately the stable, single "rule" or regularity associated with all uses of the term, of the sort that a person acquires when they "learn" the term, and *interpretation* is the context-dependent referent or semantic value that each utterance obtains, on any particular occasion[61]—then this form of context-dependence can be understood as a phenomenon of context-dependent variation for symbols or representations with context-independent meanings. I will name this widely--recognised third participatory generalisation **context-dependent interpretation**. (A more radical form of context-dependence, involving con-text-dependent meaning, will come up in the second group.)

INVOLVEMENT

We still aren't done. Even the conjunction of embodiment, embedded-ness, and context-dependent interpretation does not go deep enough. They potentially (but misleadingly) preserve a sense that cognitive creatures "look out" onto the world—that all the semantic relationships originate in heads, and are directed "agent-external" to the world that we move around and change and dwell in. So in the table I have listed a fourth and final participatory generalisation, labelled **involvement**.

The aim of INVOLVEMENT is to recognise that semantic directedness ('⇒') and causal coupling ('↔') are *orthogonal*.

To understand what this comes to, note that representational (computational, logical) systems can be understood in terms of two distinctions or "boundaries":[62]

---

[61]So when you use the term 'I,' the interpretation is *you*; when I use 'I,' the interpretation is *me*. Thus our interpretations differ. If each of us meet someone we have never met, and they use 'I,' the interpretation is *them*—a new interpretation, one we have never before encountered. But we are not mystified, when hearing that new person say the word, because we know *what the word means*. (Thus meanings can rather glibly be viewed as a rule or regularity of the form lcontext.interpretation.)

Put it this way: dictionaries give meanings, not interpretations. That is why there is only one entry under the word 'I'; not ten billion, one for each person in the universe.

[62]The discussion is this section is a radically brief summary of some of the

29. A **semantic** boundary, between a representational vehicle and its referent (what it is normatively oriented towards); and
30. A **physical** boundary, between a system's insides and outsides.

Given these boundaries, one can then identify a pair of theses on which the classical model is based:

31. An **alignment** thesis, claiming the 2 boundaries line up; and
32. A thesis of **isolation**, claiming that the 2 (allegedly-aligned) boundaries are something of a moat (causal, logical, explanatory).

Jointly, these two theses entail that all of the symbols or representations lie within the system, and all of the referents are to be found on the outside (roughly what was suggested in the original logicist figure 3). What the INTERACTION generalisation does (the idea underlying the robot reply to Searle, the idea of extending an inferential model of cognition by adding sensors and effectors, etc.) is to deny **isolation**: the idea that transactions the boundary between the symbol system and the "outside world" is closed. As far as it goes, as we have seen, that is surely correct, for any plausible notion of an embodied cognitive creature. But what that analysis fails to recognise is that **alignment** is false as well: *the boundary between symbols and their referents, and the boundary between the inside and outside of a system, are orthogonal*.[63] Not only are there (in the real world) internal symbols with external referents, as imagined on the classical image (thoughts about a friend or enemy), but also internal symbols with internal referents (introspection and self-knowledge), external symbols with internal referents (the advice of friends and psychiatrists), and external symbols with external referents (roads signs directing you to the airport). Plus, there are causal transitions *between and among all of*

results of AOS·II.

[63]This is one of the primary results of the analysis of formal symbol manipulation «ref AOS·II».

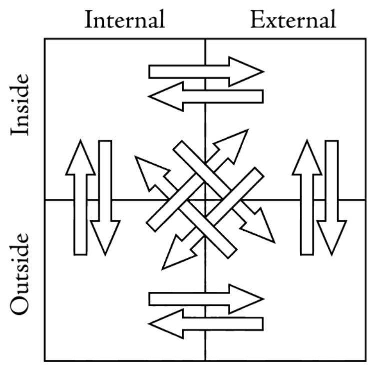


Figure 16 — Participatory Transactions

these four kinds. Figure 16 gives an indication of the structure of this terrain, with four kinds of representational example, and sixteen different types of thereby engendered causal transition. For example, a plausible normative constraint on the process of *reading* might be not that one *represents* the text being read (as many classical analyses suggest), but rather that one "internalise" the text, by constructing internal representations whose semantic content is the same as that of the external representations with which one is interacting (see figure 17).

I make no claim that even these four mandates are enough to ensure the kind of "being in the world" that everyone takes to be constitutive of a situated, embodied view. But as we will see, they are enough to cause profound consequences to the theoretical frameworks in terms of which to understand the overarching representational mandate, of local effective processes governed by overarching but non-effective norms.

6b Ontology

One feature of the separation of realms (as we have seen) is characteristic of the logicist picture is its foundational ontological presupposition that the character of the semantic realm (what objects, properties, relations, etc., constitute it) and the character of the syntactic or effective realm are established independently—and also "extra-theoretically," in the sense of being assumed to be fixed, prior to and independent of the characterisation of the agent as cognitive. This structural character reflects, in technical guise, a guiding simplification that undergirds much of the analytic philosophy on which traditional cognitive science rests: an assumption that the theory of representation

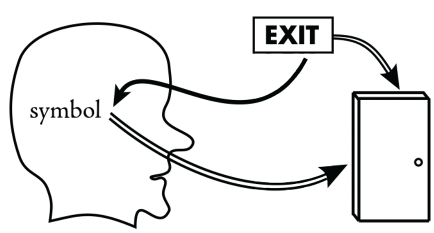


Figure 17 — Norms on Reading

(how creatures take the world) and the theory of ontology (what the world is like) are *independent*.

If any general theme underlies the sorts of shift I am recommending as necessary in order to do justice to an embodied perspective (beyond the explicit emphasis on concrete materiality, participation, etc.), it is a move to dismantle and defuse all sorts of sharp independence underwriting traditional logicism.[64] No independence goes deeper than the just-alluded-to separation of representation and ontology. More strongly, though nothing we have said so far argues for it, I want to start with a move to generalise—to be honest, to deny—this metaphysical assumption. That is, I want to endorse what I will call **entanglement**: a recognition that representation and ontology interrelated—that how we represent the world to be, and what the world is like that we thereby represent, cannot be given independent explanations.

Three immediate comments.

First, it is critical to realise that embracing ENTANGLEMENT is (in and of itself) no endorsement of radical solipsism, idealism, or any other metaphysical stance that *fuses* representation and represented. As mentioned earlier, the idea that if two things are not the same, then they must be independent, is *exactly the kind of ideological commitment to independence and sharp distinctions that I am at pains to deny*. All that is claimed, by the ENTANGLEMENT mandate, is that a fully general (rehabilitated) approach to representation must allow representation and ontology to be at least partially interdependent.

Second, in spite of those very mildly conservative observations (essentially, recognising that some vaguely realist intuitions must or at least may need to be retained), I would be the first to admit that dismantling the analytic assumption that ontology and representation are independent is an *extraordinarily expensive move*. The theoretical consequences are staggering, with implications that shake the very foundations of what it is to do science, to give a theoretical account, to know. Just a few of the most evident consequences will be touched on below; but it is in recognition of

[64]In AOS I take this pervasive sense of independence and sharp distinctions to be the deep meaning of formality (one that reaches much further into the conceptual bedrock of logic, science, etc. than the more superficial positive and negative readings we talked about in §■■).

sequences will be touched on below; but it is in recognition of the full power of ENTANGLEMENT's implications that I said above that the only ultimately palatable way to do justice to embodied or situated cognition will require complete metaphysical overhaul.

Third and finally, in spite of the expense, it is impressive how many different currents and voices in contemporary cognitive science argue, implicitly or explicitly, for exactly such a loosening of the traditional assumption, and a potential melding or meshing of representational and ontological concerns.[65] What these arguments and these voices together imply, I believe, is that embracing ENTANGLEMENT—and recognising that cognitive science must take on blatantly metaphysical and ontological issues—is the most urgent intellectual issue that faces current cognitive science and philosophy of mind.

ENTANGLEMENT is such a strong mandate that it is perhaps almost fatuous to list any others. But in the table I have enumerate five more, to give a flavour of the sorts of ontological task that await us:

33. **Entanglement:** allowance for the fact that representation and ontology inextricably interrelated
34. **Nonconceptuality:** a recognition that the world does not come "pre-parsed" into the theoretically-familiar categories of objects, properties, relations, sets, states of affairs, propositions, possible worlds, and the like, as assumed in the logicist tradition.

    I label this NONCONCEPTUAL because it is a theme of the literature on nonconceptual content[66] that *conceptual* content is content that takes the world to be structured in this way (i.e., in terms of objects, properties, relations, etc.), opening up the possibility that nonconceptual content might be content that takes it in some other way. I have argued elsewhere[67] that the warrant for nonconceptual content is ultimately ontological, not epistemological—i.e., that the raison-d'être of nonconceptual

[65]«Cite Thompson, Varela, & Rosch; Lakoff and Johnson; Haraway; Lave; Chemero; Cussins and the non-conceptual literature; Objects, etc. »

[66]«Ref Evans, Cussins, Bermudez, Peacocke, etc.»

[67]«Ref "The Nonconceptual World"»

content lies ultimately in the *world*. It is reality that is not aboriginally conceptual, that is (i.e., that is not structured in the way in which conceptual content takes it to be).

Many will assume that this ontological version of non-conceptuality is a species of *non-realism*. But that is a legitimate label only on an assumption that the conception of the world as conceptually structured (i.e., as consisting of objects, properties, relations, etc., as classically imagined has some incredible kind of pre-metaphysical claim of priority. If one assumes, as I do, that the world is not autonomously so structured, then it is *conceptual* content that flirts with be irrealist, not nonconceptual content.

35. **Abstraction:** a recognition that the commonsense ontology represented by conceptual content, as described above, involves profound capacities for *abstraction* in cognitive creatures, which cognitive science needs to explain.

36. **Deixis:** a recognition that fundamental facts about the nature of physical existence, having to do with the incremental, differential character of physical law (i.e., the ontological facts that warrant our expressing the laws of physics in the form of differential equations) imply, as a consequence, that any physically possible form of representation will (at least in any simple form) be originally *deictic* or indexical.

    It is not first-person content that is mysterious, from a physicalist point of view, in other words. Rather, the mystery—the theoretical puzzle that challenges cognitive science—is how concrete agents can achieve third-person reference or content.[68]

37. **Strong context-dependence:** a recognition of the possibility that not just interpretation (in the sense described in §■■), but *meaning as well*, may be context-dependent. It is not just that different utterances of 'I', 'now', etc., have different interpretations on different occasions, in ways governed by a stable, context-independent regularity. Rather—at least in general—it may be that *even what ordinary words (or cognitive symbols) mean* may depend on

[68]«Ref "Who's on Third?"»

contingent or circumstantial facts about the situation in which they are used.

I call this form of context-dependence *strong* because, as with the other ontological generalisations being listed here, its implications for metaphysics are strong. But there is nothing intrinsically contradictory to the idea—or, necessarily, irrealist. Some may assume that if meanings are not entirely fixed, then they must be completely fluid—thereby taking leave of any possible realist commitment (and in the process vitiating any talk of world-directed norms). But to think that is merely another instance of a black-and-white assumption. There is no logical reason why the meaning of words cannot (in general) be partially fixed, or at least relatively stable, but nevertheless be partially bent and shaped as well, by contingencies of a discourse situation.

38. **Features:** a recognition, in consort with above-mentioned suggestions that commonsense ontologies of objects, properties, etc., may involve sophisticated conceptual abilities, that a simpler way of registering the world, in terms of (temporally-indexed versions of) what Strawson has called *features*, more like property- or relation-instances than anything with the full logical structure of objects and properties, may figure in nonconceptual representational schemes.
39. **Non-discreteness:** a recognition that the structure of the world may not in general be *digital* or *discrete*, either in the ordinary sense (in which people think computers are digital), or in the sense that Haugeland has called "higher-order discreteness"[69]—a kind of clean separation of *concepts* and *properties* that is familiar in mathematics and science, but seems radically unlikely to hold of such everyday notions as (for example) confidence, ego, chutzpah, bravado, arrogance, braggadocio, etc.

Needless to say, these are mere telegraphic labels of subjects that would require vastly more space even to convey an adequate sketch of. But they illustrate the sorts of ontological reconfigura-

[69]«Ref "Analog and Analog"»

tions of the world that we, as cognitive scientists, are going to have to deal with, if we take the embodied, participatory stance seriously.

6c Normativity

We have said little, so far, about normativity. But as mentioned in the discussion of logic, to enter the realm of representation—description, language, interpretation, truth, etc.—is to enter a world of phenomena governed by asymmetric (paired) evaluative predicates: true vs. false, good vs. bad, working vs. broken, beautiful vs. ugly—where in each case one option is *better*, or *more worthy*, than the other. Accurate descriptions are better than inaccurate ones; information is better than misinformation, helpful behaviour is better than unhelpful behaviour—and so on. In fact one very plausible definition of intentional systems is that they are just those systems that are subject to norms.

The question is what to say about how to generalise the normative structure implicit in logicism in such a way as to incorporate the full range of norms that are appropriate to embodied cognitive agents.

For starters, we can generalise soundness and completeness in terms of a more general characterisation of **ends**. As described above, once states of a system can both engender causal consequences, in virtue of their effective structure, and stand in some kind of semantic relation to (potentially distal) states of affairs, the issue arises of whether, if an operation happens, or behaviour takes place, the result does or does not meet any applicable governing norm. There are, in general, two ways to fail. This leads to a natural reconstruction of the two traditional norms on inference:[70]

40. **Soundness:** wanting what you get
41. **Completeness:** getting what you want

The more substantive question has to do with what it is you want.

To get at this, consider logic's emphasis on truth. Truth is justifiably famous—but not particularly general. Within the logicist

[70]This informal but perspicuous formulation is due to John Etchemendy.

framework, moreover, it has been treated as a **static norm**, in the sense of applying to (passive) *sentences* or *claims*—i.e., to *states*.[71] Full-blooded intentional systems, however, are *dynamic*; hence governed by **dynamic norms**—norms that govern process.[72] In logic, the operative dynamic norm is derivative—defined in terms of a static norm. Reasoning, deduction, inference to the best explanation, etc., are all mandated to *preserve* or *produce* truth or explanation, where it is (critically) assumed that what it is to be true, and what it is to be an explanation, can be defined independently of, and prior to, the processes of their preservation or production.

This explanatory strategy—of starting with a (presumptively autonomous) static norm, and then defining dynamic norms in terms of it—has been picked up by other intentional sciences. Economic models of rationality and decision-making, for example, often use the dynamic norm of *utility maximisation*—where utility is (once again) presumed to be static, prior, and autonomous. But the general strategy of defining dynamic norms in terms of static norms doesn't generalize. And no computer scientist believes it. On the contrary, what practical experience with computing has taught us is that you it is vastly more general to proceed in the opposite direction: taking the semantic content (meaning) of a symbol or expression or data structure to be determined (even to exist) depending on *how it is used*—i.e., on the role it plays in the overall system of which it is a part. Rather than *define dynamic norms in terms of static ones*, that is, programmers *define static norms in terms of dynamics ones*—in a (perhaps unwitting) endorsement of the Wittgensteinian maxim that "meaning is use." And so this I have listed as our second normative generalisation: that we shift our original explanatory dependence from static to dynamic norms.

If we get our static norms derivatively from our dynamic ones, where do we get the original dynamic norms? What are they like?

---

[71]By static norms I don't mean norms that don't change, over time; evaluative metrics on book design, or on human beauty, may evolve considerably, but would still be counted as static, on my typology, because what they are evaluative predicates on—books or motionless bodies) are essentially static things.

[72]«Say: should be (or change to): 'statical' and 'dynamical'»

What governs, what puts value on, what evaluates, the use—i.e., the life and times, the activity—of general intentional processes? Though the question isn't usually asked so baldly, a variety of alternatives are being explored in contemporary cognitive science. But one dynamic norm is currently receiving by far the most scientific attention—in cognitive science, ALife, evolutionary epistemology, research on autonomous agents, and biology: **survival**.

It is clear how to get a norm out of survival: a process or activity is deemed *good* to the extent that it is *adaptive*—i.e., to the extent that it aids, or leads to, the long-term survival of the creatures that embody or perform it. This idea of resting normativity on evolution is seductive. It has been used to define a notion of *proper function*, for example, in terms of which to decide whether a system is *working properly* or is *broken*. Thus the *function* of the heart is to pump blood, and not to make a "lub-dub" sound, because hearts were evolutionarily selected for their capacity to pump blood, not for their sound-making capabilities. Similarly, the function of sperm is to fertilize eggs because that is why sperm have survived (even if only a tiny fraction of them ever serve this function).

Most interesting for our purposes, however, is the use of this same idea to define semantic content (meaning, reference, representation, truth). The representation in the frog's eye *means* that a fly is passing by, some people claim, because it leads the frog to behave in an adaptive way towards that fly (namely: to stick its tongue out and eat it) in a way that contributes to the frog's (not the fly's) evolutionary success. Similarly, the shadow on the ground conveys information *about* the hawk in the sky to a mouse just in case it plays an evolutionary adaptive role of counterfactually covarying with the presence of hawks in a way that allows mice to escape. That is, modern philosophy of mind has begun to change from logic, in taking the static norm of reference and truth to derive from the dynamic norm of leading to an adaptive or evolutionarily successful life.

Have we reached the end of the line? Will evolutionary survival be a strong enough dynamic norm to explain all the norms that apply to cognitive agents: justice, altruism, authenticity, caring, freedom, and the like? Personally, I doubt it. But in a way that is just the point. For what is at stake, for cognitive science, is not what will ultimately sub-serve the norms we need in order to understand human activity, but to understand *what the dynamic norms are in terms of which human activity is conducted and understood*. And that, I hope, is obvious: dynamic norms on human activity govern *what it is to live*—what it is to live well, to do good, to be right. That is: **ethics**. And not just ethics, but whatever governs whatever you do: ethics, curiosity, eroticism, the pursuit of knowledge for its own sake…and so on and so forth, without limit.

In sum, taking on full-fledged dynamic normativity is an unimaginably consequential move. It implies that any fully rehabilitated account of representation—any transformation broad enough to incorporate arbitrary embodied and embedded intentional systems, and thus to treat meaning along with matter and mechanism—will also, thereby, **have to address mattering as well**. Put it this way: in spite of logical practice, it won't generalise to bite off *truth* and *reference*, and glue them, piecemeal, onto physical reality, without eventually taking on the full range of other norms: *ethics*, *worth*, *virtue*, *value*, *beauty*. By analogy, think of how computer science once thought it could borrow *time* from the physical world, without having to take on *space* and *energy*. It worked for a while, but soon people realised what should anyway have been predictable: that time is not ultimately an isolable fragment—not an "independent export"—of physics. By the same token, it would be myopic to believe that the study of intentional systems can be restricted to some "safe" subset of the full ethical and aesthetic dimension of the human condition—and especially myopic to believe that it can traffic solely in terms of such static notions as truth and reference, or limit itself to a hobbled set of dynamic norms (such as survival). To believe that would be to be an ostrich, not a hero.

Moreover, to up the ante (in case this all seems too mild), something else, if anything even more expensive, is implied by these same developments. (Moreover, this is where the story starts to fit together, though it is also what mandates the development of new metaphysical foundations.) I said above that the classical model assumed that the meaning of symbols and representations could be assessed in terms of the objects and properties in the world that they corresponded to, independent of how those symbols and representations were used. But I also said, in the discussion of ontology, that many modern cognitive scientists no longer believe the classical model—in part because the physical world does not supply the requisite objects. That means, as we have already admitted, that it is incumbent on a theory of representation to explain the objects that figure in the (conceptual) content of a creature's representational states. What we didn't say in that ontological discussion, however, is that those objects are to be explained in terms of the normative structure governing the representations whose contents contain them. And those norms, we have just admitted, are ultimately grounded on *dynamic activity*.

It follows that the *material ontology of the world*—what objects and properties there are, for a given creature (not just what objects and properties the creature *takes* there to be, but what objects and properties there *actually are*, in the world, for that creature)—will, on the generalised account, be seen to be a function of that creature's projects and practices. For high-level social entities this isn't surprising: date-rape didn't exist, I take it, for the aboriginal singers of the Australian song-lines; the strike zone (a favourite object) isn't part of the furniture of the world, for earwigs. But the present claim is more radical: it suggests that what is true for date-rape and strike zones is true for food, clothing, rivers—perhaps (who knows?) even for the number four.

Ontology is inextricably linked to epistemology, in other words; that much we said with ENTANGLEMENT, above. What we are adding, now, is that epistemology is inextricably linked to ethics. These are conclusions I am happy with; but they are nothing if not strong. What is striking about them in the present context is that we have come to them by making two seemingly innocent moves: (i) by understanding that material ontology involves conceptual abstraction; and (ii) by giving dynamic norms explanatory priority over static ones.

We can summarise this conclusion etymologically.

A material object is something that matters

It must matter, in order for the normative commitment to be in place for the objectifying creature to take it as an object: to be committed to it as a denizen of the world, to hold it responsible for being stable, obeying natural laws, and so forth—and to box it on the ears, when it gets unruly. It is no pun, in other words, or historical accident, that we use the term 'material' as a term for things that are concrete (made of "matter") and also as a term for things *that are important*—as in 'material argument,' or 'material concern.' In fact that is one way to see where the embodied cognition movement is headed: whether it knows it or not, it is going to have to heal the temporary rift that for 300 years has torn matter and mattering apart.

7 Application to embodied cognition

One task remains. We need to understand how our proposed rehabilitated model deal with embodied cognition. That, after all, was our original goal: to combine the best in representational and nonrepresentational accounts, in order to avoid the fundamentalist excesses of figure 1.

Three preliminary remarks.

First, it is important to be clear on the question being asked. Many discussions of the relation between "new" and "old" cognitive science compare a proposal for a new "embodied" approach to representation as traditionally conceived. Thus van Gelder and Port contrast a dynamical systems approach, which they recommend, to their conception of the classical model (which they call "computational"), which they criticise. That is not the contrast I am addressing.[73] Rather, setting aside any vestige of the classical view, now, I want to understand the relationship between:

42. Various proposed *non-representational alternatives*, of which dynamical systems theory is one candidate, though there are others; and
43. The rehabilitated conception of representation being developed here.

For what kinds of system is each framework most appropriate? What kinds of insights and understandings are expressible in each framework's terms? What kinds of behavior warrant the admittedly more complex analysis provided in terms of a reconstructed notion of representation? How well does the rehabilitated notion of representation deal with the ■■ characteristics cited in §■■ as distinctive of the embodied view? Those are the sorts of things we want to know.

Second, it's not 100% clear what a "dynamical system" is. At the most general level, dynamical systems theory is a body of mathematics, applicable to any situation in which a system which can be described in terms of temporally-varying instantiations of

[73] I am especially not interested here in the issue of whether their view can legitimately be called computational—which I think it cannot.

measure properties—causal, semantic, emotional, whatever. By itself, that is, nothing in the term "dynamical system" necessitates the characterised properties being in any sense physical or effective. Thus a committed Cartesian could talk about God's waning love in dynamical terms (figure 18). I take it that the presumption in cognitive science, however, is that a dynamical systems account of a system's behaviour is understood to be a description of its causal (effective) behavior. That is what I will assume in the following.

$$\frac{\partial(\text{love-of-God})}{\partial t} < 0$$

Figure 18 — Dynamics

Third, a reminder about the (non-effective) nature of semantics. As we have said, it is a something of a metaphysical theorem—at least for physicalists—that systems work, mechanically, solely in virtue of their total effective (causal) structure: the effective structure of their internal arrangements, in interaction with the effective (causal) structure of the environment they are embedded in. This is a general claim, which holds of absolutely everything that there is: representational and nonrepresentational alike. So the following is not a possible objection to a representational (or other kind of intentional) analysis: "What do you mean, semantics? All that exists, for this system—all that there is—is a pattern of causal transitions and structural couplings to the immediate environment! How could there be anything else? Look at the system; attach any instruments you can devise. Show me something more than that!" This objection fails because it clearly assumes (e.g., in its reliance on *instruments*) that "all that exists" means "all that exists, *causally*"—all a meter could detect, all that involves the expenditure of energy, all that traditional sciences recognize as real. But all parties agrees with that; that was the exact import of our reconstruction of the negative reading of formality as a claim that semantics is not effective. We have already admitted that semantics cannot be detected by a (causal) instrument. To suppose that it *could* be would be to suggest that representation violates physicalism, which no one is suggesting.

Rather, what the representationalist (intentionalist) is claiming is something else: that an account of a system's local, causal interactions *does not exhaust the constitutive facts about that system*—the facts that would need to be accounted for by an explanatory theory.[74] For remember what we said about semantics: they operate as *non-effective governing norms*. In order to show that a system is not semantical, therefore, one must show that it is not so normatively governed. That is not quite as easy to do as a simple causalist might imagine.

7a First pass • Formal

The first thing to say is that the reconstructed representational account we are sketching is extraordinarily broad. Indeed, all it really comes to, so far, is that a local, causal, effective account must be given, of how the system works; plus a potentially non-local, non-causal, non-effective account of semantic interpretation; and that the two be tied together by constituting norms. By hypothesis, the view of dynamical systems we have endorsed is merely one way of giving the first of these: a causal account of behaviour.

Being a dynamically-described causal system, however, by itself has *no bearing whatsoever on whether the thereby-described system is representational*. That is because, from the point of view of pure mechanism, the new representationalism imposes no apparent constraints! Representation, as we said at the beginning, is (in its

[74]By analogy, think about all the possible cursor shapes that can be displayed on your computer. On most operating systems, cursors are arbitrary 16 × 16 bit binary patterns, which a program can set arbitrarily, so as to draw the familiar shapes we all know: arrows, hourglasses, cross-hairs, etc. Since there are 16^2 or 256 bits, each of which can be on or off, there are $2^{256} \approx 10^{77}$ different possible shapes—or about 100,000 times as many as there are electrons in the universe. Of these, we use a few hundred, or at most a couple of thousand.

Suppose one wants to provide a theory of cursors. One theory might simply say that cursors are 16 × 16 bit patterns, and describe how they are set and manipulated by hardware and software. In terms of the local pattern of causal behaviour, that account may be complete. But something may be left out. For example, suppose (falsely) that the only cursors that are ever drawn are shapes that resemble naturally-occurring artifacts. A full theory of cursors, therefore, would have to include a theory of *what it is to resemble a naturally-occurring artifact*. That additional theory would not be a theory that added or changed—especially "in the small"—how the cursor works, causally. But it would nevertheless reconstruct constitutive patterns of cursors that the purely causal story would not.

full potentiality) an extraordinarily broad notion. So the question on the table is going to boil down to the following: in what circumstances is it productive—valuable, explanatory, and *true*—not only to give an account of how a system works, mechanically, but to tie that (normatively) together with an account that interprets the system?

More strongly, the conception of representation that we have been developing was explicitly designed to include the ability to treat the sorts of behaviour that embodied cognition takes to be essential. In particular, consider the list of eight contrasting pairs of properties, listed at the beginning of §2g (page 10), of what distinguished embodied cognitive systems from classical (allegedly "computational") ones. Of these, the first—a shift from pure abstraction to concreteness, or an endorsement of the importance of EMBODIMENT—has not only been dealt with, but has underwritten the entire story we have been telling—about effectiveness, representation's raison-d'être, etc. Whatever else is true of our reconstruction, in sum, it puts concrete materiality squarely on center stage.

The third (I will return to the second in a moment), that the system NOT BE SEPARATED from its semantic realm—is part and parcel of what we dubbed a participatory view (cf. earlier remarks on perception, action, tracking, introspection, cross-cutting boundaries, etc.). Similarly for the fourth requirement, that a system be dealt with as ENGAGED with its environment. Finally, the fifth and sixth requirements—that we deal with DYNAMICS, and treat CONTINUOUS behaviour—have also been made room for (both were illustrated, for example, in the discussion of clocks, including the "clock" equation). Similarly, the sixth characteristic, that embodied systems be understood as CONTEXT-DEPENDENT, has been fully embraced. Not, let me hasten to say, that the postulated reconstructive framework provides theoretical tools for dealing with any of these aspects. On the contrary, tremendous work remains to be done to understand how to treat such features adequately. The point is only that there is nothing in a representational approach, per se, that stands at odds with any of them.

The eight listed characteristic was its selection of NAVIGATION, rather than deliberative, ratiocinative thought, as the paradigmatic "cognitive" activity of an embodied view. As I hope is clear, this is not a *requirement* of the reconstructed view; its aim was to be neutral on such decisions—providing the wherewithal to treat of both thought and navigation (and a host of other activities). So while the requirement is not exactly met, nevertheless I count this greater generality a feature. And in a sense the same is true of the previous three: while continuous, dynamic, context-dependent representations have been embraced, nothing prevents the treatment of discrete, static, or context-independent ones. This counts to the view's benefits: its aim was to be catholic, able to deal with the full range of possibilities, not to take an ideological stand on either side.

Turn then back to the second pair, having to do with the linguistic, explicit nature of the representational vehicles, on the classical side, which were rejected on the dynamical side. This is a somewhat subtler case. There are two distinct issues at stake.

The first has to do with how the reconstructive account deals with content. As indicated in the discussion of ontological generalisations, it is no part of representation, as we have reconstructed it, to be especially committed to explicit, conceptual, or linguistic content. On the contrary, we have made explicit gestures towards non-conceptual content, which stands as a strong candidate for a form of non-linguistic or non-explicit content. But as in the previous cases, the aim for the representational *framework* is for it to be neutral on the question—exactly so as to allow the theorist to explore different kinds.

The second issue does not have to do with content that is not linguistic, but rather with systems or behaviours that *do not have content at all*. That is, how are we to treat systems that (in spite of the breadth of our rehabilitation) are *genuinely non-representational*? It can hardly be counted against the rehabilitated account that (by itself) it does not deal with them; that was not its aim. For cognitive science, though, we do need to understand the powers and limitations of non-representational systems—which finally brings us back, full circle, to the first strategy mentioned at the very outset: of amalgamation.

7b Second pass • Substantive

Formally, we have concluded, nothing in the list of characteristics of embodiment militates against a representational account. But that is an admittedly thin result. After all, the rehabilitated account was expressly designed to accommodate this list. The remaining question is the substantive question: what (given that we have gone to all this work) does a representational account buy you—and when are such analyses warranted?

This, finally, is where the rubber meets the road.

Start with the most basic stipulation of the "embodied cognition" movement: that cognition has evolved in response to, and must be understood in terms of, the material conditions and capacities of the cognizing organism. *Start with the body.* The body is a mechanism. So am embodied approach must start with the mechanical—which is to say, effective—capacities of the organism. This much is gospel.

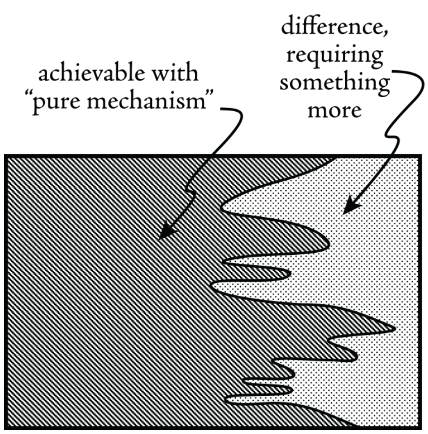


Figure 19 — Brooks' Challenge

By a **purely mechanical system** I will mean a systems whose constitutive regularities are exhausted in terms of the causal/ effective interior structure, and the causal/effective relations that it bears to its environment. Physics, and its immediate higher-level natural sciences, such as chemistry, thermodynamics, etc., I take it, study purely mechanical systems.[75] Dynamical systems theory, as we are characterising it, developed as a mathematical framework in terms of which to analyse the behavior of such systems. As Bechtel has noted, dynamical systems equations are in a sense covering law equations, more than mechanical accounts of how the systems work—a fact that will prove to be of some importance, in a moment—but for now we can continue to assume that the regularities that the dynamical equations account for are behavioral regularities, regularities that *have* an (immediate) causal explanation.

One of the insights of the embodied cognition movement—

[75]Literally: they study phenomena as purely mechanical.

reaching back as far as Raibert's pogo-stick robots[76]—is that we do need to understand bodies, and their natural dynamics, mechanically. As much was admitted on even the simple amalgamationist research strategy with which we started. Something else researchers have repeatedly discovered—epitomized in Braitenberg's book— is that as astonishing amount of behavior can be generated merely by placing a mechanism, of some functional or causal capacity, into a structured environment.

Moreover, it is not just that a great deal of behavior can be so explained, but for reasons of economy, evolutionary plausibility, and sheer good sense, it is best to try to explain as much behavior as one can, in this way. This strategy has been explicitly endorsed by Rod Brooks, who formulates it as something of a maxim:

> *Explain everything you can purely mechanically. Only use representation for the "residue"—for that last increment of cognition that cannot be explained purely mechanically.*[77]

This strategy is figuratively depicted in figure 19. The overall rectangle is meant to indicate the full suite of capacities required for general intelligence; the white central region indicates the range of capabilities that can be explained in purely mechanical terms.[78] The shaded re-gion—the difference between the two—is meant to indicate Brooks' "delta" or "residue"—the range of capacities that do require, for their deployment, representational powers.

The way we can get at our question, therefore—of what it is that generalised, reconstructed representation is good for—is to inquire about the nature of the white region, and the nature of the shaded "delta." That is, we face two questions:

44. What can be done with a pure mechanism?
45. What requires the additional resources of representation?

And finally, we are ready to reply.

The answer was implicit in §■■'s discussion of what is and what is not effective. Remember that the constraints of materiality or mechanism are the constraints of *physical being*. More particularly, they are the constraints of *effectiveness*—that was the

---

[76]«Ref»
[77]«Is there a quote I can use? Check his article in Mind Design II.»
[78]I am not making any supposition in the area of this inner curve.

ticularly, they are the constraints of *effectiveness*—that was the whole point of identifying effectiveness as a critical subject matter. But what is effectiveness like? And what can it do? Well, among other things, as we saw, effective properties are local properties, due to the fundamental locality of physical law. That leads to the following general claim. What can be done, purely mechanically—and what can be explained, therefore, purely mechanistically—are two things:

46. Regularities having to with effective properties of the system itself (i.e., its inner constitution), and
47. Effective properties of the environment in which the system is deployed.

But what are the effective properties of the environment? They, too, are intrinsically local. It follows from the nature of physical law, that is, that:

> *With respect to pure effectiveness, what a system can deal with, mechanically, is its own (internal) effective state, and whatever impinges on its surface.*

The picture, in other words—and by no means is this surprising—is very much along the lines of that of Maturana and Varela's **structural coupling**. A system (according to them) consists of an organised amalgamation of parts, whose effective properties come together to give the system some behavioral repertoire, which is then "coupled" into the immediate environment. What the system "does," as a result, is: (i) potentially adjust its effective internal arrangements (i.e., adjust its "state"), and (ii) potentially adjust (push and pull) on the impinging lamina of forces and fields that press in on its surface. Except that the "pushing" and "pulling" are symmetrical: neither affects the other any more than it affects them. This is why the Maturana/Varela image is apt: a system adjusts its internal state, and is "structurally coupled" to its environment. Re pure causality or pure effective mechanism, *that is all*. That is all that is going on. And given our background assumptions of physicalism, so long as we focus only on causal aspects of the system, that is all there is to *any* system. The locality of physics prohibits more.

We can summarise this as something of a maxim:

> *The life and times of a purely mechanical system is wholly and entirely exhausted by what happens to its internal effective arrangements, and what happens at its immediate periphery.*

That's all.

Two points.

First, not only purely mechanical systems, but all systems *qua* mechanical systems, *are always 100% coupled to their environments* (in this sense). The nature of the environment may change. But whether the system is coupled to it may not. The reason is simple: physics does not allow disengagement.

Second—and this is what matters most—it is not the *world* that such systems are engaged with. Rather, what they are coupled to a 3dimensional laminar surface of forces and fields, pokes and pressures, that is literally and constantly in the system's face. Qua physical mechanism, that is, there is no door over there across the room, no room downstairs, under the floor, no food around the corner in the cafeteria, no warm and snuggly bed, back home. *Those things are distal.* And distal things are inaccessible, as such, to pure mechanism.

So we have the answer to Brooks' paired questions.

Start with the first. What can you do, purely mechanically? The literal answer is this: you can deal with what is purely effective. What is purely effective is constrained, among other things, to be what is entirely local. So what you can do, purely mechanically, is (at most) deal with what is purely local—locally onboard you, or locally right there at your periphery. You can't even deal with *everything* that is local. Only with that vanishingly small percentage, overall, of those local properties that happen to be effective.

7c The role of representation

Is dealing with what is local, and effective, the sum total of intelligence? No. Part of what it is to be a cognizing creature is to inhabit, live in, deal appropriately with, *the world*. Perhaps not the entire world, to start with—maybe just a bit of the world, around your natural habitat.

What representation is *for*, therefore, is to *deal with the world*. To know that there is a universe out there! To deal with what is distal—with things not at your immediate proximal periphery, but some distance away: across the room, down the street, around the corner. To understand that things don't cease to exist, outside the door, around the corner…at the limits of your senses.[79]

Look around you. What do you see? It's amazing—you see chairs, tables, people, perhaps; maybe a mountain or a stream. Perhaps the inside of a car. None of these things is at your periphery. In fact—stunningly—you *can't* see anything, if it is pressed right up against your eyeball. That is because the content of our experience is the world *is at the end of those double-tailed arrows*.

Not only is experience representational, in other words, but the content of experience is invariably something we are *not* coupled to.

The final answer of this long journey, that is, is something of an ironic opposite to that proposed by Maturana and Varela. Qua pure mechanism, they are right: what it is to be a mechanism is to be structurally coupled to a manifold surround. But that is not what the world is *like* for a cognizing creature. Re what it is *like*, logicism was closer to the answer. What you represent—what you think about—is *not* what you are coupled to ('$\leftrightarrow$'), not what is effective, but what you are semantically and normatively oriented towards ('$\Rightarrow$')

What the world is *like*, that is—for us, and for any system that represents—is *how we represent it as being*, where to represent is to exploit the plasticity of that same causal coupling and locally impinging surround, so as, without violating the pre- and proscriptions of physics (it really is a magic trick) to stand in appropriate relation to what one is *not* causally coupled to. Moreover, it is exactly that fact that we are oriented towards the world, conscious of the world, committed to the world, that makes us intelligent.

[79]As Strawson put it: "How do we know that our senses fail, rather than that the world fades?" '«Ref»

D · Digitality

— Were this page blank, that would have been unintentional —

8 — The Devil in the Digital Details

Digital abstraction and concrete reality†

1 Introduction

It cannot be denied that computing and information technology have had—and are continuing to have—a monumental impact on the creative and documentary arts. The only plausible precedents are such epochal transformations as the invention of writing, or the press.

Less clear, however, is *what it is* about computing that is responsible for this upheaval. What makes information technology *special*, so that rendering art and music in computational terms wreaks such havoc with our understanding of *identity*, *materiality*, *ownership*, *originality*, *performance*, and perhaps even *value*?

One obvious place to look is to the notion of **information**. But information a curious notion. From one perspective, the concept of information can seem so vapidly general as to verge on the banal. Yet at the same time, specific technical notions of information are being intensely researched in as many as a dozen fields. I do not deny that the role of information in the arts a critical topic—one that would warrant its own book or conference. Still, information alone cannot explain our current predicament; no one could be so arrogant as to claim that information was invented in our own era. In one form or other, information has been around for millennia—perhaps since the dawn of time.

Another idea about what is fueling the computational revolution, and its impact on our lives, is the notion of **digitality**. Intuitively, it makes sense to ask whether digitality might be key to what makes information technology special, since, at least in

†An earlier version of this paper was published … «Ref: la Calcografía Nacional Simposio internacional arte gráfico y nuevas tecnologías, 2002.»

automatic machinery, digitality seems genuinely novel. The notion of digitality has also captured the public imagination. Whereas talk of the "Information Age" seems almost passé, scholarly texts and popular accounts still trumpet the coming of the "Digital Age." Best-sellers such as Negroponte's *Being Digital* are just the tip of the iceberg.[1]

In the creative arts, digitality's impact has been immense. But scale of impact has not been matched by depth of understanding. It is widely agreed that the *rendering into digital form* of images, sounds, records, and ideas has unleashed considerable conceptual confusion. Somehow or other, in ways we need to understand, digital images, recordings, and texts break the bonds of time-honored norms of *identity*, *production*, *ownership*, *reproduction*, etc. If we can understand what it is to be digital, therefore, we should thereby get a leg up on disentangling some of the most vexed issues about the nature of art in the twenty-first century.

These, then, are the goals of this paper:

1. Constitutively, to understand *what it is to be digital* or discrete—as opposed, say, to being *continuous*.
2. Pragmatically, to understand *what digital systems are good for*—and what they are not good for; and
3. Consequentially, to understand *the impact of digitality* on our understanding of: creation, ownership, identity, materiality, reproducibility, and the like.

2 Properties

Three properties of digital systems are immediately identifiable: their **perfection**, their **abstractness**, and their **dynamics**. Describing these three does not constitute a theory of digitality. It merely spells out what a theory of digitality must explain.

[1] In 2002, when this paper was first written, Barnes & Noble reported more than 8,000 books with the terms 'digital' or 'digitality' in their title; as of January 2010, the number had increased to 13,574. Even when restricted to its "books" category, Amazon claims almost two million results on the same search, and Google Scholar between four and five million. It is hard to know what those results mean—if they mean anything at all; but the even the smaller numbers indicate a term that seems not to be losing its popular cachet.

2a Perfection

Digital systems are, in a remarkable sense, perfect. When encoded digitally, a system can be *flawlessly copied*, without error, an infinite number of times. No loss, no corruption, no friction, no accumulating impact of dirt or rust. "Perfect sound forever," said Sony, in when it introduced the compact disc. Even if we know better, now, there was something right in their proclamation. No scratches, no noise, no irritating static. And no decay. In Bangladesh, religious manuscripts perpetually disintegrate, inexorable victims of insects and humidity. If only we could record them in digital form, we are told, they would be immune to rot. *Scanning as transubstantiation!* This is truly heaven on earth: the abstract purity of Plato's realm rendered incarnate, in an endless string of 0s and 1s.

Something special is required, for this digital perfection to be achieved. There must be a determinate set of judgments, or properties, or types, in terms of which the system can be completely characterized. Constitutively, that is, in order to be digital, a phenomenon must succumb to a finite series of informationally-complete black-and-white judgment calls. This holds of such ordinary "digital" notions as: scoring a basket in basketball; moving the pawn to K4; writing down the letter 'A'; making a copy of a text, a poem, or a musical score; cutting a board between six feet and six feet one inch long. All these things can be determinately accomplished—without error, ambiguity, or matter of degree.

This is where "information" comes in—a digital system is a system about which *complete information can be given in such (finite, black-and-white) terms*. If we know the answers to all those "yes/no" questions, we have "captured" all that matters about the system. Thus a chess game can be restarted, even if the board is dropped, if we know exactly which pieces were on what squares—a finite, compact list. By contrast, such systems as the state of a billiards table, a haunting smile, a painting or a musical *performance*, or cutting a board *exactly six feet long*, are non-digital because there is no finite, absolute, discrete set of facts of the matter that fully "capture" what is going on.

Sure enough, we can approximate the state or character of a non-digital system (a billiards game, a painting, an image), by using ever finer samples, to any degree that we choose; more on that in a moment. But the phenomenon itself, *at the level at which it is the phenomenon that it is,* is not discretely constituted. Unlike chess, that is, such non-digital phenomena as billiards and paintings are not defined—cannot be wholly and completely accounted for—in terms of a finite set of well-defined "yes/no" questions.

Music is an interesting case, in this regard—because of the difference between a score and a performance. Musical *scores*—at least traditional scores, composed of the familiar suite of notes, staves, markings, etc.—*are* digital, plus or minus a bit. They can be perfectly copied. If one score gets wet, or starts to decay, or has coffee spilled on it, a new one can be made without (so we have chosen to ascribe value) desecrating Beethoven's composition. But the performance itself—which, in virtue of its concreteness, invariably adds an unutterable wealth of detail to the sparse information of the score—is not discrete in the same way. In fact this is one way to understand what performers do: they fill in the infinitely rich detail between the skeletal sparseness of a digital score, and the ultimate thickness of a concrete, continuous (i.e., non-digital), musical utterance.

Issues of ownership and value in music can be made intelligible in terms of this divide. When we credit a (classical) composer with being the "creator" of a work, we do not view the work in question as a concrete in-the-world musical utterance, but rather as a work *under description*—a work "abstracted" according to the conventions that dictate what aspects of a work are captured in a traditionally-notated (digital) score. This "restriction of credit" to an abstracted version makes room for subsequent performances to be viewed as loci of genuine artistry, creativity, originality, etc., in their own right. When the Guarneri Quartet performs late Beethoven chamber music, Beethoven is given credit for the "digital" content of the score; the players, for the non-digital aspects of the ensuing performance. Similarly, one way to understand Factum Arte's project[2] is as giving us a "score" of the tombs in the Valley of the Kings, to be "interpreted," in the future, in different performances, by different Egyptologists and archeologists.

[2]«Explain Factum Arte»

2b Abstraction

A second manifest property of digital systems is that they are at least apparently *abstract*. Programs, bit maps, digital data—none of these things weigh a certain amount, or have energy or momentum, nor can they be eaten by moths, or otherwise decay. Digital data often *relates* to physical stuff, in the way that an arrangement relates to what is arranged, or a configuration relates to what is configured. But qua arrangement or configuration, digital entities are more like things which are truly abstract, such as numbers and pure ideas, than they are like concrete paintings or hand-hewn log cabins.

One way you can tell when something is (at least relatively) abstract is when it can be realized in a wide variety of materials. Famously, chess games do not have to be played with wooden or ivory pieces; salt shakers would do, or people, or a spate of suitably hovering helicopters. Similarly, letters can be formed of ink, or pencil, or jet contrails, or by arrangements of sports teams band-members during intermission. Similarly, it is because they are "abstract" that chess, unlike billiards or fencing, can be played by mail—or over the internet. Digital systems, to use a technical term of art, are *medium-independent*, in a way that non-digital systems are not.

Issues of medium-independence, it should be noted, lie at the heart of raging debates about the possibility of Artificial Intelligence. Are your thoughts digital, like moves in a chess game—implying that your identity could be uploaded onto a digital computer? Or is your mind more like billiards: inexorably tied, at the level at which you are you, to specific irreproducible facts about your material embodiment? The fate of our children depends on the answer.

2c Dynamics

The third obvious property of digital systems, along with their perfection and abstraction, has to do with their *dynamics*. On the face of it, the most obvious dynamic property of digital systems is their *stability*. "Perfect sound *forever*," was Sony's claim. And think of those manuscripts in Bangladesh; what digitality promised was protection against the ravages of time. Indeed, purity, perfection, and *stability*—a kind of eerie immutability or invulnerability to the ferment of life, to the eruptive activity of concrete existence, seems almost defining of the digital realm.

Yet if perfection and epochal stability are the marks of the digital, so too, curiously, is change. On the web, you can create, duplicate, modify with unprecedented abandon. Switching between a zero and a one takes so little energy it is essentially free. Want to adjust that memo you posted last night? No problem! Click, click, click; just one more email. And it's not just we people who change things; convergent networks and routers are in the business of moving things around, not keeping them fixed. Computing itself, in fact, once one thinks about it, is the epitome of change. "Mathematics *plus time*," it has been called. Whitehead redux: it is the *processing* of symbols, not the symbols themselves, that ultimately matters. In fact the symbols and media are increasingly dynamic: streaming video, QuickTime movies, virtual reality enactments, all pouring by at megabytes per millisecond. And what is true of the technology is equally true on the human side: eruptive start-ups, multi-mega-mergers, dot-com demise—a dizzying pace of change. For a revolution based on stability, the digital world sure moves fast.

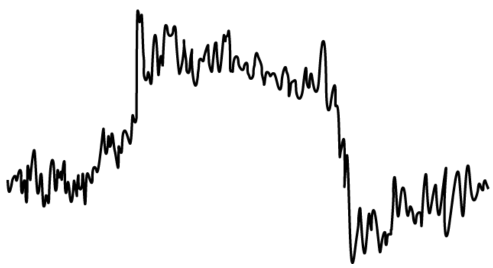


Figure 1 — Pulse in an electronic circuit

Fixity *and* fluidity, in other words—digital dynamics crucially involves both. And both in ideal form. If you want stability, it will stay. If you prefer change, it will change—in exactly the ways you specify. *Perfect* dynamics—that is what powers the digital miracle.

3 Physical realisation

Of these three properties—perfection, abstraction, and dynamics—it is the perfection that is ultimately the most important.

It is odd, moreover—absolutely astounding, in fact—that such perfection is pragmatically achievable, in this our messy world. It is astounding because (this is our first crucial insight) nothing, in the end, is *really* digital. Attach an oscilloscope to a digital circuit (figure 1), and all you see are splattered variations of bewildering complexity. Seriously: how long would it take a Martian to figure out that these intricate whiplashes of electronic alternation are, in fact, *digital*: "naught but 0s and 1s"? The discovery would merit a Nobel prize. Why? Because electrical signals, all those signals running around inside your personal computer, are not, *in fact*, digital. Rather: the parcels and patches of concrete reality that we call "digital," like all patches of concrete reality, are really continuous.[3] And not just continuous, but, like everything that exists, perfused with an unutterable richness and texture and complexity of fine-structure that stupefyingly defies finite description. What makes them digital—or rather, more accurately, what *allows us to call them digital*—is that they are continuous patches that we can treat *as if* they were digital, without getting into trouble.

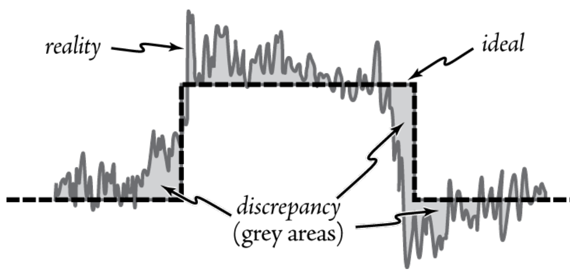


Figure 2 — Discrepancy from Digitality

Or so the story goes.

3a Discrepancy

Pure digitality is a myth—an abstraction in terms of which, with Orwellian abandon, we (re)interpret reality. As indicated by the dashed line in figure 2, the austere digital ideal is never achieved. Rather, reality differs from the ideal by an unavoidable *discrepancy* (indicated in grey). Sometimes, as we will see—far more often than people realize—the discrepancy cannot avoid making an impact. To illustrate, I will presently argue that it is conceptually impossible for two pressings of the "same" audio CD to sound identical. But it is not the discrepancy that is mysterious. What is magic is that sometimes—in so-called "digital circuits"—*the discrepancy doesn't seem to matter*.

We are all familiar with computers that have gigabytes of

[3]Ignoring quantum mechanics—which does not bear on current computers.

memory, run at billions of instructions per second, and are linked to an untold myriad of other computers on the network. If every one of these systems—every bit, step, move, fragment—is "fallen," failing to meet the inaccessible standard of digital perfection, why does the discrepancy not bring the whole thing down? Or to put it positively: how do we build perfection, on top of such inexorably messy foundations? How on earth does the digital idealisation work so well—how can it work at all—if reality is so unerringly defiled? And work well it does; those gigahertz processors and terabytes of memory really do achieve their digital goals.

Figure 3—Demise of a CD

The answer, or anyway the beginning of an answer, has to do with *containment*. Rather than eliminate discrepancy (a hopeless task), digital circuits *control* it. The genius of digital engineering involves figuring out ways to ensure that the *discrepancy does not propagate*. Whenever a signal gets (dare we say it?) *noisy*, we reshape it, clean it up, put it back on the strait and narrow—with Stalinesque efficiency. Memory on your laptop computer is "refreshed" 50 or 60 times a second, in order to stay stable. If it were not, then, like those Bengali manuscripts, it too would rot away, decay, collapse in frangible chaos. It takes work (and battery power) to prop up a digital myth—even to maintain the digital illusion of doing nothing at all.

What is stunning—and after thirty-five years in the field I am still amazed—is that we have figured out how to build devices to maintain the illusion—for a while. In the end, they, too, will fail (figure 3). Not even digitality can forever escape damnation by those deuced moths and rust. But this side of heaven, digitality comes as close to perfection as we can get. And we can get stunningly close—as close, in fact, as we want. Just tell the engineers what error rates you can accept: one in 10^7? one in 10^{20}? one in 10^{25}? Whatever you want; no problem.

3b Discrepancy and noise

What about noise? Is discrepancy noise? Often—but not always. A small company in California listens to random cell phone calls, throwing away the signal. It turns out, for every handheld unit, that its discrepancy serves as a kind of analog "signature." Given inevitable contingencies of manufacture and materials, each device is slightly different. Those differences are reflected in the particular shape of the error or discrepancy signal. (Tolstoy should be happy: we diverge in our own peculiar ways.) The company's job is to monitor the character of the discrepancy, and sound an alarm when it inexplicably changes. That is how phone companies detect when a phone number has been stolen and implanted on a different unit. There is money in being a discrepancy sleuth.

Admittedly, discrepancy is sometimes painful. It degrades the music, distracts the image, crashes the machine. But it can be valuable. Cell-phone discrepancy is useful, because it *correlates*—with particular, concrete handsets. The same is true of pirated software: discrepancies on CD-ROMs contain tell-tale traces of illegal duplicating factories.

Does that mean noise is uncorrelation? No, that can't be right, either. Sometimes lacking correlates is a priceless advantage: a childhood hideout, the Kohinoor Diamond, that night in Kathmandu.

3c Abstraction

These remarks about discrepancy and error, singularity and correlation, tie directly into what I said earlier about the distinction between digital phenomena, such as chess and musical scores, and non-digital phenomena, such as billiards and paintings and musical performances.

The point is simple. *Digitality is not a property of entities per se.* Nothing either is, or is not, digital, intrinsically. Rather, whether something is digital or non-digital is *relative to a level of abstraction*—relative to a level of description at which it is characterized.

It follows that the perfection of digital systems, though *real*, is not *absolute*. Rather, *the perfection, too, is level-specific*—relative to a level of description or abstraction. The digital miracle, therefore, is not that perfection is achievable *at the physical level of abstraction at which the world is, in fact, messy* (that genuinely *would* be impossible). Rather, the miracle is that messy physical stuff can be arranged so that, while staying physically messy, it can nevertheless *implement perfection at a higher, digital level.* (In terms of physics, this involves non-linear phenomena, attractors, and a host of other technical notions.)

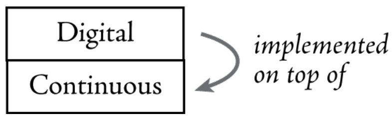


Figure 4—Implementing Digitality

Perfection, we might say, cannot be achieved *in* the physical realm; that would contravene friction, thermodynamics, and those moths and inexorable rust. Rather, as indicated in figure 4, digitality can be achieved *on top of* the physical world—by building it up, at a higher level of abstraction, on top of the underlying messiness.

This "level-specificity" of digitality's perfection is going to matter a very great deal.

3d Digital implementation

We finally have enough equipment to understand compact discs—to say nothing of Adobe Photoshop, digital cameras, and scanned paintings. (Note: I will mostly talk here about music and CDs, because their traditional medium—sound—is a single-dimensional variable progressing through time, which makes for easier pictures. But the points I will make apply equally to two-dimensional static phenomena, such as pictures and paintings as traditionally conceived, as well as to two and three dimensional dynamic media and representations, such as video and virtual reality.)

The picture we have reached is one of a messy, continuous, underlying physical substrate, on top of which we can implement digital perfection. So far so good. If the phenomenon we are interested in—chess, say, or written (at least printed) language, or musical scores—is *itself* digital (i.e., constituted in terms of a finite set of black-and-white, "yes/no" distinctions), then we are essentially done. But what about phenomena that are *not* intrinsically discrete—such as musical performances, or paintings? How can we achieve perfection in their case?

We cannot. That is an intrinsic truth. But we can *fake* it.

A moment ago, I talked about implementing digital perfection on top of a messy, continuous physical substrate. For music, paintings, and other continuous phenomena, we can pull the inverse trick: recursively implement messy physical performances on top of a lower, *digital* level. That is, we can construct the three-level structure shown in figure 5.

What makes this all work—or at least what makes it work as well as it does or can work, which is something I am going to want to pursue in a moment—is our ability to make the implementing digital distinctions be *sufficiently fine-grained*, with respect to the upper-level continuous phenomena, that they are not noticeable, or anyway not unduly noticeable. This is called *sampling*, and is the fundamental strategy behind CDs, digital cameras, bitmapped images, and the like. The details are familiar: you make a digital approximation to a continuous signal so that the *discrepancy*, as defined above, stays small. You can make it as small as you like, by using higher-and-higher sampling rates, more and more megabytes of storage.

Digitality is *continuously implemented*, in other words (i.e., is implemented on top of a continuous substrate), as we saw before; that is the relation between the lower two levels of the figure. Continuous artwork, in turn, is *digitally implemented*, in the sense of being implemented, in turn, on top of a digital substrate; that is the relation between the upper-two levels. When properly executed, the digitality of the middle level will be largely, or at least relatively, unnoticeable, at the top level. This is how the top-level phenomenon (the music, painting, image, whatever), unlike the chess position or poem, in spite of being implemented *on top* of digitality, need not *itself* be understood *as* digital. Imagine a Thelonius Monk CD: the growl, the bending of a note, a sigh. These phenomena are not *themselves* discrete. Their continuity is preserved, more or less, in spite of the digital *implementation*.

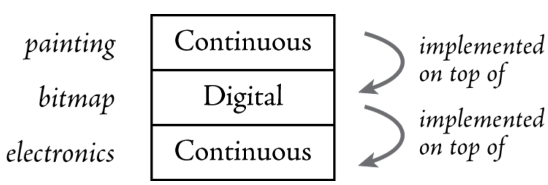


Figure 5—Interposing Digitality

Why go to all this work? Why implement continuity on top of digitality, and then implement the digitality on top of more continuity? Because, by interposing perfection, between the bottom-level messy *realisation* and the top-level messy *phenomenon*, you can largely *insulate* the continuous richness of the upper level *phenomenon* from the continuous richness—which is to say, the moth and rust and disintegration—of the bottom level. The result is that the particularity of the "performance," as it were, is insulated from the particularities of the recording—which in turn gives you extraordinary portability, stability, and immunity to decay.

Or anyway that is the theory. As we will see in a moment, reality is more complex. Still, this analysis answers our second question, by showing us the "why" of digitality: it is an engineering strategy, pure and simple, for insulating the continuity of one phenomenon from the continuity of another, in terms of which the former is carried. As John Haugeland has put it:

> *"Digital, like accurate, economical, or heavy-duty, is a mundane engineering notion, root and branch. … It only makes sense as a practical means to cope with the vagaries and vicissitudes, the noise and drift, of earthly existence."*[4]

In an ultimate sense, as I have already suggested, it is the lower half of figure 5—the implementation of digital (i.e., perfect) systems on continuous substrates—that is the miracle. Why it is that the world is such that, as far as we can tell, digitality is the only way to achieve perfection, and why it is, correspondingly, that the world, at the messy physical level, is such that digital perfection can be achieved on top of it—these are the sorts of metaphysical question that keep me awake at night. They are questions that no one, I believe, has yet satisfactorily answered. But in terms of concrete, pragmatic impact on our lives—our third opening question—what matters is not so much digitality per se, and its possibility, as what we have revealed here: the consequences of *digital implementation*. It is the full three-layer structure of figure 5, not just the two-layer structure of figure 4, that is transforming the world of art.

[4]Haugeland, John, "Analog and Analog," *Philosophical Topics* (Spring 1981); reprinted in J. I. Biro & Robert W. Shahan, (eds.), *Mind, Brain, and Function: Essays in the Philosophy of Mind,* Norman, Oklahoma: University of Oklahoma Press (1982), pp. 213–25; quote is from p. 217.

Negroponte's book was mistitled. It should not have been called *Being Digital.* It should have been called *Being Digitally Implemented.*

4 Mediation

Conceptually, most of the official story is in place. But it is instructive to pursue an example, to understand its profound limitations. Because the theory, as I have presented it so far, is not quite right. It is not *bad*; no one would turn down owning digitality's patent. But as usual, the devil is in the details.

To see what's wrong, I want to show how this analysis, if followed out with relentless logic, contravenes what is almost universally assumed: that all instances of the same digital "signal" are *absolutely*, not just *relatively* identical. The example is taken from on-going debate in high-end audio circles about whether it is possible for two different pressings of the "same CD"—i.e., two different token polycarbide discs, each of which contains "exactly the same sequence of 0s and 1s," to sound different.

According to the official story, they must sound the same. According to me—according to reality, that is, I will claim—they must not. And as usual, there is nothing special or peculiar about this result. The conclusion will hold of any digital implementation whatsoever.

4a Different bit streams

As I've said, the debate concerns pressings of "exactly the same" CD, where the two tokens have "identical" bit streams. It should be noted, however, that this is a difficult case; there are huge issues, in the art world, about much easier cases, when the CDs—or digital representations in general—encode *different* bit streams. Here, what is important to realize is that *identity*, like digitality (and like just about everything else we are talking about), is level-specific. In cases of music and art, unlike texts and scores, where the "original" is continuous, identity at the top level need not correspond to identity at the middle level. Or so it is argued. And so people interpose digital "watermarks," or lossy compression (e.g., to MP3), or digital stamping, and so on—claiming that they can do this without altering the upper level. I don't want to consider these examples, here, except to say that such changes *have* to make *some* difference. If the upper level is *continuously identified*, at the top level (as in a performance, or painting), then any difference at the middle level is a real difference. What's at stake is not whether there will be an upper-level difference (there will be), but whether that upper-level difference matters.

4b The debate

But turn to the example at hand: of whether there are or even can be aural differences between pressings of the *same* CD—i.e., between two pressings that encode the very same, identical bit stream.

The debate takes predictable form. So-called "golden-eared audiophiles" claim to hear differences between and among such different pressings. With great vehemence, self-styled "rationalists" deny the objectivity and validity of these golden-ears' subjectivist claims. "It is *impossible* for the two pressing to sound different," they cry. "They cannot sound different, because they are digital, and, as digital recordings, they are *identical*."

The rationalists are wrong. To think that two pressings of the same identical bits *must* sound identical is simply a conceptual—perhaps ideological—mistake.[5]

To see why, we need to understand the impact, on the continuous (audio) signal at the top level, of what I called the "discrepancy," at the middle level of our three-level diagram (figure 5), between concrete reality and digital abstraction. In particular, consider again figure 2 (page ■■), showing the inevitable discrepancy between the real in-the-world voltage and the digital abstraction superimposed upon it. This image depicts a "digital" *signal*, such as a track on a CD, a bitmap downloaded from a digital camera, the output of a Photoshop session tweaking a digital

[5]Similar disputes, I might note, arise throughout the community: about the rationality of colouring the edge of your CDs with green felt-tipped markers, of using isolation transformers on the CD player's power cord, etc. My general reaction, in such debates, is not only to feel that the golden-eared audiophiles are phenomenologically correct, but also that their conclusions, far from involving anything mystical, are straightforwardly scientifically explicable, if only one's conceptual analysis is sufficiently powerful. Coloring the edge of your CDs is perfectly rational, it turns out, if one is scientifically awake.

scan. That is: it is a picture of (a piece of) reality *within the digital realm*. But the image is reminiscent of something more familiar: pictures, such as the one in figure 6, of how we *sample* or *scan* or *convert* continuous phenomena (paintings, music, 3D-scenes etc.) into digital form. In these input cases, too, as always, the digital abstraction (the "information that can be encoded") diverges from the infinitely-rich concrete reality.

Every digital craftsperson is taught this: that digital encoding—the so-called "analog to digital" (A-D) conversion process that takes place in every scanner, digital camera, DAT recorder, etc.—will invariably miss all sorts of small or miniscule variations, subtleties, and nuances in the original or "source" phenomenon. This is the realm of bit-depths, sampling rates, compression, etc. Suppose figure 6 depicts the (continuous) acoustic intensity of a live music performance. The areas marked with in gray are those aspects of the original that the digitisation process will fail to capture. By the same token, digital cameras and scanners analogously "abstract away" from: (i) any fine-grained structure of the original image or scene that is too small to be "caught" within the temporal or spatial sampling rate; and (ii) all variations in intensity that are less than one bit's worth of gradation in the system's dynamic range.

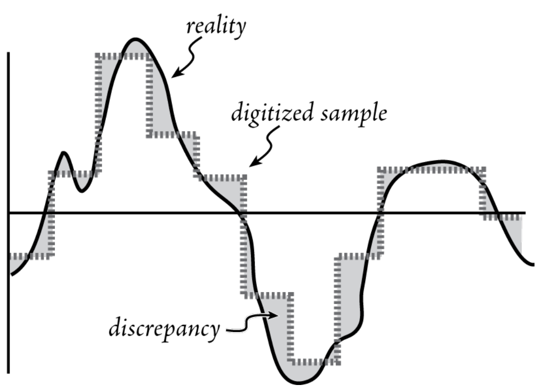


Figure 6—Sampling of a continuous original (i.e., top-to-middle-layer discrepancy)

And what is true of input is equally true of output—though the point is less familiar. Suppose we print or render a bit map or other digital encoding: on a television screen or monitor, cheap ink-jet printer, or expensive imagesetter. Printers, monitors, etc., produce *real images*: concrete, continuous, full-blooded denizens of the world. These output images, too, being actual, will, like everything else, have an infinitely rich and detailed fine-structure. They, too, will look like figure 6. Or consider listening to a CD. Once again, the digital-to-analog (D-A) converter will take as input a digital signal, and produce as output something that is continuous, analog, and (as usual) infinitely-detailed. Being actual, these outputs, like everything else in the universe, will in fact have an infinitely rich fine-structure.

It follows, from all this, that if two images (sounds, whatever) are produced from one digital source, that they can potentially differ in some or even all of their fine-structure, in all their discrepancy. In fact they can—and will—differ in an infinity of ways, in spite of having been produced from the same bit stream, because the "digital" bit-stream doesn't determine that in one sense superfluous but in another sense absolutely necessary fine-structure. This is the point that underwrites Lowe's work on *Digital Prints*.[6] Starting with multiple copies of an "identical" bitmap, he printed eighteen high-quality prints, using eighteen different printing/rendering processes. They look different—radically different, even, when examined closely. And from what we have said we can easily see why. They look different because they differ *in their fine-structure*—in their *discrepancy from the* (common) *digital abstraction*.

4c Fine-scale interactions

So prints, outputs, sounds, all differ. What does this have to do with different pressings? Because of this punch line:

> *The discrepancy intrinsic to the (continuous) physical realisation of a digital signal (i.e., the discrepancy endemic to the relation between the middle and bottom layers of figure* 5*) invariably influences the variation at the top layer of the resulting performance (i.e., the discrepancy endemic to the relation between the top and middle layers of the figure).*

Not only *can* it have an influence; it *must* have an influence. It is a theorem of physics.

Why? Because, as we said at the outset, the "digital" signal is *not really real*. All that "really" exists is the underlying, physically messy carrier. The so-called "digital signal" is only an idealizing abstraction.

The point is that D-A converters, the devices that *produce* a per-

[6]«Reference—and explain.»

formance, given a digital signal, do not—and can not—work as indicated in figure 7. This figure illustrates how people *think* things go—but it is a fantasy, based on the idea that the digital abstraction is real. Rather, the way they *really* work is indicated in figure 8. They work this way because this is all that really exists.

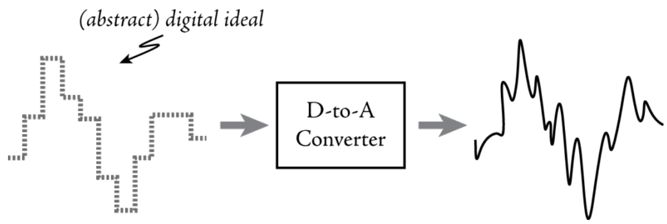


Figure 7—How digital-to-analog converters do **not** work

It is perfectly obvious, in fact, that the fantasy could not be real. Just think of what it would require! It would mean that an engineer would have to build a concrete, physical device that (i) responded to *the non-existent digital signal that the actual analog signal was ideally meant to encode*, but (ii) that *ignored the actual variation or "discrepancy" in the actual, real physical signal*, which is the underlying physical realisation of that digital abstraction.

And that, needless to say, cannot be done.

I have talked of three kinds of fine-structure, each more detailed than is captured in any governing digital abstraction: (i) fine-structure in the original input, if there is one, before it is entered (converted, scanned, etc.) into a digital realm; (ii) discrepancies within the digital realm, in the fine structure of the signals that "carry" or "encode" the digital abstraction; and (iii) fine-structure in the output (prints, sounds, images) produced from those digital encodings. The striking fact is that it is a fundamental theorem of physics that these fine-structures not only will, but must, influence each other.

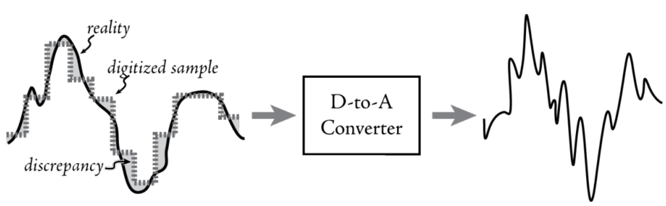


Figure 8—How digital-to-analog converters **do** work

To see why, think about the encoding process. Suppose we start with a routine continuous signal—an acoustic wave, painting or image, or 3D-scene. We've all been told, thousands of times, that analog-to-digital encoding processes take continuous signals as input, and produce digital signals as output. But do they *really* produce digital outputs? *No—of course not!* Digitality, as we've said, is an *abstraction*. Analog-to-digital converters, in contrast, are concrete: made out of physical stuff—the same stuff that we are made of, the same *continuous* stuff of which field-theoretic physics holds true. As a result, it would be contrary to the laws of physics for them to produce something abstract. Rather, like all physical processes, take in, *and produce*, concrete, continuous, signals or waves. More specifically, what analog-to-digital converter really does is: (i) take as input a continuous, concrete, real-world signal, and (ii) produce as output *another* continuous, concrete, real-world signal, where (iii) the output signal, *if interpreted under a digital abstraction*, can be seen to "encode" the digitised version of the input.

Analog-to-digital conversion processes, in other words, don't mediate between what is concrete and what is abstract. No real-world process could do that; it would be magic. Rather, A-D conversion is a *concrete-to-concrete transformation*, both ends of which are as a result genuinely continuous.

Exactly the same moral holds true, of course, at the output end: in processes of digital-to-analog conversion. Just as with A-Ds, D-A converters don't *really* take a digital signal as input, and produce a continuous one as output. Rather, they (i) take as input a continuous signal that supports a particular digital abstraction (i.e., lies within its acceptable constraints), but that, like all "digital" signals, is complete with discrepancies and fine-structure; and (ii) produce as output another continuous signal, the continuous signal which the digital abstraction of the input encodes. Just as in the input case, that is, output D-A conversion is a process of *concrete-to-concrete* mediation.

Once we have recognised the inalienable concreteness of the signals at both ends of A-Ds and D-As, it immediately becomes clear that it is an absolute necessity—a veritable theorem of physics—for the fine-structure (or discrepancy) in the inputs of A-Ds and D-As to have an affect on the fine-grained structure of the outputs. In particular: the fine-structure of the "digital" input to a D-A will effect on the fine-structure of the continuous output. It *must* have such an impact, because (as usual) physical devices are continuous. Sure enough, engineers can strive mightily to minimize the effect. But there is no way, in this world we inhabit, for an engineer to build a concrete physical device that (i) responds *to the (non-existent) perfect digital idealisation* that a signal is "intended" to implement, but nevertheless (ii) to ignore the fine-structure of the incoming signal *as it actually is*.

This is why, ultimately, no two pressings of the same CD will (or even could) sound exactly the same. Or rather, to put the point more exactly: this is why no two pressings of *the very same (digital) bit stream* will ever lead to *exactly the same (continuous) acoustic wave*. They will sound different even if we assume, for simplicity, that they are played on the same stereo system, in identical states. They will sound different because, although each CD will carry the same digital idealization, each will do so complete with its own unique fine-structure—i.e., with its own distinctive way of diverging from the putative digital ideal. After all, the fact that they are the "same" CD means no more than this: that if we were to abstract away from the two infinitely-rich continuous patterns, in the way mandated by the digital idealisation, the two pressings would be discovered to "carry" the same stream of digital bits. Any aspect or fine-grainedness of structure that is not relevant to this digital abstraction is free to differ between the two CDs. And as we have already seen, the D-A converter is mandated by the laws of physics to respond differentially, in the two cases, *to those different fine-structures*. Perhaps not *very* differently; but nevertheless *some* differently. So when the continuous signal is extraced from the D-A, sent to the power amplifiers, and propagated to the speakers, it will carry its own distinctive characteristic signature. There is no way in which it could be any other way.

And finally, to bring this back to images, the same holds true of printing. It is not just that two printings of the same bit-stream (even: of the very same CD encoding that bit stream) can produce different concrete images, when printed on different printing devices, as Lowe showed so compellingly. It is also that two different CDs of that "same" bit stream, when printed on the *same* printer, will *also* produce different prints. This will be true independent of how the digital bit stream was produced: entirely within the digital realm (Photoshop or painting programs), or scanned or sampled from a continuous original.

5 Conclusion

What have we learned?

Six things, already. And a seventh lies just below the surface, with which I will conclude.

1. Digitality is not an intrinsic property of anything. Whether or not something is digital is a higher-order characterization of it: a characterization of a characterization. It is *characterizations* of objects, "takes" or cuts on objects, that are, or are not, digital—not objects per se. As a poem or score, a text may be digital, even if as an arrangement of ink, it is not. As a CD, a recording may be digital, even if, as a reflector of laser light, it is not.
2. Some objects—such as musical scores—are digital at the level at which we identify them as the sorts of thing that they are. That is why we say that a score can be "perfectly" copied, or think that we know *exactly* what sonnet Shakespeare wrote. It is not that we (or anyway most of us) *have* the fully-concrete sonnet that issued from his pen. Rather, what society or culture or history has settled on, about sonnets and scores, is that what constitutes their identity, as the kind of object that they are, is their characterisation under a given set of descriptors or types, which can be exhaustively specified in terms of a finite set of "yes/no" decisions.
3. In spite of the identification of some classes of thing (such as scores) as constitutively digital, nothing *actually* is—or anyway, nothing concrete, nothing actual. The physical world is messy, and so any material thing, as a material thing, is, far from being perfect, in fact a messy, decaying, piece of stuff.
4. Phenomena that are not only (of course) not digital per se (nothing is), and that are also not digital as physical entities (as we have just seen that nothing is that, either), and that are not digital at the level at which we take them to be constituted—such as paintings and musical performances—can be *digitally implemented*, at some loss, but with

the benefit that one thereby largely insulates their high-level continuity from the low-level continuity of the substrate—paving the way for extraordinary longevity, transportability, reconfiguration, modification, etc. This is the realm of the digital CD and the digital image; a digital implementation of a continuously-constituted phenomenon.

5. In spite of the undeniable success of this three-level strategy (of digitally implementing continuous phenomena), the underlying discrepancy is never avoided entirely. Because, as we have seen, physics is continuous, the discrepancies from the ideal in the lower level of implementation (digital on top of a messy substrate) can never be wholly isolated from the discrepancies at the upper level (the loss or violence to the continuous upper-level phenomenon that comes from digitally sampling or representing or encoding it).

   Furthermore, it is a theorem of physics that this unavoidable underlying discrepancy will always influence the output.

6. Because nothing physical is in fact digital, and because, as we have seen, the discrepancies can never be entirely removed, it follows that *digitality itself is an abstraction*. This is the reason why, even though we *say* that a digital implementation insulates the continuity of the constituted phenomenon from the messy continuity of the implementing substrate—to say nothing of the moth and rust—it is never actually so. Sure enough, as I have just said, we can go to a lot of work to minimize the impact of the inexorable discrepancy (different pressings of the same CD can be arranged to sound *pretty much* alike). But the metaphysical truth remains: digitality is not only a property of abstraction; it itself is an abstraction. When we say, of an abstraction, that it is an abstraction—for example, when we say of it that it is a digital abstraction, as for example in the case of musical scores—we are (recursively) engaging in a higher-level abstraction of our own.

7. Finally, what goes around comes around—one more final time. Even the idea that we are *abstracting* is an abstrac-

> tion. The whole edifice of "levels of description" is a *way of describing* what we do. It is a cut, a take, on our epistemic practices.

Nothing that is actual, actually abstracts. Rather, for us to say that something abstracts—a recorder or scanner that performs an analog to digital abstraction, say, or a printer or amplifier that performs a digital to analog abstraction—is an abstraction of ours, which, *as* an abstraction, like all abstractions, underdescribes what is going on. To say that something abstracts is to do an injustice to it. To make a claim about an *actual* process is to commit oneself to an abstract characterisation of a (concrete) process—a process that, like all physical processes, mediates between *one thing that is concrete and something else that is also concrete* (or perhaps we should say, more carefully: between one thing in its full concreteness and something else in *its* full concreteness).

Put it this way: it is not just that *digitality* is an abstraction. Nor is it even, though this is also true, that the *perfection* of digitality is also an abstraction. The bottom line is stronger still: *abstraction* is an abstraction, of which digitality is an instance. As I have said, nothing that is actual, actually abstracts. We might as well get used to it.

The world is utterly and inexorably concrete.

9 — Deconstructing Digitality

— *Had this page been blank, that would have been unintentional* —

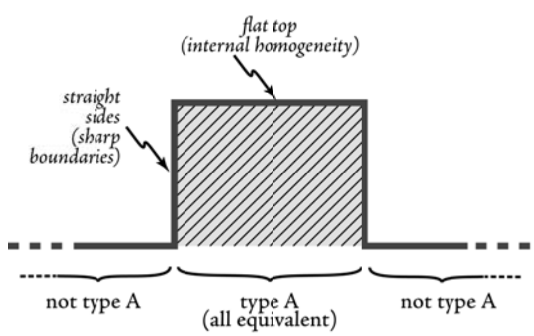
Deconstructing Digitality

Dismantling some of our preconceptions about the foundations of the Information Age†

Brian Cantwell Smith
August 16, 2007

Of all the terms associated with the computer revolution, none are more celebrated than "information" and "digital." Both have been vaulted to prominence as emblematic of our age. A search for "information" on Amazon.com returns more than half a million books; for "digital," the number is more than twice as high—close to 1.3 million.

The notion of information has received critical theoretical analysis in multiple disciplines—from biology to engineering to philosophy to sociology. Digitality, on the other hand, remains remarkably unreconstructed. Perhaps digitality is taken to be simple, or computers' digitality to be obvious. Whatever the reason, questions about digitality are rarely asked. Not that digitality is unimportant. Arguably, the invention of the digital computer was the major development in the history of computing. Sure enough, there are analog computers, too: old ones, of resistors and capacitors; and new ones, such as artificial retinas and cochleae. But think of what digitality unleashed: universal machines, programming languages, implementation and data structures—to say nothing of e-mail, the Internet, compact discs (CDs) and virtual reality. Somehow or other, digitality—or "discreteness," to use an equivalent term—lies at the core of the computer revolution.



More abstractly, computers' presumed discreteness, or "absoluteness," plays a major role in our computational *Zeitgeist*. That computer science is a "formal" discipline, that computing is amenable to mathematical analysis, that computer science is a science—all these classifications rest on the premise that the appropriate theoretical concepts for studying computing have a formal, or discrete, character. Similar assumptions underlie the widespread view that computers are nothing more than dry and desiccated machines. Indeed, it is exactly the alleged contrast between the cut-and-dried, neat and sharp categories of the formal computational world, and the messy, contested, inevitably metaphorical and, ultimately, "wet" categories of human life-as-lived that drives the wedge, many people would say, between the monstrously mechanical and the sacredly humane.

But is it true? Are computers, in fact, digital?

And what does "digital" mean, anyway? What would be it for the myth to be true?

1. Abstract Perfection

A first cut at the nature of digitality is best conveyed with a picture.

As suggested in figure 1, two things are required. The first, depicted by the vertical edges, has to do with a digital state's boundaries: they must be absolutely sharp. Whether a system is in a given state—on or off, 0 or 1, yes or no—must be a totally and completely definite question. Either it is, or it is not—with no room for ambiguity or degree. Digitality, thus, manifests what we never find in nature: an absolute, perfect, 90° cliff.

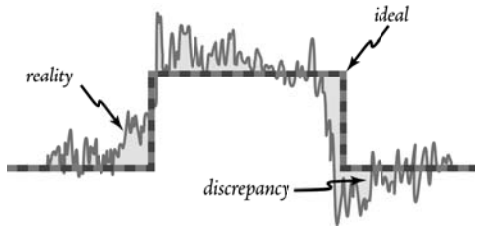
The second aspect, depicted by the flat top, is that digitality requires utter internal homogeneity or uniformity, with no internal variation. All instances of a digital type must be exactly equivalent. One "0" state is as good as another "0" state—completely interchangeable. Again, there are no matters of degree; there is no possibility for the system to be partly 0, or mostly 0, or vaguely 0, or more-or-less 0. Everything is absolute, determinate, and clean.

Needless to say, nothing in the real world is quite so neat. But that is all right. In fact, the construction of digital systems is expressly aimed to accommodate such cases. Departure from the ideal is not so much forbidden (which would be difficult to achieve, let alone sell for cents per megabyte), as almost magically rendered irrelevant. That is, the idea is not that things are discrete in some absolute or ultimate metaphysical sense, but that they are fashioned so as sustain a digital level of description.

Rather than attempting to eliminate variation, engineers build digital systems by arranging things so that the inevitable individual variations do not matter, such as

 Extracted and abbreviated from "Indiscrete Affairs," forthcoming.

voltages wandering up and down around some standard. To whatever extent is necessary, offending properties are cleaned up, boxed in, confined to certain limits, kept from spilling outside a protected region. As a result, errors neither accumulate nor propagate, and results do not get out of hand. The trick is to ensure, with respect to the overall or future state of the system—i.e., with respect to everything that matters about the system at the digital level of abstraction—that all present and future behaviour, such as whether the system will be in state B, depends only whether the system is now in state A1 or A2 or … or Ai, not on the way in which it is in one or another of those states. As long as that condition is met, any potentially distracting variations will be locally contained—washed away, made invisible. As a result, the relation of the system to the (digital) property of being in state B is reduced to a single "bit" of information. Yes or no. On or off. Black or white.



You can see what is going on in figure 2. Taking an electrical pulse as paradigmatic, the green line indicates what the electrical circuit is actually like. The dotted red line indicates the "digital idealization." The yellow region indicates the "discrepancy" or "departure from the ideal"—the difference between idealization and actuality.

The amazing accomplishment, for digital systems, is that they are built to work as if they were red, instead of what they actually are, which is green. In constructing the rest of the system, that is, or in analyzing its behaviour, you can assume that it is red—in spite of the fact that the red line does not exist! This is a more impressive achievement than may be obvious—easily, in my view, worth a passel of Nobel prizes. It is certainly far from obvious that such a construction is possible. If you were to build a building with this kind of error between how it was supposed to be and how it was actually built, it would likely fall over.

Contrary to popular myth, in fact, the lowest levels of computers, far from being adamantine 0s and 1s, are not all that stable. Situations regularly occur where the implementing physical parameters get out of hand, wrecking any simple digital abstraction. Compact disks are a dramatic example, where a fingernail scratch can leave a wake of devastation hundreds of bits wide. Cosmic rays and the conveyor-belt motors at security checkpoints similarly can produce decay, to say nothing of a background slow drift and general disintegration in underlying materials. In a curious sense, in fact, modern digital media are more vulnerable than traditional non-digital ones. As is often pointed out, high-quality paper can last for hundreds or even thousands of years; disk drives are lucky to last 10. Optical media do better, but only somewhat, at best lasting a few decades.

How is the digital abstraction maintained, given these inevitable processes of dissolution? An extraordinarily impressive surrounding structure of routines and mechanisms prop up the digital abstraction. Compact disks employ staggeringly complex error recovery schemes to preserve and even recover the idealized digital "signal" in the face of catastrophic tracks of microscopic destruction. Laptop memory is rewritten every 15 milliseconds, in order that rapidly accumulating "bit-rot" does not take over. Internet packets are checked and resent when they have eroded en route beyond the point of digital recognition. Disk headers are stored redundantly; fragile memories are backed up on disks; mission-critical applications are run in parallel on identical computers, in case one fails. The full gamut of such coding strategies and error recovery schemes is extraordinarily impressive. Certainly the popular idea that a visitor from Mars could examine a single CD and simply "read off" the music is a severe stretch, if not an outright error.

What is digitality for? Why all the fuss? Why construct a system that—at least at this abstract level—is so pure, so crystalline, so fixed? John Haugeland gives an apt answer. Digitality, he writes, is "a practical means to cope with the vagaries and vicissitudes, the noise and drift, of earthly existence" ("Analog and Analog," Philosophical Topics, Spring 1981). Discreteness, that is, more than anything else, is about protection—protection from the ravages and uncertainty and exigencies of the local surround. Winds might blow; the power supply might suffer a brownout; moth and rust might corrupt; someone at the next table might say something distracting. If you are a digital system, you need not care. Your constitution guarantees that you will not be buffeted unseemly by such local aberrations. In fact, you will not be unseemly at all. In a certain "abstract" sense, digital systems are intrinsically perfect.

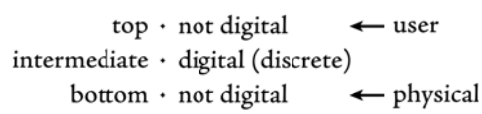
2. The User Experience

How do we experience the digital? At one level, the answer is obvious: we construct programs, automate processes and transformations, store data, send e-mail, interact with other users, manipulate "information." All of these things "exist"—are coherent and intelligible—at the digital level of abstraction. But that is not all. Something else we do, as quickly as we have achieved the digital level, is do our best to hide it.

Think again about CDs—but this time, about the music. For example, think of a recording of Charlie Parker.

Or a scanned original of a hand-written Walt Whitman poem. Or a late-night phone conversation with a lover. In each case, the medium or substrate will be digital in several respects: frequency, volume, hue. Yet, it does not follow that the music itself, or the nuances of the image, or the inflection in the caller's voice, are thereby themselves rendered phenomenologically discrete. Rather, what these examples show is that you can implement or encode or represent something non-digital on a digital substrate, but continue to experience it as continuous.

This fact about the relation among one and the same system at three distinct levels of description, only one of which is digital, is as (if not more) important to the computer revolution than the simple fact that there is one level of abstraction at which most computers can be taken to be digital, even if from a physical perspective they are not. The situation is depicted in figure 3. Even if it has grown familiar to the point of the banal, it is still amazing that we can construct a single system—one and the same "thing," a single patch of metaphysical reality—that can be analyzed, simultaneously and correctly, at three different levels of abstraction: (i) a top level, such as music, poetry, and the like, implemented (encoded, represented, constructed, etc.) on top of (ii) a "digital" level (the non-physical abstraction depicted as a red line in figure 2, which obeys the criteria of perfect discreteness), implemented, in turn, on top of (iii) a bottom physical level, at which it is not discrete.



Three levels of a computational system

Arranging things in this triple-decker fashion simultaneously gives you the best of all possible worlds. It is fortunate that the lowest level, the level of the physical substrate, is not digital, since that means we can actually build things out of circuit components, metal parts, light guides, slightly varying components, and so forth—i.e., stuff made out of the messy, decaying, material clay supplied to us as the basis of all that exists. If we arrange that layer properly, however, mechanically and dynamically, we end up with a device that, at a higher level, supports the digital abstraction, with all of the resulting perfection discussed earlier: freedom from buffeting, protection from the ravages of time, insulation from unwanted or unwarranted influence. The astonishing part is that this protection from the world's dishevelment apparently extends upwards to all levels implemented on top of it. And yet—and this is the crucial part—this immunity of upper levels from buffeting and decay is accomplished without requiring that the higher level phenomena (the music, the meaning, the caller's sotto voce intimations) themselves be rendered experientially digital or discrete. In virtue of being "digitized," that is, the music, meaning and intimacies need in no way be neatened, straightened up, clarified or disambiguated. No boxing on the ears is required in order to force them into the strictures of the discrete.

When we talk about "digitizing" music and art, in other words, strictly speaking we are using shorthand for "digitally encoding." To render the music itself digital would mean taking away from the Bird the ability to transform one melody continuously into another, or to build gradually from a whisper to a growl, or to have every performance of the "same" tune be unique. Fortunately, CDs don't require that.

The simplest way to understand the achievement of the digital age, therefore, is the three-level structure depicted in figure 3. This is what our future rests on: an intermediate level of digitality, sandwiched between a lower, non-digital level of the brutely physical, subject to inexorable material buffeting and decay, and an upper, non-digital level of music, meaning, social praxis. Between the two lies the abstract, but terrifically consequential, intermediate, digital level, which, by virtue of its achievement of almost magical perfection, affords the upper level complete protection from the ravages of the underlying lower-level physics, thereby enabling arbitrary mobility, perfection and replication, without requiring that that upper level itself be digital.

The protection of the digital without the price of the digital—that is what the intermediate level provides to everything above it. Moreover, given that we have the intermediate level of digitality, we can use it to harness the almost arbitrary powers of algorithms, programming, data, and information processing, in order to engender limitless patterns of transformation and interaction, configured so as to instill arbitrary creativity in the uppermost level.

It is a three-level confection of historic power—with society, needless to say, dining out on the results. And remember: the different "levels" are not separate, modular pieces of an integrated whole. They are all the very same system or phenomenon, analyzed at different levels of abstraction.

3. Conceptual Discreteness

From what has been said so far, you might take the conclusion to be this: that (i) while nothing is physically digital, (ii) we can, nevertheless, build physical things to sustain a digital (i.e., "computational") level of abstraction, (iii) on top of which we implement all kinds of non-digital things. Doing so gives these implemented things an unprecedented degree of stability and mobility—even virtual perfection. Society's slogan should be "The Digitally Implemented Age," not "The Digital Age." And that's where things would stop.

It is not a bad, as a first, cut, but even it is wrong. And this time, it is a major falsehood—or perhaps we should say, an expensive falsehood. Getting over it will cost a great deal of the modern intellectual tradition.

The problem is that there is a more abstract form of digitality—what Haugeland calls "higher-order digitality"—that applies, not to the specific waveforms and measurable quantities of a concrete phenomenon, but to the very concepts themselves, in terms of which things are explained. Thus, consider force, mass, velocity, charge—staple concepts in physics. Specific forces and velocities can be as continuous as you please (23.759 kilograms, 0.3335640951981521 x 10-8 seconds, etcetera). However, the concepts in terms of which such things are analyzed are as pure, discrete and distinct as any digital states: nothing is ½ of a force and ½ of a mass, or partway between a momentum and duration. The concepts of physics are like the monoliths at the opening of the movie *2001*: unadulterated and distinct.

In contrast, consider arrogance—and the boundaries between it and pride, egocentrism, self-confidence, braggadocio, and the like. Sharp edges do not apply. Nor is the issue just epistemic, of judging whether someone is one or other; the point is that the concept does not (and could not) not be sufficiently precisely determined for there to be an exact metaphysical answer as to whether someone is arrogant or not. Moreover, the internal structure of arrogance is not uniform, either—implying that the concept is not internally homogeneous. People are more or less arrogant, arrogant in this or that particular way—in ways that make a difference, with respect to their arrogance.

The problem is that actual computer systems deployed in real-world situations betray the fact that a large number of computational categories, in spite of being built on top of our now-familiar abstract form of discreteness, are more like arrogance than they are like mass. Consider four notions fundamental to the analysis of any real-world computer system:

1. *Subject/object*—and allied notions of *representation/represented*, *symbol/referent*, *sign/signified*, and so on;
2. *Form/content—syntax/semantics*
3. *Inside/outside—internal/external*, *intrinsic/extrinsic*
4. *Abstract/concrete*

In each case, concrete, lived experience (rather than theoretical constructs built on assumptions to the contrary) shows that they are far from being neat and clean, "clear and distinct"—i.e., digital—concepts. That is not to say that these (or a host of other such) distinctions are useless, inapplicable or untenable. The point is just that, at best, they demarcate a complex, intermediate region or territory—not a "gradual" or "continuous" or "smooth" compromise, but rather a turbulent locus of ferment and activity, a place where things are stretched and pulled and splintered into a thousand other considerations, considerations that no longer line up and pull in one direction, nor line up and pull in the other, but sunder, cross-fertilize and lead to more distinctions—all the way (as it is said) up to "the edge of chaos."

Ultimately, instead of being discrete, the situation begins to resemble that depicted in figure 4.

And so it goes—to deeper levels and broader scopes. Not only do specifically computational properties fail to be discrete, but the same moral applies to more general distinctions, of which computer systems are sometimes used as models: between nature and society, the sciences and the humanities, subject and object, mind and body. Computers are wonderfully disruptive precisely because, if properly understood, they make a sham of the ultimate sharpness of every one of these classical dualisms. Computers are symbol manipulators par excellence, but does that mean they validate those who claim that language is merely an endless play of signifiers? No, they do not. They spend too much time mucking around in their own (semantic) task domains. In fact, they tell the lie to that postmodern mantra.

Ultimately, in fact, it is wonderful historical irony. Computers are supposedly objective, scientifically "OK"—intellectually respectable, naturalistic, not spooky. It is in virtue of this pedigree that they are *echt* denizens of the modern academy. But this alleged respectability, so innocuously garbed in the idea that computers are "mere machines," may turn out, historically, to reflect no more than sheer prejudice.

Loosed into the wild, computers play the trumpet outside the digital walls of Jericho. The boundaries of conceptual discreteness are tumbling down.

4. Conclusion

Why does it matter whether the digital level of abstraction is "real"? That much of what we call digital is neither physically nor experientially digital, but only digitally implemented? That the concepts and categories of computing are not conceptually discrete?

In part, the answer stems from a point with which we started—that notions from the computer revolution, such as digitality and information, have assumed such importance in our collective imaginary. As said there, many people assume there is a fundamental (discrete!) divide between people and computational "machines"—that the latter, by virtue of a presumptive neatness, formality, and cut-and-dried conceptual structure, have no purchase on the contested and metaphorical "wetness" of human existence.

I would be the last to claim that anything anyone has built so far can manifest care, chuckle ironically or make a surreptitious gesture. But it is not a fact from which I would extract metaphysical comfort. We have a long history, after all, of striving to maintain the human as fundamentally distinct from the other systems with which we share our habitat: the heavens before Galileo, the animals before Darwin. Reaching for non-discreteness as a way to secure us from the encroachment of the Information Age is just as likely, in my view, to be grasping at metaphysical straw.

Any importance (and humility) that we humans are worth must stem from concrete facts about our actual existence, not from any presumptive immunity from being reproduced—or perhaps more elementally, from belonging to the world.

— Were this page blank, that would have been unintentional —

10 — Indiscrete Affairs†

Everyone knows that computers are digital. Or at least that *most* computers are digital. Sure enough, there are exceptions: analogue computers, resembling old telephone exchanges, for solving differential equations; ultra-modern analogue VLSI chips that mimic the human cochlea and retina; continuous Turing machines theorized in mathematical papers in computer science; incipient dreams of organic and quantum computers not based on zeros and ones. Still, the invention of the digital computer is widely taken to have been one of *the* major developments in the history of computing. Think of what came along with it: abstract symbols, universal machines, programming languages, data bases, digital controllers—and the internet. To say nothing of CDs and DVDs, personal computers, e-mail, mobile smartphones, electronic gaming, and virtual reality. Somehow or other, **digitality**—or **discreteness**, to use a term that for present purposes I will take to be equivalent[1]—lies at the core of the computer revolution.

More abstractly, computers' presumed discreteness, or "absoluteness," plays a major role in our computational *Zeitgeist*. That computer science is a "formal" discipline, that computing is amenable to mathematical analysis, that computer science is a science—all these classifications rest on the premise that the appropriate theoretical concepts for studying computing have a formal, or discrete, character. Similar assumptions underlie the widespread view that computers are nothing more than dry and desic-

†An abridged version of this paper, with the title "Deconstructing Digitality," was published in *Idea&s*, a semi-annual magazine of the Faculty of Arts & Science at the University of Toronto, «ref yr & pp».

[1]Distinctions between digitality and discreteness can and perhaps even should be drawn—but to do so would run beyond the scope of this paper.

cated machines. Indeed, it is exactly the alleged contrast between the cut-and-dried, neat and sharp categories of the formal computational world, and the messy, contested, inevitably metaphorical and, ultimately, "wet" categories of human life-as-lived that drives the wedge, many people would say, between the monstrously mechanical and the sacredly humane.

But is that correct? Are computers, in fact, digital?

And what does "digital" mean, anyway? What would be it for the myth to be true?

1 Perfection and Protection

What does 'digital' mean? That is difficult to say[2]—but perhaps less difficult to picture.

As suggested in figure 1, two things are required. The diagram is essentially metaphorical—using a square wave as something of an icon of the more abstract conception of digitality. The first requirement, signified by the flat top in the middle, is that, to be digital, or to exemplify a digital property—i.e., to occupy the 'digital' region represented by middle square—requires a kind of *homogeneity* or *internal uniformity*. If a computer is in a digital state (0 or 1, paradigmatically, at the "lowest" computational level[3]), then there is not supposed to be any state-internal varia-

---

[2]Good philosophy of digitality is thin on the ground. The two main writers are Nelson Goodman (see for example chapter 4 of his *Languages of Art*, «ref») and John Haugeland (*Artificial Intelligence: The Very Idea* «ref», Introduction to *Mind Design II* «ref», and "Analog and Analog" «ref»). Of the two, Haugeland does a better job of articulating the consequences of digitality (reliability, resistance to degradation, support for perfect copies, etc.), whereas Goodman deals more with what it is to be digital—what something must be like, apparently, at least in this world of ours, to achieve the standards of reliability, unambiguity, copyability, etc. that Haugeland articulates. Goodman also distinguishes *syntactic* from *semantic* discreteness. But questions remain. In order for something to be discrete, for example, must there be a continuous background metric (spatial or temporal?) with respect to which the digital phenomenon is discriminated?

[3]The traditional labeling of the two binary states in a computer is not without problems. They are normally be understood in terms of (or by analogy with) the *numbers* 0 and 1, though it makes more sense of elementary coding and arithmetic practices to associate them with binary *numerals* ('0' and '1'). For an exploration of these and related issues see Smith

tion: no matters of degree, no possibility for the system to be *partly* 0, or *mostly* 0, or *vaguely* 0, or *more-or-less* 0. The machine is either in state 0 or it is not—black and white, cut and dried. Everything is nice, determinate, and clean.

So that is digitality's first aspect: complete (for the relevant purposes) internal homogeneity. The second aspect, signified by the vertical edges in the diagram, has to do with a digital state's boundaries: they must be absolutely sharp. Whether or not a system is in a given state—on or off, 0 or 1, yes or no—must be a totally and completely definite question. Either it is, or it is not, with no room for ambiguity or matter of degree. Systems outside the indicated region in figure 1 are not in state *A*, as surely and perfectly and absolutely as systems inside the region are in that state. Thus the structure illustrates what is never found in nature: an absolute, perfect, 90° cliff.

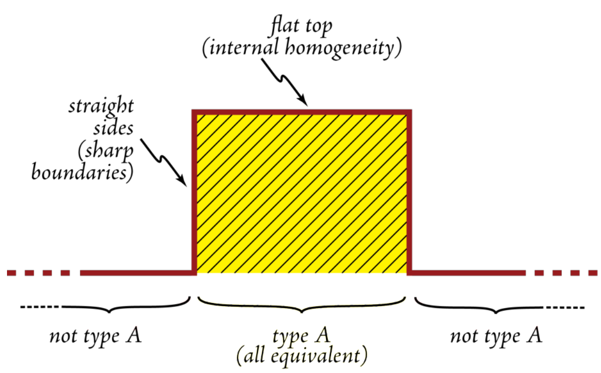


Figure 1 — Classical Digitality

Needless to say, nothing in the real world is quite so neat. But that is all right. In fact that is why digitality is such a metaphysically powerful invention: it is expressly aimed to accommodate such cases. Departure from the ideal is not so much forbidden (which would be difficult to achieve, let alone sell for cents per gigabyte) as it is somehow, almost magically, rendered irrelevant. I.e., the idea is not that things are discrete in some absolute or ultimate metaphysical sense, but that they are *fashioned so as sustain a digital level of description*. Rather than eliminating variation, which would be impossible, we build digital systems by arranging things so that the inevitable individual variation does not matter—such as voltages wandering up and down around some established standard (2.3 volts, 1.6 volts, whatever). To whatever extent is necessary, offending properties are cleaned up, boxed in, confined to certain

(forthcoming)—especially Volume ■■ (Digital state machines).

limits, kept from spilling outside a certain protected region, so that errors do not accumulate or propagate, or the results get out of hand. The trick, that is, is to ensure, with respect to the overall or future state of the system—i.e., with respect to everything that matters about the system at the digital level of abstraction—that all present and future behaviour, such as whether the system will be in state *B*, depends only whether the system is now in state A_1 or A_2 or…or A_i, not on the *way* in which it is in one or other of those states. As long as that condition is met, then any potentially distracting variations will be locally contained—washed away, made invisible. As a result, the relation of the system to the (digital) property of being in state *B* is reduced to a single "bit" of information. Yes or no. On or off. Black or white.

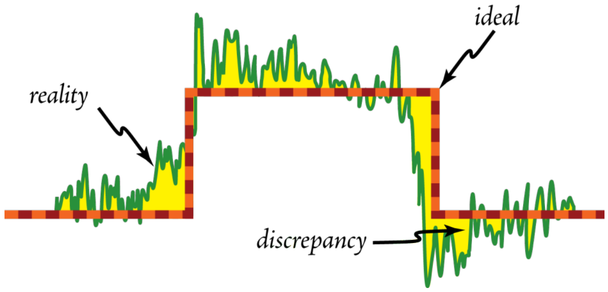


Figure 2 — Digitality as Ideal

You can see what is going on in figure 2. Taking an electrical pulse as paradigmatic, the green line indicates what the electrical circuit is actually like. The dotted red line indicates the "digital idealization." The yellow region indicates the "discrepancy" or "departure from the ideal"—the difference between idealization and actuality.

The amazing accomplishment, for digital systems, is that they are built to work *as if they were red, instead of what they actually are, which is green*. In constructing the rest of the system, that is, or in analyzing its behaviour, you can *assume* that it is red—in spite of the fact that the red line does not exist!

That digital systems can be assumed to be operating in terms of the digital ideal, instead of their concrete continuous messiness, which only approximates the ideal, is a much more impressive achievement than may be obvious—easily, in my view, worth a passel of Nobel prizes. It is certainly far from obvious that such a construction is possible. Normally, though idealizations in engineering are ubiquitous, discrepancies from ideality mount up in their impact. If you were to build a building with this kind of error between how it was supposed to be and how it was actually built, it would likely fall over. If it were a nuclear power plant, it would leak. Digital computer systems, on the other hand, are constructed so that—even with hundreds of millions of parts, changing states billions of times per second, there is not a single case in which, at the relevant level of abstraction, the discrepancies ever "push the system over the edge" into another digital state.

The crucial phrase in that last sentence is 'at the relevant level of abstraction.' Contrary to popular myth, the very lowest levels of computers, far from consisting of adamantine 0s and 1s, are not all that stable. Situations regularly occur where the implementing physical parameters get out of hand, wrecking any simple digital abstraction. Compact disks are a dramatic example: a fingernail scratch can leave a wake of devastation hundreds of bits wide. Cosmic rays and conveyor-belt motors at security checkpoints similarly can produce similar decay, to say nothing of a background slow drift and general disintegration in underlying materials. In a curious sense, in fact, modern digital media are more vulnerable than traditional non-digital ones. As is often pointed out, high-quality paper can last for hundreds or even thousands of years; disk drives are lucky to last ten. Optical media do better, but only somewhat, with current estimates of their longevity running only for a few decades.

How is the digital abstraction maintained, given these inevitable processes of dissolution? An extraordinarily impressive surrounding structure of routines and mechanisms prop up the digital abstraction. Compact disks employ staggeringly complex error recovery schemes to preserve and even recover the idealized digital "signal" in the face of catastrophic tracks of microscopic destruction. Laptop memory is rewritten every fifteen milliseconds, in order that rapidly accumulating "bit-rot" does not take over. Internet packets are checked and resent when they have eroded en route beyond the point of digital recognition. Disk headers are stored redundantly; fragile memories are backed up on disks; mission-critical applications are run in parallel on identical computers, in case one fails. The full gamut of such coding strategies and error recovery schemes is extraordinarily impressive. Certainly the popular idea that a visitor from Mars could examine a single CD and simply "read off" the music is a severe stretch, if not an outright error.[4]

What is digitality for? Why all the fuss? Why construct a system that—at least at this abstract level—is so pure, so crystalline, so fixed? Haugeland gives a particularly apt answer. Digitality, he writes, is:

> "*a method for coping with the vagaries and vicissitudes, the noise and drift, of earthly existence.*"[5]

Discreteness, that is, is more than anything else about **protection—**protection from the ravages and uncertainty and exigencies of the local surround. Things might get cold; winds might blow; the power supply might suffer a brown-out; moth and rust might corrupt; someone at the next table might say something distracting. If you are a digital system you need not care; your constitution guarantees that you will not be unseemingly buffeted by such local aberrations. You will not be unseemly at all, in fact. In a certain sense, digital systems are *intrinsically perfect*.

2 The User Experience

How do we experience the digital? At one level, the answer is obvious, or anyway familiar: we construct programs, automate processes, store data, send e-mail, post messages on social networking sites, interact with other users, manipulate "information." All of these things "exist"—are coherent and intelligible—at the digital level of abstraction. But that is not all. Something else we do, as quickly as we have achieved the digital level, is do our best to hide it.

Think again about CDs—but this time, about the music. For example, think of a recording of Charlie Parker. Or of a compact disc of Thelonius Monk—of *Ruby My Dear*, say, or *In Walked Bud*, or *Straight No Chaser*. Or of a scanned original of a handwritten Walt Whitman poem. Or a recording of a late-night

---

[4]For an account of how this is actually achieved, see «ref AOS volume V.»

[5]Haugeland, John, "Analog and Analog," *Philosophical Topics*, Spring 1981.

phone conversation with a lover. In each case, the medium or substrate will be digital in several respects: frequency, volume, hue. Yet, it does not follow—and this is the point—that the music itself, or the nuances of the image, or the inflection in the caller's voice, are thereby themselves rendered phenomenologically discrete. Rather, what these examples show is that you can implement or encode or represent something non-digital on a digital substrate, but *continue to experience it as continuous*.

This fact—about the relation among one and the same system at distinct levels of description, only one of which is digital—may in the end be as important to the computer revolution as (or even more important than) the simpler fact that there is one level of abstraction at which most computers can be taken to be digital. ⟦…not just below…⟧ And then, "underneath" the digital abstraction, there is another physical level, at which the system or machine is again not digital.[6] The situation is depicted in figure 3. Even if it has grown familiar to the point of the banal, it is still amazing that we can construct a single system—one and the same "thing," a single patch of metaphysical reality—that can be analyzed, simultaneously and correctly, at three different levels of abstraction: (i) a top level, such as music, poetry, and the like, implemented (encoded, represented, constructed, etc.) on top of (ii) a "digital" level (the non-physical digital abstraction or idealization depicted as a dotted line in figure 2, which obeys the criteria of perfect discreteness), implemented, in turn, on top of (iii) a bottom physical level, at which it is again not discrete.

top	· not digital	←	"user"
intermediate	· digital (discrete)	←	"computational"
bottom	· not digital	←	"physical"

Figure 3 — Three levels of abstraction

Arranging things in this triple-decker fashion simultaneously gives you the best of all possible worlds. It is fortunate that the lowest level, the level of the physical substrate, is not digital, since that means we can actually build things out of circuit components, metal parts, light guides, slightly varying components, and

[6]More likely, it will be coherent or intelligible at the level of Maxwell's (continuous) equations.

so forth—i.e., stuff made out of the messy, decaying, material clay supplied to us as the basis of all that exists. If we arrange that layer properly, however, mechanically and dynamically, we end up with a device that, at a higher level, *supports the digital abstraction,* with all of the resulting perfection discussed earlier: freedom from buffeting, protection from the ravages of time, insulation from unwanted or unwarranted influence. The astonishing part is that this protection from the world's dishevelment apparently extends upwards to all levels implemented on top of it. And yet—and this is the crucial part—this immunity of upper levels from buffeting and decay is accomplished without requiring that the higher level phenomena itself (the music, the meaning, the caller's sotto voce intimations) themselves be rendered (at least experientially) digital or discrete. In virtue of being "digitized," that is, the music, meaning and intimacies need in no salient way themselves be neatened, straightened up, clarified or disambiguated. No boxing on the ears is required in order to force them into the strictures of the discrete.

When we talk about "digitizing" music and art, in other words, strictly speaking we are using shorthand for "digitally encoding." To render the music itself digital would mean taking away from the Bird the ability to transform one melody continuously into another, or to build gradually from a whisper to a growl, or to have every performance of the "same" tune be unique. Fortunately, digital music does not require that.

Overall, I believe that the simplest way to understand the achievement of the digital age is in terms of figure 3's three-level structure. This is what our future rests on: an intermediate level of digitality, sandwiched between a lower, non-digital level of the brutely physical, subject to inexorable material buffeting and decay, and an upper, non-digital level of music, meaning, social praxis. Between the two lies the abstract, but terrifically consequential, intermediate, digital level, which, by virtue of its achievement of almost magical perfection, affords the upper level complete protection from the ravages of the underlying lower-level physics, thereby enabling arbitrary mobility, perfection and replication, without requiring that that upper level itself be digital.

The *protection* of the digital without the *price* of the digital—that is what the intermediate level provides to everything above it. Moreover, and non-trivially, given that we have the intermediate level of digitality, we can use it to harness the almost arbitrary powers of algorithms, programming, data, and information processing, in order to engender limitless patterns of transformation and interaction, configured so as to instill arbitrary creativity in the uppermost level.[7]

It is a three-level confection of historic power—with society, needless to say, dining out on the results. And remember: the different "levels" are not separate, modular pieces of an integrated whole. They are all the very same system or phenomenon, analyzed at different levels of abstraction.

… Figure out how to incorporate the following section into the foregoing (it is from a different version) …

3 Sustaining the digital abstraction

Are actual computers digital? Do they meet this ideal standard?

In one sense the answer is yes—but to a much lesser degree, and in a much more complex way, than is normally imagined. Contrary to popular myth, the lowest physical levels are not all that stable. Situations regularly occur where the implementing physical parameters get out of hand, wrecking any simple digital abstraction. Compact disks are a dramatic example, where a fingernail scratch can leave a wake of devastation hundreds of bits wide. Cosmic rays and the conveyor-belt motors at security checkpoints can similarly produce decay, to say nothing of a background slow drift and general disintegration in underlying materials. In a curious sense, in fact, modern digital media are more vulnerable than traditional non-digital ones. As is often pointed out, high-quality paper can last for hundreds or even thousands of years, hard disks are lucky to last ten. Optical media

[7]For example: digital "filters" and algorithms are now regularly employed, at the digital level at which music is encoded, to perform adjustments that are intelligible at the higher, implemented level—such as subtracting a soloist (for Karaoke), compensating for room acoustics, adding echoes or other fabricated artifacts, etc.

do better, but only somewhat, a best lasting a few decades.

Given these inevitable processes of dissolution, a surrounding structure of routines and mechanisms put in place to preserve—and prop up—the digital abstraction. Optical disks (such as CDs, DVDs, and Blu-Ray) employ phenomenally complex error recovery schemes so as to preserve and even recover the idealised digital "signal" in the face of microscopically devastating tracks of destruction. Internet packets are similarly checked and resent when they have eroded en route beyond the point of digital recognition. Disk headers are stored redundantly; fragile memories are backed up on disks; mission-critical applications are run on multiple 'identical" computers in parallel, in case one fails. The full gamut of such coding strategies and error recovery schemes is extraordinarily impressive. Certainly the popular idea that a visitor from Mars could examine a DVD, for example, and simply "read off" the music is a severe stretch, if not an outright error.[8]

In general, that is, the "digital" level of abstraction-the level at which two copies of the "same" CD are identical, for example—is higher (more abstract) than the level at which they are physical tokens. It also takes clever design and on-going work to maintain. This is one reason why different pressings of the "same" CD can sound different, different digital pressings of the same print look different, etc. We see and hear at the lower, continuous, physical level-and so we are vulnerable to what is digitally ignored: the ineliminable roughness, the necessity of approximation, contingent particulars of the given concrete token.

4 Conceptual discreteness

From what has been said so far, you might take the conclusion to be this: that (i) while nothing is *physically* digital—i.e., discrete at the underlying physical level, (ii) we can, nevertheless, build physical things to sustain a digital (i.e., "computational") level of

[8] The situation is more than a little bit reminiscent of what has happened with regard to our understanding of DNA. Whereas it was first (mistakenly) thought that dna "contained" all the information about the structure of the phenotype, it has more recently been recognised that this idealisation is quite severely awry. Only within the context of a surrounding pool of RNA, proteins, etc.—all structures "encoded for" by the DNA itself, of course—can the "code" within the DNA be interpreted or effective.

abstraction, (iii) on top of which we implement all kinds of non-digital things. Doing so gives these implemented things an unprecedented degree of stability and mobility—even virtual perfection. Society's slogan, on this view, should be "The Digitally Implemented Age," not "The Digital Age." And that is where things would stop.

It is not bad, as a first cut—but even it is wrong. And this time, it is a major falsehood—or perhaps we should say, an *expensive* falsehood. Getting over it will cost a great deal of the modern intellectual tradition.

The problem is that there is a more abstract form of digitality—what Haugeland calls "higher-order digitality"—that applies, not to the specific waveforms and measurable quantities of a concrete phenomenon, but to the very concepts themselves, in terms of which things are explained. Thus, consider force, mass, velocity, charge—staple concepts in physics. Specific forces and velocities can be as continuous as you please (23.759 kilograms, $0.3335640951981521 \times 10^{-8}$ seconds, etc.). However, the concepts in terms of which such things are analyzed are as pure, discrete and distinct as any digital states: nothing is ½ of a force and ½ of a mass, or partway between a momentum and duration. The concepts of physics are like the monoliths at the opening of the movie *2001*: unadulterated and distinct.

To make this concrete, I will call a concept **higher-order digital**, or **higher-order discrete**, just in case, to continue using the vocabulary form figure 1: (i) it is internally homogeneous, in the sense that there is no matter of degree, no "internal" structure, to its exemplification; and (ii) its boundaries are absolutely sharp, in the sense that whether or not something exemplifies the property is a clean, pure, absolute, binary, determinate, yes-no issue. These properties are to be contrasted with being **first-order digital** or **first-order discrete**, which would hold in case the concept or notions takes the entities that fall within its extension to be discretely divided. Thus in classical physics, the notions of mass and velocity are first order continuous but higher-order discrete, since both masses and velocities can come in any real measure,[9] but as

[9]Remember that this is classical dynamics, not quantum mechanics.

already noted there is no such thing as being somewhere between a mass and a velocity. The informal division of the day into morning, daytime, evening, and night, however, is a system of concepts that in contrast are first-order discrete but not higher-order discrete,[10] since they *do* divide the day into four discrete chunks, but not in an absolutely principled and dichotomous way; whether a given time is night or morning (such as in the early dawn light) is not an absolute question; it is not meaningless to say call such a time *partly night*, and *partly morning*.

For a more complex example, consider gender. "Being male" would be higher-order discrete just in case: (i) there were no facts of the matter, indeed no coherence to the idea, about *how* male something or somebody was; (ii) there was no internal structure to a given particular person's *being* male; (iii) if the "way that Andy is male" and the "way that Bill is male" were wholly interchangeable; (iv) just in case some things were male, and some things were not male, but no things—because of the verticality of the edge or boundary—were *ambiguously*, or *vaguely*, or *partially*, or *unstably*, or *contestedly*, male. And as the articulation makes clear, these absolutist criteria are not conditions that the (at least present-day) concept of gender meets.

By the same token, consider the notion of *arrogance*—and the boundaries between it and various nearby notions, such as *pride*, *egocentrism*, *self-confidence*, *braggadocio*, and the like. Once again, sharp edges do not apply. Nor is the issue just epistemic—an issue of uncertainty, of unclarity in the judging whether someone is one or other. More strongly, the point is that the concept itself is not—and could not be—sufficiently precisely determined for there to be an exact metaphysical answer as to whether someone is arrogant or not. Moreover, the internal structure of arrogance is not uniform, either—implying that the concept is not internally homogeneous. People are more or less arrogant, arrogant in this or that particular way—in ways that make a difference, not only in general, but in particular *with respect to their arrogance*.

[10] I do not say they are higher-order continuous. Articulating the conditions on conceptual continuity is a more difficult project than can be taken up here. One consequence of this way of analysing things, however, is already evident: that digitality (discreteness) and continuity are not precise opposites, nor do they form a mutually exclusive exhaustive pair.

It might be thought that these examples are useful because of the ways in which they contrast with the computational situation. Computers, many people think, are distinctive exactly because, unlike people and perhaps other naturally occurring organisms, they *do* exemplify such perfected qualities: neatened-up categories, binary distinctions, clean edges. Many people think, in fact (including John Haugeland, in the paper cited above) that computers are *deeply* digital—not just made up ultimately of zeros and ones, in the sense discussed above, but much more generally that *whatever* properties computers have, in virtue of being computational, they have in a perfectly determinate manner. They either are push-down automata or not, universal or not, terminating or not. They either will or will not run Microsoft Word, are or are not connected to the internet, will or will not reboot after a crash. In no case—or so at least the official story claims—will the answer to a constitutive computational question be "sort of" or "somewhat" or "more or less."

It is exactly because of this presumptive (higher-order) absoluteness, moreover, that computer science is widely thought to be a formal discipline, that the study of computers is considered to be scientific, etc. At the same time, the same presumptive (higher-order) absoluteness is what makes computers, in many people's eyes, dry, desiccated, and inhuman. I.e., it is exactly the contrast between the cut-and-dry, neat, sharpened categories of the formal computational world and the messy, contested, inevitably metaphorical, and ultimately "wet" categories of human life-as-lived hat drives the wedge between the (monstrously) mechanical and the (sacredly) humane.

5 Computational categories

From what has been said so far, you might think that computers, *qua* computers—i.e., computers *at the computational level of description*—would all be digital, even if we use them as a substrate or vehicle or representation for other non-digital phenomena, from music to thunderstorms to politics to the digestive processes of T-cells. Or, to put the same point another way, you might think that all computational properties would be (higher-order) digital—clear, distinct, sharp-edged, as metaphorically intimated in figure 1. Not only *could* you think that; many people *have* thought it; I myself thought it, for many years. It is a very common view. But it is wrong. At the higher-order level we are now talking about, it is simply false that computational properties are discrete. It is a major falsehood, too—or perhaps we should say it is an expensive falsehood. Getting over it will cost us all of modern metaphysics.

To see why, it is useful to consider a variety of notions in terms of which computation is classically analysed. In each case, I will argue the same thing:

1. The intellectual mythology we have inherited, what it is fair to call the **formal tradition**, in terms of which we presently understand computers, has viewed this distinction (i.e., whatever notion we are discussing) as higher-order discrete.

2. In point of fact, however—in the actual, lived cases of what I will call **"computation in the wild**"—the distinction is not discrete.

3. Not only *is* the notion not discrete; it is *crucially* not discrete. The fact that the systems we build are possible, useful, realisable, interesting, and economically viable depends on the fact that the distinction in question is not, when you actually look at it, sharp-edged, cut-and-dried, determinate-i.e., is not a black-and-white yes/no affair.

These are strong claims, which ultimately require strong arguments. But it is not hard to develop an intuitive feeling for what is going on.

The problem is that actual computer systems deployed in real-world situations betray the fact that a large number of computational categories, in spite of being built on top of our now-familiar abstract form of discreteness, are more like arrogance than they are like mass. Consider four notions fundamental to the analysis of any real-world computer system:

1. **Subject/object**—and allied notions of representa-

tion/represented, symbol/referent, sign/signified, and so on

2. **Form/content**—syntax/semantics
3. **Inside/outside**—internal/external, intrinsic/extrinsic
4. **Abstract/concrete**

In each case, concrete, lived experience (rather than theoretical constructs built on assumptions to the contrary) shows that they are far from being neat and clean, "clear and distinct"—i.e., digital—concepts. That is not to say that these (or a host of other such) distinctions are useless, inapplicable or untenable. The point is just that, at best, they demarcate a complex, intermediate region or territory—not a "gradual" or "continuous" or "smooth" compromise, but rather a turbulent locus of ferment and activity, a place where things are stretched and pulled and splintered into a thousand other considerations, considerations that no longer line up and pull in one direction, nor line up and pull in the other, but sunder, cross-fertilize and lead to more distinctions—all the way (as it is said) up to "the edge of chaos."

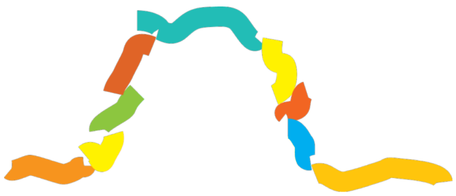


Figure 4 —Boundaries breaking

Ultimately, instead of being discrete, the situation begins to resemble that depicted in figure 4.

Start with the three distinctions listed in figure 5: (i) between a symbol and its referent, (ii) between syntax and semantics; (iii) between the inside of a system and the "external world" in which it is embedded; and (iv) between things that are abstract and things that are concrete. All four are implicated in the analysis of any interpreted or representational system, including not only computers, but also people, and at least arguably such other things as language, books, and e-mail.

The first distinction, between symbol and referent, gets at the ineliminable fact that any interpreted or intentional system describes, represents, encodes information about, or is in some other way “oriented” towards a task domain or subject matter. The second, the more abstract or conceptual split between syntax and semantics, separates concerns about how a system works (i.e., issues about “form,” material embodiment, and causal effectiveness) from more distal or interpretive questions having to do what the symbols mean or represent. The third, between inside and outside, is in some sense even more basic: it is implicit in the very idea that the system is a system or entity at all. What is “inside” the body or skin or rack panel constitutes the system itself; what is outside is labelled the context, or environment, or external world. And the fourth …

	Primary	**Allied**
1.	*symbol* , *referent*	sign/signified name/named representation/represented
2.	*syntax* / *semantics*	form/content
3.	*inside* / *outside*	internal/external intrinsic/extrinsic
4.	*abstract* / *concrete*	

Figure 5 — Traditional distinctions

Again, all four distinctions are very general, as applicable to people and human activity as to any conceivable artificial mechanism. What makes computers special, however, according to the logical and metamathematical traditions from which computer science has inherited its explanatory frameworks, is that these distinctions are thought to apply to computers *in a distinctively discrete way*. First, the inner symbols themselves are thought to be discrete. Second, the categories in terms of which they are analysed, such as meaning and semantics, are thought to have exactly the sorts of sharp boundary and internal homogeneity discussed above. Third, the three distinctions are taken to be aligned, with the symbols, especially with regards to their syntactic aspects, imagined as being on the inside of (and thus part of) the computer; and the referents, implicated in the semantics, on the outside (and therefore not part of it)—again, in a neat and uncontentious way. Fourth, the inner realm of symbols is taken to be abstract, in contrast to the presumptive concreteness of the external realm of referents. Fifth, and finally, the divide between the two realms—between the pure, inner world of discrete, abstract symbols, and the messy external world of concrete referents—is viewed as something of an explanatory moat: a gulf across which theoretical dependence does not cross. Moreover—to put the icing on the cake—it is exactly in virtue of allegedly having this neat overall structure that computers are taken to be scientific: amenable to rigorous, mathematical analysis.[11]

6 Computation in the wild

But is it true? In practice, is the computational realm so neat? No, it is not. And the reasons cut deep.

… figure out what is first-order, what higher-order? does it matter? i think so…

To see why, consider what is perhaps the simplest imaginable counterexample (too simple, perhaps, to convince anyone—but maybe still illustrative): an elementary case of counting. What counting illustrates, in a way that doing sums does not, is a computational process that *actually interacts with its subject matter*—namely, with an exemplified situation of some number *n* of objects. When you count five elements in a list, you end up with a representation or numeral '5', designating five. But what you start with is an actual *number* of elements—not a "*numeral*" of elements, a phrase whose very awkwardness betrays the fact that it makes no sense.

And counting is just the tip of the iceberg. As all practitioners know, it is impossible to separate computers from the worlds they represent. Computers are so involved in their task domains, in fact, that it is impossible to sort their interaction into the traditional categories of reason, action, and perception. It is not even enough to generalise to a broader notion of experience. Just think of e-mail, of file systems, of network traffic nodes, of display cards and window systems and run-time compilers. Computers **par-**

[11]The impact of this alleged divide gets carried over into other fields. Thus cognitive science, based on the hypothesis that minds and intelligence are computational, makes an analogous distinction between: (i) "narrow," brain-oriented, psychological phenomena, assumed to be wholly mechanistic, intrinsic, mathematically analysable-the subject of scientific psychology; and (ii) "broad," social, relational, allegedly non-psychological phenomena, usually left to sociologists or anthropologists or historians. Many critics have argued that these assumptions do not hold in the human case—and thus that people must not be computers. The claim in the text, however, is that they do not hold of computers, either.

ticipate in their subject matters: they muck around in, create and destroy, change, and constitute, to say nothing of represent and reason and store information about, a hundred realms—new realms, some of them, that owe their existence to the very computers that interact with them. In fact computers are so thickly engaged in their subject matters that it can even be impossible to draw a stable inside/outside boundary. Are the windows on the desktop inside or outside of the computer? What about the disk drive? the file system? the backup tape? the network?[12] Similarly, the boundary between sign and signified, and the corresponding theoretical boundary between syntax (in the generalised sense of the realm of the effective) and semantics (in the similarly generalised sense of a distal realm of that with which computer systems are normatively enjoined to coordinate) is about as far from sharp as it is possible to be. The two sides interpenetrate, not so much in gradual shades of gray as in a profusion of middling, "hybrid" intercalations.

And so the situation, instead of being discrete, begins to resemble that depicted in figure 3 ⟦4?⟧. At least with respect to these first three classical distinctions, that is, real-world in vivo boundaries are far from being clean and sharp. At best, the three notions demarcate a complex region or territory—far from being even "gradual" or "continuous", but rather a locus of ferment and activity, a place where things are stretched and pulled and splinter into a thousand minor considerations, considerations that no longer line up and pull in one direction, nor line up and pull in the other, but sunder, lead to more distinctions, and may even be best described as on the edge of chaos.[13]

With respect to this moral, moreover, there is nothing special about these first three distinctions. Much the same story holds

[12]Computers "shake hands" in the same medium as that in which they think. It is as if we humans, upon encountering a friend, could plug our nervous systems together directly—i.e., as if we had "ports" on our nervous systems—without having to transmit the signals through a different underlying medium. Perhaps, if we had developed to perform such feats, the individuation criteria for people would be as messy as they are for modern machines.

[13]«. Reference, if it is possible to do so coherently, some of the Santa Fe work.»

for any number of other constitutive computational properties. Thus consider *abstraction*. For many many years, it was assumed that the way to build complex systems was in terms of so-called "black boxes"—abstractions that presented a fixed and given interface to the outside world, but that completely hid within themselves all internal "details of implementation." As usual, the idea of discrete black-box abstraction had a certain theoretical appeal. But in practice it, too, has turned out to be an unworkable idealisation. As every professional programmer knows, no matter how elegant the formal or explicit interface to a virtual machine, inner implementation details invariably "shine through" and affect the systems built on top of them, in ways that often have dramatic effects on performance. To make a program run fast, that is, you don't just need to know the formal definition of C++; you also need to know (or have experience with) how it is implemented. This non-opacity of abstraction boundaries is even gaining theoretical recognition, leading to the design of fancy mechanisms that allow programmers access to the "innards" of the underlying level. Some have even suggested replacing the notion of a black-box with something like "gray box" or "glass box," in order to legitimate making the workings of the lower level visible.

In the wild, that is, what had been theoretically allegedly to be a fixed, discrete boundary turned out in practice to be something quite different: a locus of negotiation, of communication and sharing of advice, a region rather than a line, where responsibilities and information are exchanged—far more like a market or town square, where consensual agreements are hammered out and maintained in real time as things progress, rather than the pure line of fixed abstraction that the intellectual heritage imagined.

7 Logic

And so it goes—to deeper and deeper levels. Not only do specifically computational properties fail to be discrete, as we have seen, but the same moral applies to more general distinctions of which computer systems are sometimes used as models: between nature and society, between the sciences and the humanities, between subject and object, between mind and body. Computers are wonderfully disruptive precisely because they make a sham of the ultimate sharpness of every one of these classical dualisms.

Computers are symbol manipulators par excellence, for example, but does that mean they validate those who claim that language is merely an endless play of signifiers? No, they do not; they spend too much time mucking around in their (semantic) task domains. Computers are supposedly objective and natural, or at least naturalistically palatable—i.e., scientifically OK, intellectually respectable, not too spooky. But the stories we tell about them are so thoroughly peppered with intentional vocabulary (*programming languages*, *data bases*, *information highways*, *knowledge representation*, *symbol systems*, and on and on) that this alleged "respectability," intuitively reflected in the claim that computers are "mere machines," may ironically turn out to be sheer prejudice. It is particularly curious that at the very same time that their alleged objectivity recommends them, philosophically, as naturalistic (i.e., as one with the sciences), at the very same time they are candidates for a theory of what it is to be an intentional subject, because of their manifest representational character.

The failure of discreteness even applies to some of the most foundational distinctions on which all of logic, mathematics, and science are thought to rest: existential distinctions, between and among objects themselves, and logical distinctions, such as that between objects and the properties or types they are taken to exemplify. Formal logic, mathematics, science, and a good measure of modern philosophy, in particular, not only presume a background of objects with precise black-and-white individuation criteria, but even more seriously assume that the goal of scientific discourse is to delineate the objects, categories, and properties in the world so that their boundaries are higher-order discrete, in just the absolutist sense we have been wrestling with. It is this, I believe, that computational experience had shown us, and will increasingly show us, to be an impossible, out-of-date, and ultimately futile game.

At a workshop on representation a few years ago in England, a philosopher argued that philosophy, taking a lead from science, should insist on a very strict notion of object (on clear definitions, precise identity criteria, and the like). As a working scientist, I could only muse that in two decades of wrestling with the essential structure of computing, which is at least a candidate for the most important scientific, let alone intellectual, development of the twentieth century, I had never found any such distilled, lapidary objects. The identity criteria on computational objects simply do not honour this formalist ideal. Think about the property of being an x86 microprocessor,[14] necessary in order to run Microsoft's Windows operating system. Enormous effort goes into defining the exact operating specifications of such commodity chips. And yet numerous issues about what it is to be in this class remain unanswered (as clone manufacturers are continually discovering, to their dismay). Nor is there any reason to believe that the answer is temporal stable. This is the realm of copyrights, patents, and million dollar lawsuits. Among other things, the answer depends on who is asking. For it is widely recognised that to be a legitimate instance of a particular architecture is in part a commercial and political question, involving issues of market share, advertising power, and the like. No two runs of a single chip design are absolutely identical, let alone explicitly different versions, or allegedly the same version from different manufacturers. Even within single companies, in situations when all the market forces press for a common chip type, it turns out in practice to be impossible to guarantee that unresolved boundary cases will not emerge. And what is true for hardware is doubly true of software. That is why software is maintained; it takes money, power, and influence to preserve the identity of a program over time.[15]

---

[14]The processors that power most personal computers—from the Intel 8086 through the Pentium up to present day Core i6s and i7s, and similar offerings from AMD.

[15]Curiously enough, moreover, very much the same conclusions—about the lack of strict individuation criteria, and the concomitant breaking up of the object's boundary—arise in even simple cases of arithmetic.

I once designed a programming language that, unusually, attempted to maintain strict use/mention distinctions among (i) numbers, (ii) internal structures that designate numbers (internal 'numerals', essentially), (iii) external expressions (like '234') corresponding to those internal structures, (iv) distinct copies of those internal structures, (v) pointers to those individual copies, (v) and so on and so forth. By the same token, a similar set of distinctions was made among sequences, internal structures that designated sequences, external character strings that notated the internal structures that designated sequences, etc. Not only did these distinctions crosscut; they were in turn crossed with several other familiar sorts, such as between types and tokens and instances and uses. To what end? Total con-

Needless to say, what is true of computing is even more true of human experience. Suppose, for example, on a camping trip, after gazing at the sky, that you turn to your companion and say "we probably shouldn't attempt the ascent today; there are clouds covering the north side." And suppose, further, than your friend, having nothing better to do, asks the following pedantic question: "OK; you've been to college; how many clouds are there, exactly?" Your inability to answer cannot be ascribed to merely epistemic doubt or lack of knowledge. Nor does it mean there was anything wrong with your original statement. There is no reason to suppose that there need be any metaphysical fact of the matter—any metaphysical fact as to whether, in some region of the sky, the arrangement of foggy air should count as one cloud, or two. If, as I believe, this is right, then it must be that the competent use of the English plural does not metaphysically require a set of discretely countable individuals in order to be true. The same would apply to a claim that "a program still has bugs." The truth of that statement does not depend on there being strict individuation criteria on bugs—and bugs are surely as computational a concept as one could please.

The distinction between type and token is similarly crumbly, in lay experience. It is not just that the traditional two-way distinction is not adequate—between abstract type or category, on the one hand, and concrete token or instance, on the other. Nor is it enough to spawn a three-way distinction among type, token, and use—or even one able to deal with more complex cross-cutting spatial and temporal fan-outs of interpenetrating abstrac-

---

fusion! The result was impossible to use.

Semantical clarity, or at least something resembling it, was obtained at the expense of sanity. It turns out that what one wants—and as common sense anyway suggests, at least on reflection—is a system that makes *whatever distinctions are appropriate, in the moment, for the purposes that at that time are being served*. Distinctions need to be made on-the-fly, in response to particular circumstances, not inflicted, as if that were even possible, all at once, at the outset. In real systems, that is to say, with anything approaching the complexity of modern software (note that even Xerox copiers now have multiple millions of lines of code), individuation criteria, and thus object identity, are themselves context-dependent, negotiated, and maintained.

Not even ontology is sacrosanct.

tion.[16] Instead, imagine getting up one morning and saying, drearily, "oh no, I still have a headache." Or: "the fog is coming back." Or: "the wind from that direction is typically warm." How are we to understand the referent of the singular noun phrases: 'a headache,' 'the fog,' 'the wind'? There is no reason to suppose that they refer to *types*, in the sense of something that can be "tokened." Nor is there any reason to suppose that they refer to *tokens*, in the sense of something that is of a type. Rather, there is no reason to suppose that the distinction between type and token, or between object and property, in the lived world, is any more of a "discrete" way, with any more sharp and absolute and black-and-white boundaries, than any of the others we have already seen.

Jericho once again. As in figure 4, the boundaries start tumbling down.

Ultimately, in fact, it is wonderful historical irony. Computers are supposedly objective, scientifically "OK"—intellectually respectable, naturalistic, not spooky. It is in virtue of this pedigree that they are *echt* denizens of the modern academy. But this alleged respectability, so innocuously garbed in the idea that computers are "mere machines," may turn out, historically, to reflect no more than sheer prejudice.

Loosed into the wild, computers play the trumpet outside the digital walls of Jericho. The boundaries of conceptual discreteness are tumbling down.

[16]Consider a simple program for computing factorial:

```
procedure factorial(n)
   if n=0 then 1
      else n*factorial(n-1)
```

Suppose that this program is called with the argument '5'. With respect to different readings of the term 'the variable n', there is something of which there is one, something (*spatial* fan-out) of which there are three, something (*temporal* fan-out) of which there are five, and something (both) of which there are fifteen.

Why does it matter whether the digital level of abstraction is "real"? That much of what we call digital is neither physically nor experientially digital, but only digitally implemented? That the concepts and categories of computing are not conceptually discrete?

In part, the answer stems from a point with which we started—that notions from the computer revolution, such as digitality and information, have assumed such importance in our collective imaginary. As said there, many people assume there is a fundamental (discrete!) divide between people and computational "machines"—that the latter, by virtue of a presumptive neatness, formality, and cut-and-dried conceptual structure, have no purchase on the contested and metaphorical "wetness" of human existence.

I would be the last to claim that anything anyone has built so far can manifest care, chuckle ironically or make a surreptitious gesture. But it is not a fact from which I would extract metaphysical comfort. We have a long history, after all, of striving to maintain the human as fundamentally distinct from the other systems with which we share our habitat: the heavens before Galileo, the animals before Darwin. Reaching for non-discreteness as a way to secure us from the encroachment of the Information Age is just as likely, in my view, to be grasping at metaphysical straw.

Any importance (and humility) that we humans are worth must stem from concrete facts about our actual existence, not from any presumptive immunity from being reproduced—or perhaps more elementally, from belonging to the world.

8 Successor metaphysics

Enough negative claims. It is boring, ultimately, to say how the world is not. Much more important—to say nothing of more fun—to see how it actually is.

I began by saying that everyone "knows" that computers are discrete. I argued that they are wrong. But it is not digitality *per se*, that has been my primary target. Rather, this investigation grew out of what initially seemed like a much more general project: to understand *formality*, and the even more widespread consensus that computers are formal (that they themselves are formal, that they must be studied formally, etc.). It was evident to me, from the outset, that 'formal' is an amazing—and assuredly *non*-formal word—not a notion that will ever succumb to clear definition. Depending on how you count, there are anywhere from two or three to a dozen distinct meanings of the term—meanings such as "independent of semantics," "abstract," "able to be mathematically modeled," "purely ideal" (as in Platonic forms), and the like. Over many years of trying to make sense of them, it gradually emerged that what lay underneath these various readings, and tied them together into a coherent group, was their common presumption of exactly the sort of higher-order discreteness under discussion here.

Cognitive science's interpretation of 'formal' as meaning "independent of semantics," for example, turns out on sustained analysis to come to neither more nor less than the abstract claim suggested earlier: that computational systems are (allegedly) distinctive, among semantically interpreted system more generally, in that the divide between the syntactic and the semantic is sharp—engendering the claim that in the case of such systems the two realms are "*independent*." Similarly for the "abstract" reading: many things in the world, such as hospitals and birthday presents, are defined at a higher level of abstraction than the purely physical. What makes something like a number or type (but not a hospital) *formal*, in the time-honoured sense of being *abstract*, is the claim that the divide between the abstract object and any physical realisation or instantiation of it is (once again) sharp or absolute.

In fact, if I had to reduce the last century of logic, set theory, mathematics, (academic) computer science, and so forth to a single phrase, I would say the following: that

Formality is discreteness run amok.

Every one of those different readings of formality rests on an assumption about the existence of a strict, black-and-white, cut-and-dry, discrete distinction. Thus a strict *subject/object* split is presumed by the scientific method; a strict *syntax/semantics* split, alleged in cognitive science and the philosophy of mind; a strict *abstract/concrete* split, assumed in recursion theory and the theory of computability. By formal ontology, similarly, is meant ontology where the individuation criteria are discrete—the same presupposition that underlies the rather general reading of 'formal' as 'capable of being mathematically modeled.' Note that physics, the calculus, and continuous mathematics are all formal by this count, as well. Admittedly, these fields license continuous values, but the prior and constitutive higher-order questions, such as whether x is or is not equal to 0.32157, are assumed to have *precise, determinate, yes-no* answers, of exactly the sort that we have been considering here. Nor is repairing to probabilities of any help; the probability of whether a given event P will happen may be 0.62, but the boundary of that 0.62 is as sharp and discrete as any we have yet seen. So too, by the same criterion, are the assumptions of fuzzy logic—still discrete and formal. None of these "weakenings" are anything like strong enough to escape the grip of the formalist tradition.

But that observation in turns points towards the sort of picture I want to construct in its place. To see how it might work, note that if I ask you to write your name on the wall, here, next to where I am working, it does not follow that there is any ambiguity about where I am pointing, just from the fact that there is no discrete fact about my description's reference. And if there is a problem, no doubt we can talk about it, work it out. Only prejudice says that intellectual inquiry must start with the discrete, i.e., with the digital, and build everything up on top of that. Yes, that is the current practice in logic and mathematics: one sets out with discrete sets, constructs the integers out of them, defines continuity as limits of infinite series of discreteness, and models vagueness on top of that-an overall strategy captured in Kronecker's famous dictum that "God made the integers; all else is the work of man." As far as I am concerned, however, *Kronecker got it almost exactly backwards*. Discrete integers are the work of (yes) man; God made everything else.

So here is what I want. I want to start over, at a violently rupturous beginning—a feisty, obstreperous, riotous fount of an overwhelming mass of stuff. That is where we live; that is what we are made of; that is what we inhabit. And that is where I want to ground metaphysics—and to do so with no prior commitment to reductionist formality. In fact I want no prior commitment to any distinctions at all. Not to rational foundations, not to mathematics (which will anyway have to be overhauled), not to the transcendental a priori, not to the very very small. No discrete distinctions whatsoever should be presumed in advance—between might and reason; among truth, beauty, and goodness; between intentional directedness and the directedness of obligation or duty or awe. For to do any of those things would be to build in discrete formal boundaries at the outset. And that, in turn, I am convinced, experience with an in-the-wild practice shows to be a mistake. It is only through lived, complex processes of stabilisation, of domestication—perhaps of taming and of tilling—that we partially and constantly work our patterns in the flux: register objects, temporarily set up negotiable borders, live, practice, and carry on our decidedly informal and indiscrete affairs.

— Were this page blank, that would have been unintentional —

Epilogue

— Were this page blank, that would have been unintentional —

Interview

Text of an interview of Brian Cantwell Smith by **Gordy Slack***, conducted on February 16, 1997 for the Center for Theology and Natural Science of the Graduate Theological Union, Berkeley, California, in preparation for a workshop (held June 8–10, 1997) on the relation between computer science and theology. An abbreviated version was published in …*[1]

GS • Would you say a few words about your religious background?

BCS • Well, a few words aren't going to suffice, because the issues interpenetrate a lot of what I do. But let's start with just the facts. I grew up as a member of the United Church of Canada, which was a single church formed (before I was born) out of a merger of the Congregationalists, the Methodists, and half the Presbyterians. On top of that, my father was a theologian—technically also an ordained minister, though he worked as an academic, not as a preacher. In lots of ways, I've been connected to his work. In fact, even though I've worked in (and been under the influence of) the sciences, there's a sense in which you can see me as running the family store. There's a fair amount of continuity, in a lot of the basic issues that come up in my work, between what I believe and his world view: his sense of significance, his sense of what it is to be religious, the theological presuppositions and so on and so forth that I was given as a child.

---

[1]«Ref; and check that it really was abbreviated—and also compare the text there word-for-word with the one here, to see if any infelicities here were cleared up there, in which case this version should be brought up to date.»

That said, it's pretty important to know that my father's theology is radical in a lot of ways. For example, he's written books arguing against the presupposition that propositional belief is at the core of any religious tradition. You can think of propositional belief as "belief that": I believe *that X*, you believe *that Y*, etc., for any *X* or *Y*. Lots of people think that to be religious is to believe certain things like that—for example, to believe that God exists, or that someday we'll go to heaven. In fact most people in this country think that to be Christian is to believe certain things of that form. But for many years my father argued that the tendency, in the modern western Church, to reduce being religious to the assent to certain propositions, is *fatal*. You simply cannot get at what matters about the tradition in terms of propositional belief. So there's a real crisis for the church. That's what he said. And I guess I pretty much agree with him.

So: did I grow up with a religious background? Absolutely. Does that mean I believe in God? Or that I believe this or that? Probably no, to most of those questions. The idea is to get deeper than those questions, not to either assent to them or deny them.

GS • I talked to Arno Penzias last week, who's participating in this project, and he said pretty much the same thing. He's said if you ask a Jew whether they want to become a Christian, they say, "Well what do Christians *do*?" If you ask a Christian if they want to become a Jew, they say "What do Jews *believe*?"

BCS • When the Shah fell, in Iran, the *New York Times* got in touch with my father, because he was an Islamicist, and asked him what Muslims believed. His basic answer was: "If you think that's the right constitutive question, you are guaranteed to not understand the Islamic tradition."

[chuckling] I think the *Times* may have gone on to ask other people.

GS • Of course the *Times* was calling at ten minutes to deadline.

BCS • That's right. Sound bites weren't his forte.

But I thought a lot about these things, as a kid. I remember refusing to be confirmed, at age twelve, because I couldn't believe the things they were telling me at church. Later, soon after I got to college (though I was still only sixteen), I quit going to church entirely. And I haven't really had what anybody in the outside (or inside!) world would call a religious practice since then. I found it untenable for lots of reasons. But I never stopped struggling with these things. In fact the very next summer, when I was seventeen, I was back at home, and I remember asking my father what he thought it was to be religious. His answer was: "*to find the world significant*." That kind of metaphysical and theological question—what is the nature of being? what are the grounds of ethics?—those things have always mattered to me enormously. Pretty much always, but maybe especially when you're a student (me at the time, and students of mine, now) those questions, of where to find grounding, of how to anchor your life, what it is worth committing yourself to doing—they're pretty urgent.

GS • What, then, is your religious practice?

BCS • Well, in terms of what those words mean to most people, the answer is probably none.

GS • What about in your own terms? Can you distinguish between those activities you engage in that are religious and those that aren't?

BCS • No, I don't think that's right (that is, I don't really accept the question). I don't use the word "religious" much. I don't use it much myself; and I especially don't use it much in conversation (unless a whole lot of trust has already been established).

Don't get me wrong. I'm completely prepared to talk about this stuff; it's not that I feel these things are private. In fact, I'm prepared to talk to students about this in class. I think it's critical that these things not be private, in fact. The issue is: what words do the best job of communicating, to other people, the issues in this whole area that really matter? The problem with the word "religion" is that it is such a trigger, both for those people to whom it means a lot, and for those people who are allergic to it. There are lots of both kinds. And my experience is that I don't in general have any more in common with people who are pro-religion (i.e., who consider themselves religious) than I do with atheists, with people who are outright allergic to religious language. In fact I often have more in common with people who *don't* think they are religious.

I'll tell you another story about my Dad. When I told him I was quitting going to church, because I didn't believe the things that they were requiring me to affirm, he said I was probably right not to believe them. "But you know," he said, "the sad thing is that you and your friends are going to lose any vocabulary in which to talk amongst yourselves about the things that matter to you most."

Thirty years on, I can report that he was largely right. A lot of people in my generation, a lot of post—Second World War people, a lot of people like me, have lost any vocabulary that can mean, for them, what it is that the religious traditions meant to the people who thought of themselves as religious. Another thing my father used to say: "If one person says, 'I believe in God,' and another person says, 'I *don't* believe in God,' then it's *impossible* for the word 'God' to refer to the same thing, for those two people" (first order logic notwithstanding!).

GS • Right. And even moving out of the realm of logic, it's highly likely from a psychological point of view that they mean quite different things.

BCS • Yes, it's likely. Of course two people who say they *do* believe in God may also mean different things as well. And that's part of what's really been problematic. But I never answered your question. Do I believe in God? Well probably not. But I guess I think I do have a sense of what that word means to at least some people who do believe.

What about the question of whether I have a religious practice? First of all, and this is kind of important, there's no one facet of life that is reserved for "religious stuff". It's not a distinct subspecies of life to me, so it's not a practice in the sense that each morning I do *X*, or each Friday I do *Y*, or anything like that. It undergirds the whole thing. Second, there's this vocabulary problem. It's extremely difficult to find words that come anywhere close to communicating, with people I know, what it is that "being religious" means to me. One thing I find myself doing is using different words with different kinds of people. You might think that was hypocritical, or opportunistic. But I think it is actually *more* accurate, not less. Still, it's a struggle.

But, in terms of my practice—how do I live my life? what walk do I walk?—and in terms of what I take to be the issues that underlie life, then yes, religious things are kind of total. Absolutely, yes, it's important to me—in all aspects of things.

For example: take this book I've just finished.[2] People who are religious in the sense I mean that word—I'm pretty sure they will find it to be a religious book. People who aren't religious, *won't* find it religious (I hope). That's okay. And it's not because there's a secret or hidden meaning that the quote-unquote "religious" folks will see, that is invisible to the others. That would be *very bad*. That's not what I mean at all. Rather, there's something important to me, something I am trying to get across in this book. The people who don't think of themselves as religious may perfectly well "get it"; they just won't think of that kind of thing *as religious*, because they don't think that what it is to be religious is what I think it (au fond) is. (Probably, like we said at the beginning, these will be the sort of people who think that to be religious is to assent to some weird or spooky sounding proposition.) So like I said: it's not that they'll misunderstand the book; they just won't categorize that kind of understanding *as religious understanding*. And I tell you: that's fine with me. I don't care how people *categorize* it (in fact I'm rather distrustful of categories). What I care about is that we learn how to talk to each other about things that matter.

GS • Taking your father's definition of, or explanation of, leading a religious life for a moment: Do you think that there are people who don't find significance in the world? I mean can you be a human being and not find significance in the world?

BCS • I think that's a terribly important question. But before I answer, I just want to say that we have a tendency, when asking questions like this—I think it's become a kind of cultural assumption—to polarize such issues, to assume that words can be broken into opposites. So there is a tendency, in responding to a question like the one you asked, to think that, say, with respect to the meaning of life, it is something that people either "do" or "don't" find. (Feminist epistemologists talk about this in terms of dualisms or binarisms.) And I think that's a really unfortunate, deleterious

[2] *On the Origin of Objects*, Cambridge, Mass: MIT Press, 1996.

aspect of a lot of the conceptual framings that we academics use. So, I don't want to presume that "significance" is something that either you have or you don't have, in a black and white way. I don't even want to think of it as something that you have in a more or less continuous way. Simple continuity is a pretty paltry way to get at the thick meaning of a fully-lived life.

But given all that: *yes*, I do think that there's enormous dissatisfaction with respect to that question these days—people feeling that their lives are hollow or unsatisfying, people feeling anonymous, people feeling that their social and economic conditions don't give them a chance at a satisfying life, don't welcome them, don't provide them a way to participate, and so on and so forth. You know what I mean; everyone knows what I mean; it's almost platitude to say this sort of thing (though just because it is a platitude doesn't mean it isn't true).

Here's one way I get at it with students. Think about the rise of religious fundamentalism, I say to them, in this country, and in the Near East. You have the Christian right in this country, and you have fundamentalist Muslims in the Near East and North Africa. You may think of these as two separate phenomena, not as instances of the same thing (the press tends to treat them differently). But I think, in fact, there's something very similar going on in both of them. What's going on? Well, you know, no one sentence is going to avoid being glib, but we can caricature it like this: There is a deep unsatisfied hunger in a lot of people's lives, an unfulfilled yearning, where people feel that certain kinds of materialist values, certain kinds of economic values, and so on and so forth, are not, fundamentally, satisfying. Popular values don't give them the kind of anchoring, the kind of grounding, the kind of community, the sense of self-transcendence, the sense of significance, that they would like.

I believe that those movements recognize a palpable and urgent lack, a kind of hunger, a kind of yearning, a kind of frustration, in a certain sense the hollowness people feel. And the thing about these fundamentalist "religious" responses, is: *they're providing answers*. Problem is, they're providing an answer that I find *appalling*. In fact I'm scared stiff of their answers; I think it's really very dangerous, in many, many cases. I think it caters to lots of things, forms of closedness for example, and bigotry, and fascism, and so on and so forth, that I think are just terrible. But what I say to the students is, What are we on the left, what are we intellectuals, what are we academics *providing by way of response to that felt hunger, to that palpable yearning?* If we on the left, we academics, we intellectuals, don't have an answer, then we don't have much leg to stand on to criticize the answer of the fundamentalist right.

So, the question is: what would it be for us to formulate a better answer—an answer that does justice to people, in their plural ways of being; an answer that does not have all of the bad aspects of ideology and fundamentalism that I worry about, an answer that is inspiring, in the literal sense of giving people breath and hope, an answer that answers that sort of felt, that palpable hunger for anchoring, for meaning, for a sense of significance? That's what we need. That, approximately, is what I want to do for the next twenty-five years: I want to help work on formulating an answer to that question.

(Let me put in a footnote here. One of the reasons some non-religious people are so allergic to religion is because they worry about this last way of putting things. The problem, they say, is blunt: economic conditions and social injustice. Any effort to come up with a "religious" response to appalling conditions, to the absence of sustaining work, to street violence and homelessness and so forth, they view as little better than fascism. I want to say that what they say is extremely important: yes, we have to correct economic injustice too; that's part of what I take to be a condition on a palatable answer. On the other hand, I don't think economic conditions are enough. Hollow lives aren't a prerogative of the underclasses.)

GS · I wonder—since science is the place that so many people automatically look when they've turned away from fundamentalist theologies, or moderate theologies. I wonder if such significance can actually be found in science at all? I know the scientists I know best are religious, in the vulgar sense of that word, about subtracting significance from their perspective.

BCS · Well, first let me tell a story, then I'll try to answer the question. The story is about a friend of mine, who's Jewish as it happens, and a very serious Jew at that, who devotes a day or so a week to questions of Talmudic interpretation and so on. It's a very significant part of his practice. He is also a "big-S" scientist; worked for a while at Bell Labs; is now chair of a computer science department. We were good friends in graduate school, and this sort of question—about the juxtaposition of the scientific and the religious—obviously occupied us both. The funny thing was, and it struck both of us at the time, I was completely unprepared to do what it was that he seemed entirely content with. He was viewing his scientific work as in point of fact religious, in a certain (to him) satisfying sense. But somehow I just couldn't do the same thing. I was visiting him, a couple of years later, and at one point I remember bursting out laughing. "I see," I said; "I finally figured it out! For you, what you want your scientific work to be is *worship*. What I want my scientific work to be is *theology*." And we both knew exactly what we meant.

But to get to your question. One of the things that people in science have tried to do is to subtract the issue of value. That's part of the "value-free" mythology of science. Now one immediate counter-argument to the idea of value-free science is that we don't eliminate *truth*, which is a value. It's a big value, in fact. If I come up with a theory that's *false*, that's not supposed to be good! I can't defend myself against you by claiming you were supposed to be value free! So even the most traditional scientist has to agree that some norm is operating in science; mainly the norm of truth. Given that, though, it is interesting to take the Greek separation of values into truth, beauty, and goodness—the three basic normative dimensions of life—and ask why science has hung on to the ideal of truth, and let go of the ideal of beauty and goodness. It is not a trivial question, not nearly as trivial as it might look. But anyway, the classic model of science, the reigning conservative ideology, is that yes, truth (and its cohort, rationality) are relevant in science, but the *other* values, like beauty and goodness and so forth, must be kept out. Actually it's curious; that's not quite right. Recently, in mathematics for example, some people are letting down a bit with respect to beauty. Theorems of mathematics are *elegant*, they say; mathematicians are driven by the *beauty* of the abstract forms. But as for goodness; well, you're not supposed to let *that* in. Scientific theories aren't *ethical*. In sum, it is something of a default modus operandi for science, these days, to valorize truth, subtract goodness, and perhaps allow a little beauty back in, to dance over the elegance of the equations.

I have two thoughts about this. In a minute I want to say a little bit about what I think is happening to the *content* of science, at this particular point in history, because I think we're in the midst of an extremely interesting transformation. But first, I want to make it clear, at the outset, that I am very respectful of why it is that the people who want to subtract values from science want to do that. Me, I don't want to do that, as it happens; I want to argue for letting certain kinds of other values back in (especially ethical ones). But I want to do so in a way that respects why science originally threw them out.

Here's the gauntlet I'm prepared to answer up to, in other words. People who defend a "value-free" science—truth, the whole truth, and nothing but the truth—have perfectly legitimate fears of what would happen if we were to abandon that high standard. "If we let go of objective, scientific truth," they claim, "we will open ourselves back up to prejudice, bigotry, suspicion, obfuscation, lying, and of a whole bunch of other reprehensible things." I hope it is obvious that I agree that those things are terrible. So it is absolutely critical not to go back on those fears. Sure enough, we don't want to re-license inquisitions, or applaud rank subjectivity, or legitimatize the crude and unchecked exercise of political power. Yes, sure enough, it was genuinely liberating for science and rationality to free us, during the Enlightenment, from such forms of oppression. "The truth will set you free"—all that sort of stuff. It isn't garbage.

Problem is, it's not enough, either. No one who is involved in social action thinks that a theory of political power is enough; ultimately you also have to *do* something. And so, if we are to fight *for* the things we believe in, and fight *against* the things we don't believe in (note: this isn't propositional belief, we're talking about here—this is "believe" in the etymologically original sense of "caring" or "giving your heart to"), then we have to be instructed in the ways of power as well as in the ways of truth. And to do that…well, I'm just not sure it is enough to keep the *bad* things *out* of science; it might be time for us to bring some *good* things *in*. All in all, I just sort of feel as if "speak no evil; hear no evil; see no evil" is a tad out-dated, as a form of legitimation. If we are going to struggle for what we believe in, we have to have our eyes open, and be prepared to live a life that is full in terms of *all* the applicable norms and values and powers, not just truth.

So I don't want to let science slide back into a pre-rationalist era. I want the opposite: want to say, to the people who are afraid of how other forces can wreck science, something like this: "You are absolutely right. Those are terrible things. But you don't conquer your enemies by being *blind* to them, by keeping them out. Rather, they're so serious (just look at the society around us) we have to take them on explicitly."…

Anyway, sorry to run on; I just feel strongly about those things. But let me get to the second thing I wanted to talk about: about what's happening with science, as we end the millennium.

Back some time ago, I used this word "significance." There was some malice aforethought in my using that word. Since its rise in the sixteenth and seventeenth century you can sort of think of natural science as having gone through an enormous, several-hundred-year-long ascendance. It's cracking in some places. Since the war and the atomic bomb people worry about the untrammeled success of science, whether it won't do us in, and so on. But nobody could argue against its success. It's been an absolutely spectacular success story for several hundred years.

It's interesting that at the beginning of that movement there was the whole era of the alchemists, who were sort of shunned, who remain unappreciated for many hundreds of years. Once you got Newton, and Maxwell, and got science in place, then the alchemists looked like people doing all this crazy stuff. Now people are coming to realize that the alchemists were very important to the preconditions for the possibility of science. Not in any sort of transcendental sense, but in a pragmatic and perhaps even economic way—as necessary for establishing the conditions that allowed the rise of science.

I think the twentieth century is going to be recognized as the emergence of something that's on the scale of natural science. Namely…well, I don't have a very good word for this, but basically an investigation or inquiry into things having to do with *meaning* or *interpretation* or *symbols* or *representation* or *information*. If you were a philosopher you would call it the realm of the *intentional*. The realm of the "semiotic" might be a better description, except "semiotics" has such particular and strong connotations, in some quarters, that many people are as allergic to it as other people are allergic to the word "religion." But whatever we call it, it is basically a realm of the epistemic or semantic or…well, basically a realm of *meaning*. You see it in mathematics, you see it in set theory, and you see in the realm of the computer, the symbol manipulator or information processor. You see it in psychology, where people are dealing with representations and also processing information.

So my view is that, for the next couple of hundred years, we're going to have the era of epistemic or semiotic or "meaning" sciences, the way that for the last few hundred years we've had *physical sciences*. Of course the physical sciences are often called "natural" sciences. "Natural" is a funny word. I suppose it approximately means "not supernatural." So maybe this new era I'm talking about will also get called "natural science"—in an extended sense. It will certainly be a science of natural stuff, in the sense that if I say "hey, are you coming to the party?" that's a pretty natural thing to do. Meaning things, interpreting things, speaking language, figuring things out, dealing with information—no one can say that doing thing like that is unnatural.

So, let's call the new era natural science, too. That means we could say something like this: "Look, what we've had for several hundred years is physical sciences. What we are now going to have, maybe for another couple of hundred years, is a new kind of natural science, to go alongside the old one: something like semiotic or intentional science." That's not to say that these new sciences are not physical. It's not as if we're going to throw the physical out and go off into some abstract realm. The physical substrate is an absolutely critical part of meaning things, as all the discourse about materiality and the body, and so on and so forth, is so quick to emphasize. In fact materiality, in literary disciplines, is a very trendy thing.

—— *At this point there was a bit of a digression* ——

BCS • (continuing): This is a footnote, but the idea that the internet is "virtual" is crazy. Where did this idea come from? It's as material as anything; it just happens to have a different salient physics. It's a different materiality than lots of our other materiality, but it sure is material. As AOL knows only too well.

GS • I did look up John Searle in the index to "The Origin of Objects" and found you pointed this criticism at his claim that software is not material.

BCS • Yep; for sure.

—— *End of digression* ——

BCS • (continuing): Anyway, get back to what we were saying. So what is this new realm of science? Well one way to describe it is as an emerging science that deals with signs and signifying. Signs, signifying, signification—as long as you understand those words broadly enough, these things are the essential basis of anything semiotic or epistemic or intentional. I think we're on the cusp of a new era of this kind of science.

A minute ago I mentioned the alchemists. I mentioned them because I think of the world's C++ programmers as essentially *semiotic alchemists*. The original alchemists were trying to turn iron into gold; today's alchemists are trying to turn C++ code into gold. By now we have perhaps fifty years of a very widespread, inarticulate, absolutely dedicated, and rather disheveled practice of people trying to construct arbitrary things out of symbols and information. It really is a very similar situation. And I wouldn't be too surprised, once we finally getting our heads around this new stuff and understand it, if this first hundred years of inchoate programmers get laughed at and shunned, and are thought to be just all messing around, the way we shunned and laughed at the alchemists. But I bet, too, that present-day programmers are in fact, and will ultimately be recognized to be, as important as the alchemists were, in setting the stage for a profound new intellectual revolution.

So what does this have to do with religion? Here's the crunch. Signs, signifying, signification, and…*significance*! But as we saw at the beginning, significance means *importance*. What's *significant* isn't just what has been mentioned or symbolized or represented or referred to, but what *matters*.

That brings me to the million dollar question. If twentieth century developments—computing and logic and psychology and mathematics and theoretical biology and so forth—is really bringing us to the verge of a new era in science, a new era that will take on not just the physical world, but also the world of symbols and meanings and signifying, *what will this new era have to say about significance?* Is the kind of significance it will be able to study restricted to a mere truth-like semantic relation, of one thing (like smoke) signifying another thing (like fire)? Or is there a chance, when all is said and done, that we won't be able to take on significance for real without recognizing that it means importance, too? In other words: is this new era of science going to require a broadening of our sights to include not just the factual, but also the ethical?

GS · Or, I guess, could you subtract the value of significance in the scientific study of it? I suppose the last hundred years of anthropology has faced that puzzle.

BCS · That's right. In fact it's doubly true! It comes up at the meta level. You could imagine an ideological traditionalist who, wondering how to study signs, would ask the question this way: is there a right—i.e., *true*—way to study signifying? And it also comes up at the object (base) level: is *truth* the only substantive connection that connects signifying acts to the world? But I'm not prepared—especially a priori, in a prejudicial way—to restrict myself to truth alone at either level. Of course this is counter to some trends. Even truth has come under fire in lots of postmodern contexts, so that people start talking about "endless plays of signifiers, signifying nothing." It has actually proved very difficult to hang on to truth, in the face of things like cultural pluralism. Anthropologists certainly know this. How are they to assess the truth of the myths? Maybe they can think that they're not going assess the truth of the myths at the object level, but what about at the meta level—what about the stories they publish in anthropology journals about these myths? Are they meant to be true stories about myths that don't have any truth properties? Or are they just more myths, that they're spinning in the anthropology community?

So there's leakage. That's part of our present-day intellectual crisis. But I want to keep the main topic in focus. If we admit signs and signification into the realm of science, what is the full range of normative (value) consequence? Of course some people would say that this is all a pun—that it is only an etymological accident that "significance," in English, means "importance," and is also used (more technically) to refer to the property of signs, whereby the signify things. But I think that's *false*. From what I can tell from having studied intentional systems, the *truth* property and the property of *normative consequence* cannot, in fact, be wholly separated. So it is not an a priori argument on my part. What I am saying is that broadening the scope of applicable norms, at both the level of the theory and the level of the subject matter, is a *necessary* condition of this new scientific era.

In fact once you realize this, all sorts of things start making sense, on both sides of the fence (physical and intentional). For it is not just that *signifying* involves an ethical dimension. That same is true of materiality. *Material* evidence, in a court of law, isn't evidence that weighs some number of kilos, or that has an inertial mass, but evidence that *makes a difference*. Even the word 'matter' has a normative dimension, in English. Scientifically, we think of matter as "pure physical stuff." But what "matters" is also a way of describing what is important. (I bet if you looked back over a transcript of this conversation, the word 'matter' will have occurred half a dozen times already.)

You might think that this, too, is a pun. But again, I believe that is wrong. In fact one of the things I try to do in my book is to reclaim "materiality" for the kind of thing that has importance, and pull it away from pure physicality. This is because—and in this sense I'm not far from various continental traditions, and an increasing number of people in analytic philosophy—I believe that ordinary material objects are normatively constituted. To be an object is to be taken by agent or society to be something that is *valued* as an object, something that one has to *defend* as an object. I.e., to say that "A cup is a cup" is a normative statement; a statement of object identity is a statement of values, not a statement of purely physical conditions. Does that mean I want to say a cup is not purely a material object? No, what I want to say is that it *is* a material object, but that *materiality* is *normative*. So in a funny way I end up being more materialist than most people (certainly most religious people) would expect.

GS • Re-imbue matter with mattering?

BCS • Yep, re-imbue matter with importance. Put the mattering back into the matter. That's right. And then…this is the dream…maybe we can have an epistemic or intentional or semiotic "science" that actually understands "significance" in the ethical sense of importance. And a science that does so in a *good*, not just in a *true*, way!

Now I have to be careful here. Dreams can crash and burn. I don't really want to prejudge all of this. I don't want to say I have an a priori commitment to a claim that importance does in fact derive from signification, in the way that this new era of science is going to understand. Two or three hundred years from now, I can imagine, even if we have a kind of semiotic science or a broad range of sciences dealing with signification and interpretation and so on, that people of that day will say "Look, issues of mattering, issues of emotion, issues of social justice, etc., weren't more done any more justice by the 300 years of intentional sciences than they were done justice by the preceding 300 years of physical science.

But—and this is the point—I am not sure that that's right. I am not sure that it won't be the other way. That is: I want to be open-minded to the possibility that we do, in fact, need to take on importance, significance, to serious ethical considerations. In fact I have reasons for thinking so. See, one of my most basic metaphysical commitments is that, *au fond*, truth, beauty and goodness are not completely separable. Just as the physicists claim that gravity, charge, mass, etc., weren't separate, in the first 10^{-23} seconds of the universe, so too I don't think God made the world with truth, beauty, and goodness fully separated out, either. In fact I think the idea that they are distinct is rather our idea (and not necessarily the greatest idea, at that). Strangely enough, I even think you can see shadows of this being true in modern software design. Whether programs "work well", whether they're beautiful, and whether they're right—in practice these things aren't all that separable. In practice, that is, it is impossible to maintain a clean distinction between and among those norms.

GS • There are certainly good psychological explanations for why we would associate beauty and truth especially. Truth and goodness, I'm not sure I can explain either psychologically why such an association would evolve, since there are as many true things that can do harm as there are true things that can do favor.

BCS · Well, it depends on what favor means. It's certainly good for you to have a roughly realistic sense of what's going on. To live in a fantasy life with respect to your visual perception would not be a very good strategy, in heavy traffic. You'd quickly get killed. If good has to do with survival—and I bet you're running into this idea with your biologists—then evolution can be used as an anchor to tie the good and the true. This is actually a rather popular idea just now: a lot of people think that what is valuable about both biological and psychological states is that they lead to survival. As it happens, I am quite unhappy with subjugating truth to survival, because I can easily imagine situations where mass delusion would prolong survival. I.e., it is sometimes more advantageous not to understand what's going on. But here we're basically getting into the real philosophy. All I'm trying to do in this conversation is to open up the possibility of these questions. Actually that's not quite right; I'm trying to do something more. I'm trying to say that ethics may not only have to be brought into our new subject matter; it may also have to be brought into our methods. Not just *true* theories of the-true-and-the-good. True-*and-good* theories of the-true-and-the-good.

GS · The philosopher Ludwig Wittgenstein said something about how we feel that when science has answered everything it can, the questions of life will remain untouched. You seem to be suggesting that with the emerging science of semiotics, that what religion is like to you, or what you're meaning by religion in this conversation, and science may well actually meld together, and that science may begin to say some things that do touch the "questions of life." On the other hand, I see that you're remaining open-minded about that, you're not convinced necessarily that that's so, but you're opening up that possibility.

BCS · It depends on how we use the words.

There are several problems. First, I don't have a good word for this new era. Nor is it up to me, as a solitary individual, to prescribe a word. So terminology is hard. It is almost guaranteed that any expression I use—"science of semiotics," "science of intentionality," or whatever—will mean something to most people that is *not* what I intend. (Earlier I said that it was hard to know what to say to people about whether I was religious; this current topic is no easier.) So it is very difficult to know how to put this.

But I can at least say this much. *Science* is not going to shed light on these (ethical or transcendent) questions, if by "science" we mean what science has been imagined to have been, for the last 300 years. A whole panoply of assumptions underlie our present image of science, some of which we've already mentioned: about its having no values (other than truth), about its objectivity, about its formulation of the laws of nature, about certain notions of reduction, and so on. *That* conception of science is not going to touch the "questions" of life. That's what Wittgenstein said, and I agree with him.

The thing is, I don't think that conception of science is going to work to understand the era of significance, either. And so I am reluctant to say, "No, science can't touch what matters. You have to look elsewhere." Statements like that are rooted in a particular conception of science—the one we've had for 300 years—which may not last. For at least three reasons. First, if I am right that a new metatheoretic framework is going to be needed, in order to understand this new "Age of Significance"—that's really my name for it, by the way—then maybe science will simply change, to incorporate these new values. Stranger things have happened. Second, as I said above, I believe our current *conception* of science is inadequate to the task, but then our current conception of science is inadequate to explain *current* science, too—as so many people in the history of science, science studies, philosophy of science, etc., have documented in the last few decades. Even what it is that is currently known, scientifically, and *how it is that it is known*, are more politically and ethically infused than it is usually recognized in the reigning myths. So in a way the sorts of change I am envisaging may as much involve a deepening of our understanding of what's has always been the case, as they do require a brand-new conception. And third, there are some signs that things are already afoot that are going to transform science as it is into something new.

In sum, for a whole lot of reasons, I am not sure that what we call "science" is all that stable. And so it may change enough to include other norms and other values, in its methods and its subject matters, in ways that could start to incorporate Wittgenstein's "questions of life." After all, the root Latin word, "*scio*," just means "to know." It doesn't intrinsically mean a certain *kind* of knowing, only appropriate for the sorts of physical phenomena that science has classically studied. So it may be flexible enough to incorporate issues of interpretation and meaning in a truly meaningful way. I don't know, though. I think it is too early to call.

You know you can equally well ask the same question of religion. Will religion be flexible enough to incorporate what we learn about symbols, interpretation, meaning, significance? Just as science may change, so too religion might change—into something unlike anything ever imagined. Maybe, as much as a new science, we need a "new theology": unlike any religious traditions we've ever had, altered so as to capture the imaginations and inspire a world-wide community of diverse people, and brought up to date—so as to incorporate the full range human questioning into questions of ultimate significance, able to give people a reason to live and an anchor for their commitments, able to help people understand why they care about the people they care about, why they should care about things that are important—maybe we need a new theology like that, as much as we need a new science. Or maybe they are the same thing. Who knows? I don't know what will fire the imagination, calm the spirit, do justice to the world, and provide grounding for our lives. All I know is that it is urgent that we do our best to start figuring these things out.

I try to take a small step in this direction in my "Objects" book, sketching a metaphysical conception of the world that, I think, might be durable enough to underwrite both projects (or their fusion, or whatever). The basic claim is that no *other* form of metaphysical foundation (and no foundation we've had in the past) is strong enough even to underwrite science and computing and things that mundane, let alone questions of importance and ultimate significance. As I said above, even simple questions of individuation, such as what thing an individual entity is, can't be answered, I believe, except with respect to an ethical frame—which already starts to encroach on topics of traditional religious interest. Having to decide if a fetus is alive, in the case of abortion, for example, is a question of individuals—and of course it is a question that matters. If you are going to act, based on your answer, you need to know what your commitments are. I don't believe any science can answer whether there's a person there, without recognizing that it is an ethical—perhaps even sacred—question. I.e., there's not going to be a non-ethical science that can do justice to the requisite notion of individual. And basically, I don't think there is any other notion of individual. That's the only notion of an individual there is.

So I don't know whether we should call it science, or religion, or philosophy, or metaphysics. I guess I don't even really care (except of course that what you do call it has enormous political ramifications). I'm more interested in what it's going to be *like*.

GS • Is it possible that a computer scientist, in trying to develop a machine that could recognize individuals, would provide the answer to the question of what constitutes and individual?

BCS • I think it's unlikely. First of all, I don't think that's a question that has a black and white answer. If it has an answer at all, it's not in any ordinary sense of "answer." What I think is true, is that if computer scientists write programs which make decisions based on judgments of individuality, and those systems are deployed in society, then those systems *are thereby intrinsically implicated in questions with that kind of ethical weight*. The question is, what responsibility do you bear as a programmer, or as a computer scientist, in constructing systems that make that kind of decision?

I've been talking philosophy, all this while—but the questions aren't always that philosophical. So for a change of pace, let me come at some of these issues from a very different (and much more pragmatic) perspective. About fifteen years ago, a bunch of us were involved in starting an organization called Computer Professionals for Social Responsibility (CPSR). We were concerned about a lot of things, but what focused the organization at the beginning were questions about nuclear war: Reagan's Star Wars Initiative, and issues about launch and warning. It wasn't easy to figure for sure what was the case, but we worried that a lot of the Pershing II Missiles in Eastern Europe were set on automated (i.e., computer-based) launch and warning status, since you basically have eight minutes from the launch on the Soviet side to get those missiles out of the ground. Our question was: can you trust a computer system to make the right decision in eight minutes? That is, we wanted to get the right question on the table: are you prepared to threaten civilization as we know it, in that kind of time frame? It wasn't an abstract question of whether computers could or could not be trusted. People would ask, If you don't trust the computer would you rather have a person do it? Our answer was: *no*; neither a person nor a machine should do it; it is not a question that should be answered in eight minutes, *at all*. It just shouldn't be done. It's not the kind of judgment that can be made in that amount of time. Why? Because it's a sort of judgment that has such profound consequences. Anyway, this was basically our line.

Very soon, we encountered left-wing fundamentalists, who said, "You should never trust a computer with human life." But I don't believe that. I land at the San Francisco Airport in the fog all the time. I'm glad there aren't pilots peering out the windows trying to find the runway. In fact I think that being landed automatically by radar, or at least substantially assisted by radar, is quite possibly far and away the best thing to do in that situation. But if that's true, then you have to face up to the question: "What *can* you trust computer systems with?" Very quickly, that brings you up against questions of what it is to trust, what kinds of decisions there are, how we can understand issues of that sort, and so on. Talk about biological taxonomization being hard! Taxonomizing the ethical structure of the sorts of decisions that computers are implicated in is terrifically difficult. The thing is, though—and this is the point—it is something that we are tacitly doing already. *All the time.* We are doing it because computers are already deployed, throughout society, often in so-called "mission-critical" applications.

All I want is for our imaginations, and our understandings, and our insight, to be up to these decisions that society is inevitably taking. I don't think we're going to stumble on the right answer by fortuitous accident. And I am concerned that computer science is intrinsically implicated in the answer. And if computer science (which I'm part of) is implicated in the answer, then I think we damn well better figure out what we're doing.

Computers you know, are rather diabolical things. Although they were originally invented by a mathematician, they aren't theoretical objects any longer. They are actual; they are participants, here in the world, along with us. They have material properties. They have economic properties. They affect political decisions. They are implicated in ethical decisions. And so on and so forth. They are wonderfully historically ironic, in fact—in the sense that they're implicated in all kinds of issues that transcend anything frameable in the theoretical frameworks of the people who invented them. So our responsibility, as computer scientists and philosophers and social theorists and the like, is to come up with an *understanding* of computers that is up to the challenge that they intrinsically pose.

GS • Technological progress depends a lot on looking at things in new ways, in honoring innovation, and in trying on different pairs of glasses, so-to-speak, until you've seen things in a light that enables you to do new things. A lot of religion as it's practiced, has a reverse emphasis. It emphasizes the importance of seeing things in a traditional way, of reminding oneself how things are to be understood, of reminding oneself why certain things are good and other things bad. How do you move back and forth between this striving for new interpretations, and at the same time honoring the past and the significance that we obviously inherit from it?

BCS • Well, as I'm afraid you will predict, I take exception to the question. I think it's false on both fronts. Sure enough, science is supposed to look at things in new ways, but there's a tremendously conservative structure underlying how you are supposed to look. You're supposed to have causal explanations of a certain sort. You have to have P be less than .05. You have to know whether a thing has been experimentally verified or not. The canonization of the scientific ways of looking at things is pretty strong. Within that, of course you're looking for new things. But again, although you are looking for things in new ways, what you're looking *at* is not supposed to change. In fact that is encoded in the famous scientific "empirical method": the basic assumption that the world is out there, independent of what you're doing. It has presumptively been there forever, that kind of stuff. Science as we know it, that is, presumes a kind of absolute "givenness" to the structure of the world. The world of science is not our *creation*. And so on and so forth. There is a tremendously canonized conceptual structure to science, in terms of what you understand, and what you are supposed to do to understand better.

Also, note that it is only the research scientists—quite a small segment of society, if you think about it—who are supposed to be doing this novel stuff. Mostly—in its application to build bridges and develop new drugs—the science itself is supposed to hold pretty stable.

On the other hand, it is my impression that anyone who is serious about the religious traditions has recognized that religion, too, can get old and encrusted. The history of the religious tradition is full of fights against the evils of stagnation and unimaginative bureaucratization. Similarly, consider interpreters of the Talmud, speculative theologians, mystics and religious writers. There are a tremendous number of religious traditions that emphasize the constant renewal and reinterpretation that is required in order to keep a tradition vital.

It is too bad, I think, that in discussions of science versus religion, people often select a Nobel Prize-winning physicist from Bell Labs, and then contrast their sense of science with a layperson's belief in the catechism or reincarnation or something like that. If we are to have Nobel scientists representing science, we should have great theologians, and ask whether the great theologians aren't looking for new ways as much as the new scientists are. Or else ask whether people in the street who have put their fingers in the outlet if they are trying to invent new ways to understand electricity. By and large they're not. If it's a hundred and twenty volts, it's going to hurt you. If it's twelve volts it won't.

Once you've got the thing at the same level on both sides (as I hope we will at these conferences), then there's no reason, or at least there is less reason, to suppose that there should be any less room for increasing and deepening and opening oneself to new ways of understanding on the religious side than on the science side. By chance, I just saw the film "Open City," made during the war in Italy (partly by Fellini), in which a priest collaborates with a profoundly good but otherwise non-religious fellow, in protecting various people against the German occupiers. At one point another priest challenges the first priest, asking him how he can collaborate with a non-believer. And of course the first priest says the evident platitude: that the alleged "non-believer" is seeking the truth and doing good, and that, as far as he knows, that's what it is to be Christian. Surely any Christian worth their salt is going to recognize the truth in that.

GS • Yeah, on the other hand, it does seem to me that in general, the theological perspective, even among those theologians worth their salt, seems to invest a lot of hope in something that has already happened, and in taking to heart lessons already spelled out in the past. And, in general, it seems that science, especially in the culture of technology, looks for salvation in a future. It does seem that science looks forward in some sense, and religion backward, for its inspiration, if not for its power.

BCS • Sure, institutionally there is truth to that. Certainly the myth of scientific research is this constant emphasis on the "new, new, new." And admittedly, too, the religious myths don't have this "ever new" emphasis. But some of them nevertheless emphasize searching—though it is more of the personal variety. Many years ago I was married to a Quaker, for example, and for a while attended Quaker meetings. You know George Fox's notion—that there is "that of God in every person," with the concomitant rejection of the priesthood and so on—that each person's salvation is for him or her to find. So in this sense the notion of searching is as religious as it is scientific.

On the other hand, you are surely right that searching is not as heavily institutionalized on the religious side as in science. But that doesn't mean that that is okay. So much the worse for theology, in fact, I would say! Surely it has to change too, to come to understand better as urgently as science does.

Look, it's not that I think scientific and religious practice are (or even should be) identical, that there no distinctions in the world, that everything should be reduced to one grand "Om." But it strikes me as tragic, if it is true, as you suggest, that the religious traditions aren't out there trying to figure out new things. Think of the urgent problems they face. How can they simultaneously have faith in their own traditions, and yet recognize the validity of other religious traditions? Can they help the rest of society develop a way to incorporate the generosity and justice of pluralism without compromising excellence, standards, and value? I.e., how can we have a pluralist world view that is neither vacuous nor shallow? Presumably it is too late, in history, for any religious leader any longer to say (or believe) anything of the form: "we're right; and you're wrong." And yet, at the same time, it would be terrible if religious leaders were to water down conviction to something like "It doesn't matter what you believe; we all have our stories." Both of those positions—both of those limit cases—are profoundly untenable. But what is a viable middle ground? Or is it even a question of "middle"?

Formulating it this way, moreover, shows how intertwined the issues are with intellectual and scientific ones. For there is no greater problem facing the university, I believe, than essentially the same one: how to combine appropriate respect for pluralism with deep recognition of value.

GS • That's a key issue for a lot of the scientists in this project. They have these two very powerful ways of gaining access to the world, but what do they say about each other? How do they coexist?

BCS • I think that's absolutely right. What I'm saying is that, as well as being an issue *between* science and religion, it is also an issue *internal* to science itself, and also internal to religion itself. What are the Christians and the Muslims and the Zionists going to say to each other, for example? All of us have Abraham in our background. It's not as if we're as distant, culturally, as each of us is to Buddhist or Hindu traditions.

Admittedly, the problem may not be as acute for *individuals*. Few of us, individually, belong to more than one religious tradition; and few of us, too, practice more than one science. On the other hand, quite a few of us are scientists and also have, in one way or the other, religious sensibilities. So it may be that issues of pluralism arise more acutely for individual people across the science-religion boundary, rather than within either side. Still, it is important to recognize that the issue itself—the issue we are dealing with at this conference—is not unique to our setting. It is one of the day's great questions—a perfect example of a sort of questions that the two traditions could collaborate on *in general*.

I even wonder whether it might not be a more profitable topic—if only because it would deepen the collaborative sense of "we." You and I are sitting here, at the moment, having this interview, looking at each other. But if we were to sit side by side, and look out on those mountains over there, and talk about whether California is going to fall in the ocean, or whether the coastal commission is doing an adequate job protecting these hills, our sense of having in common something, something larger than us individually, would constitute a bond. Perhaps CTNS could someday have a conference on how to combine a sense of norms or standards with an adequate sense of pluralism, and people could speak to that common problem from both the science side and the religious side.

GS • Let me change direction for a second and talk about God. Does the idea of God work into your view of life at all? I know it's a word that you use occasionally. There was a quote in your book that I thought was quite lovely. You write that "the world has no other." Unless the world itself is defined as God—a definition that might wear out pretty quickly for its simplicity—is there any room in this perspective for God?

BCS • I heard it said, once, that one of the most politically shrewd ideas of Christianity was the construction of the trinity. The idea was that many people—pastors and parishioners both—had a great deal of trouble with one of the three, but most felt comfortable enough with the other two, leaving them with a majority. I remember asking some ministers about Jesus, God, and the Holy Spirit; some of them said they just couldn't figure out about the Holy Spirit, they were just kind of put off by that, but that God and Jesus were fine. Other people, other pairs.

GS • If you answered yes to at least two of the three above questions you belong.

BCS • Something like that. And sure enough, my reactions are asymmetrical. I am very resistant to the notion of Jesus. I admit it; I get quite put off. But God and the Holy Spirit don't trigger that kind of allergy. That is not to say that either notion figures in either my language or my thoughts—internal or public. But I feel as if I know what the tradition I come from was getting at, with those notions. And that I feel appreciative of.

Was it Tillich who said God was the ground of being? To the extent that I have any use for the word "God"—or perhaps what I mean is that to the extent that I *understand* the word "God," since I don't really use it—it is as a word for everything. For me, it is a reminder—it connotes the *moreness* of everything. I'm not sure, but I think it is part of the muezzin's cry to say something like "I know that Allah is greater than I know him to be." There's a wonderful humility implicit in that phrase. So to the extent that the word "God" means anything to me, it absolutely does *not* mean anything like a person or anthropomorphized figure. It doesn't mean anything that has agency in the world, that is separate from the world in any way. There are Kabalistic stories, I understand, about how at the beginning of the universe God had to evacuate a space within himself in order to make room for the world to exist. That's certainly wonderful poetry, and it makes a wonderful point, but I don't believe it. I suspect my notions are much closer to Buddhist notions than anything recognizably Christian—except that I don't know how rare it is in Christian theology to take God to mean something like the "ground of being." "The world" is kind of a cheap way to refer to everything there is. "God" is an expensive way to refer to everything there is (and many people are allergic to it). So I don't know.

GS • Clearly you don't use that definition, since you say in your book that "there's nothing larger than the world."

BCS • Well, again it's just this problem of communication. If you have a people who have a roughly common sense of the totality then it's useful to have a word that doesn't name the totality, because names don't work that way. Names require a figure/ground separation, this is not going to be a figure because there's no ground. But if you have a kind of shorthand way of orienting towards everything, then in fact maybe the word "God" is a good word. But in 1997, in post-industrialized western U.S., using "God" as a word to allow people to remind themselves to orient in total probably doesn't work very well. I don't know that we have any *other* word that does work in toto. And that, I think, is what is urgent. I'm not really interested in whether I believe in God. I probably don't, in the sense that I don't assent to the proposition that most people would think those words express. But what matters to me is not the future of that word, selling it short or buying it long. What matters to me, throughout all of this, is *what terms those people who don't find religious vocabulary serviceable are going to use to mean such things*. What words are going to carry that kind of meaning for us? How are we going to speak? how are we going to talk to our friends about what matters to us?—if we've rejected that dimension of our cultural heritage which has propped up that ultimate question? It is pretty undeniable that the religious traditions have been the locus where most ultimate questions get framed, for most civilizations, over most of their histories.

GS • And where significance is derived, too.

BCS • Ahh, yes—but it's tricky. Whether significance has been *derived* from there, or whether it's just that the religious side of the house is where significance has been *recognized* and *affirmed*, isn't so simple a question. But I think it's more the latter. That is, it seems to me closer to the tradition not to say that you derive your significance from church, but that going to church *reminds you* of your significance.

GS • But they might say that you derived your significance from God.

BCS • Well they might. But then the question is, What is God, such that you derive your significance from Him? And on that, people vary. Some people of course are reputed to think of God as a delineated individual, of a sort that is different from trees. But I just don't understand that. This goes back to your earlier suggestion that science searches for new ways of understanding, whereas the religious traditions don't. It seems to me urgent for the religious traditions to recognize that the word 'God' isn't doing much, these days—not only for people outside the religious communities, but even for people within the religious communities, if it is taken to mean something separate. I doubt that they'd want me as a theologian, but that is what I'd be tempted to argue. That the idea of a "separated" God just doesn't make sense, in the context of our twentieth-century understandings of the world. In fact it seems to me dangerous. To license it—without some pretty fancy concomitant explanation—is liable to engender a sense that religious understanding can part company with other (e.g., scientific) understanding and not be responsible for showing how that can be so. That is, it is in danger of not taking responsibility for showing *that the world is one*. And that just seems to me shabby. Showing that the world is one is exactly the kind of ultimate question that religious traditions should be focused on.

That's a great question: what could a conception be, what could a practice be, that would enable people to orient towards the grounds of ultimate significance in a way that's modern? If theologians are not thinking about that, they sure ought to be. That's certainly what I am trying to do in *Objects*, but it is of course one person's paltry start. And words are a problem. We can't solve this thing alone.

GS • If to be religious is "to find the world significant," God might be defined as that which makes the world significant. But there may not be that much you can say beyond even that.

BCS • That's not too far from Tillich's conception of the ground of significance. But I confess to having trouble with the way you put it (that God *makes* the world significant): it sounds *causal*, as if God is the cause, and the world's being significant is the effect. I.e., as if God made the world significant the way GM makes Chevrolets. People like thinking that way; they are happy with cause and effect; cause and effect seem to be part of the great science we all inherited. But I don't like it because it makes God and the world two. And the minute you have *two*, I don't think you are in the realm of God any more. So if I were to say anything (not all that likely), I would say God is more like *the world in all of its significance*, or something like that.

Moreover, it is my sense that most religious traditions, if you push, don't say that "something makes things significant," but rather that things *are* significant in virtue of their existence. Significant in and of themselves. If that's not Christian, then I guess I'm not a Christian; it's not for me to say what that tradition is. Though I do think that any attempt to formulate what the word "God" means that tries to specify it in articulated terms is going to fail. What's most important, if we are going to keep that three-letter word around at all, is surely not something *articulated*. If people could have a sense of what it is to live life in such a way as to take the significance of the world seriously, and find significance for themselves therein, then I think a practice could grow up in which people used the word to remind each other of that common orientation. But it is the *orientation* that matters; not the *formulation*.

I suppose all I'm saying here is that no one thinks that religious language is enough to make anyone religious. (That's one reason why religious and non-religious people don't share enough language for there to be a sentence they can both entertain, such that one agrees with it and the other disagrees.)

Language is a very big problem. When I first moved to California in 1981, I looked at a bunch of churches. I was put off, though, by prevalent tendencies for the services to orient towards personal psychology and social justice. I felt that I could get better personal psychology from psychiatrists, and better social justice from political action groups and political science. So I didn't go back. The two traditions that had the most power (though I didn't take either of them up, either) were Quaker and High Episcopal. Some people found that odd, because Quakers and Episcopalians are often thought to be at the opposite ends of the Protestant spectrum. But they had one crucial thing in common: they didn't try to translate religious language into propositional form. Quakers, of course, did this by not putting weight on formulation at all (they're silent). And the Episcopalians were okay as well, because it turns out that the 1929 Book of Common Prayer, which they use, is so ritualized, and so poetic, that in point of fact it is capable of much more radical theological interpretation than the supposedly more liberal mainstream churches. So except for these two I was disappointed; the attempts to modernize had ended up being radically restricting, because they tried to formulate particular, concrete, modern interpretations of things that I thing aren't so effable. I think that's a mistake.

On the other hand I have great respect for how hard it is to say any of these things in a way that *is* tenable.

Poetry is some help. A poem can orient you towards things that it itself doesn't have to name. Plus, people understand that even though a poem is not factual, it's also not thereby false. There's a lot of that, I think, in traditional religious language. But at the same time poetry is too marginalized, right now, to play as important a role as we need. I don't think, given the scientific, technological, economic, and political state of the industrial west, that poetic language alone is going to allow people to forge a requisitely strong common sense of purpose, and adequately give voice to the things that matter to us, individually and collectively.

So what language *will* work? I tell you: I don't know. This is an absolutely urgent question, without evident answer. One thing I know: we can't presume that we know how language works, and then, using that presumptive understanding, try to figure a language that will articulate our sense of significance. Current theories of language are too rooted in the prior scientific (formalist) era. But language—fortunately, language is not hemmed in by what we think of it. It's fertile, fecund, and not, I think, exhausted. So I'm still optimistic. Maybe we can find—even hammer out—some language that will go the distance.

GS • That might be a good place for us to stop. Thank you.