

HOW WORDS DO THINGS WITH US

Language, like consciousness, only arises from the need, the necessity, of intercourse with others.

KARL MARX, 1846

Consciousness generally has only been developed under the pressure of the necessity for communication.

FRIEDRICH NIETZSCHE, 1882

Before my teacher came to me, I did not know that I am. I lived in a world that was a no-world. I cannot hope to describe adequately that unconscious, yet conscious time of nothingness. . . . Since I had no power of thought, I did not compare one mental state with another.

HELEN KELLER, 1908

1. REVIEW: E PLURIBUS UNUM?

In chapter 5 we exposed the persistently seductive bad idea of the Cartesian Theater, where a sound-and-light show is presented to a solitary but powerful audience, the Ego or Central Executive. Even though we've seen for ourselves the incoherence of this idea, and identified an alternative, the Multiple Drafts model, the Cartesian Theater will continue to haunt us until we have anchored our alternative firmly to the bedrock of empirical science. That task was begun in chapter 6, and in chapter 7 we made further progress. We returned, literally, to first principles: the principles of evolution that guided a speculative narration of the gradual process of design development that has created our kind of consciousness. This let us glimpse the machinery of consciousness from inside the black box — from backstage, one might say,

in homage to the tempting theatrical image we are trying to overthrow.

In our brains there is a cobbled-together collection of specialist brain circuits, which, thanks to a family of habits inculcated partly by culture and partly by individual self-exploration, conspire together to produce a more or less orderly, more or less effective, more or less well-designed virtual machine, the *Joycean machine*. By yoking these independently evolved specialist organs together in common cause, and thereby giving their union vastly enhanced powers, this virtual machine, this software of the brain, performs a sort of internal political miracle: It creates a *virtual captain* of the crew, without elevating any one of them to long-term dictatorial power. Who's in charge? First one coalition and then another, shifting in ways that are not chaotic thanks to good meta-habits that tend to entrain coherent, purposeful sequences rather than an interminable helter-skelter power grab.

The resulting executive wisdom is just one of the powers traditionally assigned to the Self, but it is an important one. William James paid tribute to it when he lampooned the idea of the Pontifical Neuron somewhere in the brain. We know that the job description for such a Boss subsystem in the brain is incoherent, but we also know that those control responsibilities and decisions have to be parcelled out somehow in the brain. We are not like drifting ships with brawling crews: we do quite well not just staying clear of shoals and other dangers, but planning campaigns, correcting tactical errors, recognizing subtle harbingers of opportunity, and controlling huge projects that unfold over months or years. In the next few chapters we will look more closely at the architecture of this virtual machine, in order to provide some support — not proof — for the hypothesis that it could indeed perform these executive functions and others. Before we do that, however, we must expose and neutralize another source of mystification: the illusion of the Central Meaner.

One of the chief tasks of the imaginary Boss is controlling communication with the outside world. As we saw in chapter 4, the idealization that makes heterophenomenology possible assumes that there is someone home doing the talking, an Author of Record, a Meaner of all the meanings. When we go to interpret a loquacious body's vocal sounds, we don't suppose they are just random yawps, or words drawn out of a hat by a gaggle of behind-the-scenes partygoers, but the acts of a single agent, the (one and only) person whose body is making the sounds. If we choose to interpret at all, we have no choice but to posit a person whose communicative acts we are interpreting. This is not quite equivalent to positing an inner system that is the Boss of the body.

the Puppeteer controlling the puppet, but that is the image that naturally takes hold of us. This internal Boss, it is tempting to suppose, is rather like the president of the United States, who may direct a press secretary or other subordinates to issue the actual press releases, but when they speak, they speak on his behalf, they execute his speech acts, for which he is responsible, and of which he is, officially, the author.

There is not in fact any such chain of command in the brain governing speech production (or writing, for that matter). Part of the task of dismantling the Cartesian Theater is finding a more realistic account of the actual source(s) of the assertions, questions, and other speech acts we naturally attribute to the (one) person whose body is doing the uttering. We need to see what happens to the enabling myth of heterophenomenology when the complexities of language production are given their due.

We have already seen a shadow cast by this problem. In chapter 4, we imagined Shakey the robot to have a rudimentary capacity to converse, or at least to emit words under various circumstances. We supposed that Shakey could be designed to "tell us" how it discriminated the boxes from the pyramids. Shakey might say "I scan each 10,000-digit-long sequence . . .," or "I find the light-dark boundaries and make a line drawing . . .," or "I don't know; some things just look boxy. . . ." Each of these different "reports" issued from a different level of access that the "report"-making machinery might have to the inner workings of the box-identifying machinery, but we didn't go into the details of how the various internal machine states would be hooked up to the printouts they caused. This was a deliberately simpleminded model of actual language production, useful only for making a very abstract thought-experimental point: if a sentence-emitting system had only limited access to its internal states, and a limited vocabulary with which to compose its sentences, its "reports" might be interpretable as true only if we impose on them a somewhat metaphorical reading. Shakey's "images" provided an example of how something that really wasn't an image at all could be the very thing one was actually talking about under the guise of an image.

It is one thing to open up an abstract possibility; it is another to show that this possibility has a realistic version that applies to us. What Shakey did wasn't real reporting, real saying. For all we could see, Shakey's imagined verbalization would be the sort of tricked-up, "canned" language that programmers build into user-friendly software. You go to format a diskette and your computer "asks" you a friendly question: "Are you sure you want to do this? It will erase everything

on the disk! Answer Y or N." It would be a very naïve user who thought the computer actually meant to be so solicitous.

Let me put some words in the mouth of a critic. Since this particular imaginary critic will dog our discussions and investigations in later chapters, I will give him a name. Otto speaks:

It was a cheap trick to call Shakey "he" rather than "it"; the trouble with Shakey is that it has no real insides like ours; there is nothing it is like to be it. Even if the machinery that took input from its TV camera "eye" and turned that input into box-identification had been strongly analogous to the machinery in our visual systems (and it wasn't), and even if the machinery that controlled its production of strings of English words had been strongly analogous to the machinery in our speech systems that controls the production of strings of English words (and it wasn't), there would still have been something missing: the Middleman in each of us whose judgments get expressed when we tell how it is with us. The problem with Shakey is that its input and output are attached to each other in the wrong way — a way that eliminates the observer (experiencer, enjoyer) that has to lie somewhere between the visual input and the verbal output, so that there is someone in there to mean Shakey's words when they are "spoken."

When I speak, [Otto goes on] I mean what I say. My conscious life is private, but I can choose to divulge certain aspects of it to you. I can decide to tell you various things about my current or past experience. When I do this, I formulate sentences that I carefully tailor to the material I wish to report on. I can go back and forth between the experience and the candidate report, checking the words against the experience to make sure I have found *les mots justes*. Does this wine have a hint of grapefruit in its flavor, or does it seem to me more reminiscent of berries? Would it be more apt to say the higher tone sounded louder, or is it really just that it seems clearer or better focused? I attend to my particular conscious experience and arrive at a judgment about which words would do the most justice to its character. When I am satisfied that I have framed an accurate report, I express it. From my introspective report, you can come to know about some feature of my conscious experience.

As heterophenomenologists, we need to divide this text into two parts. We put to one side the claims about how the experience of speaking seems to Otto. These are inviolable; that is how the experience

seems to Otto, and we must take that as a datum demanding an explanation. To the other side we put the theoretical claims (are they the conclusions of tacit arguments?) that Otto makes about what this shows about what is going on in him — and how it differs from what was going on in Shakey, for instance. These have no special standing, but we will treat them with the respect due all thoughtful claims.

It is all very well for me to insist that the Middleman, the Internal Observer in the Cartesian Theater, must be eliminated, not found, but we can't just throw him away. If there isn't a Central Meaner, where does the meaning come from? We must replace him with a plausible account of how a meant utterance — a real report, without any scare-quotes — could get composed without needing the imprimatur of a solitary Central Meaner. That is the main task of this chapter.

2. BUREAUCRACY VERSUS PANDEMONIUM

One of the skeletons in the closet of contemporary linguistics is that it has lavished attention on hearing but largely ignored speaking, which one might say was roughly half of language, and the most important half at that. Although there are many detailed theories and models of language perception, and of the comprehension of heard utterances (the paths from phonology, through syntax, to semantics and pragmatics), no one — not Noam Chomsky, and not any of his rivals or followers — has had anything very substantial (right or wrong) to say about systems of language production. It is as if all theories of art were theories of art appreciation with never a word about the artists who created it — as if all art consisted of *objets trouvés* appreciated by dealers and collectors.

It is not hard to see why this is so. Utterances are readily found objects with which to begin a process. It is really quite clear what the raw material or input to the perception and comprehension systems is: wave forms of certain sorts in the air, or strings of marks on various plane surfaces. And although there is considerable fog obscuring the controversies about just what the end product of the comprehension process is, at least this deep disagreement comes at the end of the process being studied, not the beginning. A race with a clear starting line can at least be rationally begun, even if no one is quite sure where it is going to end. Is the "output" or "product" of speech comprehension a decoding or translation of the input into a new representation — a sentence of Mentalese, perhaps, or a picture-in-the-head — or is it a set of deep structures, or some still unimagined entity? Linguists can de-

cide to postpone an answer to that stumper while they work on the more peripheral parts of the process.

With speech production, on the other hand, since no one has yet worked out any clear and agreed-upon description of what initiates the process that eventually yields a full-fledged utterance, it is hard even to get started on a theory. Hard, but not impossible. There has been some good work on the issues of production quite recently, excellently surveyed and organized by the Dutch psycholinguist Pim Levelt, in *Speaking* (1989). Working backwards from the output, or working from the middle in both directions, we get some suggestive glimpses into the machinery that designs our utterances and gets them expressed. (The following examples are drawn from Levelt's discussion.)

Speech is not produced by a "batch process" that designs and executes one word at a time. The existence of at least a limited look-ahead capacity in the system is revealed by the way stress gets distributed in an utterance. A simple case: the stress in the word "sixteen" depends on context:

ANDY: How many dollars does it cost?

BOB: I think it's sixTEEN.

ANDY: SIXteen dollars isn't very much.

When Andy gives his second speech, he must adjust his pronunciation of "sixteen" to the word (DOLLars) that follows. Had he been going to say:

SixTEEN isn't very much.

he would have given the word a different stress pattern. Another example: notice how different the stress is on the two occurrences of "Tennessee" in

I drove from Nashville, TennesSEE, to the TENnessee border.

Spoonerisms and other speech errors show quite conclusively how lexical and grammatical distinctions are observed (and misobserved) in the course of designing an utterance to speak. People are more apt to say "barn door" when they mean "darn bore" than they are to say "bart doard" when they mean "dart board". There is a bias in favor of real (familiar) words over merely pronounceable (possible but not actual) words even when making a slip of the tongue. Some errors are suggestive about how word-selection mechanisms must operate: "The competition is a little stouger [stiffer/tougher]," and "I just put it in the oven at very low speed." And think of the transposition that must

be involved in producing an error like "naming a wear tag" for "wearing a name tag."

Thanks to ingenious experiments that provoke such errors, and intricate analyses of what does and doesn't happen when people speak, progress is being made on models of the highly organized mechanisms that execute the ultimate articulation of a message once it has been decided that a particular message is to be released to the outside world. But who or what puts this machinery in motion? A speech error is an error in virtue of being other than what the speaker meant to say. What taskmaster sets the task relative to which errors such as the examples above are judged?

What, if not the Central Meaner? Levelt provides us with a picture, a "blueprint for the speaker":

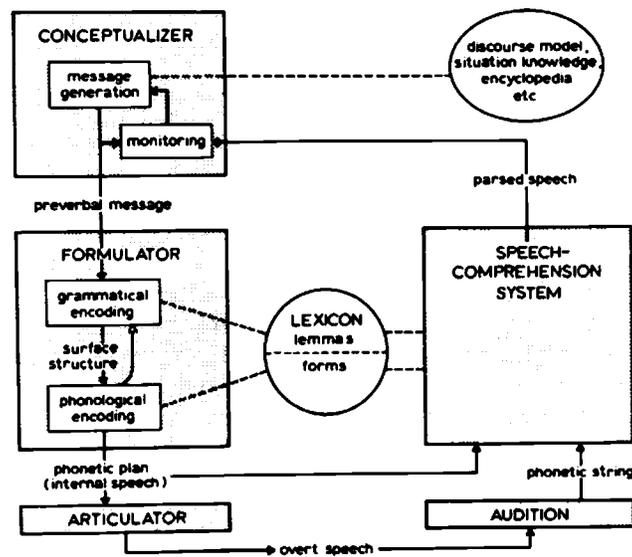


Figure 8.1

In the upper left-hand corner a functionary who looks suspiciously like the Central Meaner makes his appearance in the guise of the Conceptualizer, armed with lots of world knowledge, plans, and communicative intentions, and capable of "message generation." Levelt warns his readers that the Conceptualizer "is a reification in need of further explanation" (p. 9), but he posits it anyway, since he really can't get the process going, it seems, without some such unanalyzed Boss to give the marching orders to the rest of the team.

How does it work? The underlying problem will be clearer if we begin with a caricature. The Conceptualizer decides to perform a speech act, such as insulting his interlocutor by commenting adversely on the size of his feet. So he sends a command to the bureaucracy under his sway, the Public Relations Department (Levelt's Formulator): "Tell this bozo his feet are too big!" The PR people take on the job. They find the appropriate words: the second-person singular possessive pronoun, *your*; a good word for feet, such as *feet*; the right plural form of the verb *to be*, namely *are*; and the appropriate adverb and adjective: *too big*. These they cunningly combine, with the right insulting tone of voice, and execute:

"Your feet are too big!"

But wait a minute. Isn't that too easy? When the Conceptualizer gave the command (what Levelt calls the *preverbal message*), if he gave it in English, as my caricature just suggested, he's done all the hard work, leaving little for the rest of the team to do, except to pass it along with trivial adjustments. Then is the preverbal message in some other representational system or language? Whatever it is, it must be capable of providing the basic "specs" to the production team for the object they are to compose and release, and it must be couched in terms they can "understand" — not English but some version of Brainish or Mentalese. It will have to be in a sort of language of thought, Levelt argues, but perhaps in a language of thought that is used only to order speech acts, not for all cognitive activities. The team receives the preverbal message, a detailed Mentalese order to make an English utterance, and then it fills this order. This gives the subordinates a little more to do, but just obscures the looming regress. How does the Conceptualizer figure out which words of Mentalese to use to give the order? There had better not be a smaller duplicate of Levelt's whole blueprint hidden in the Conceptualizer's *message generation box* (and so on, *ad infinitum*). And certainly nobody told the Conceptualizer what to say; he's the Central Meaner, after all, where meaning originates.

How then does the meaning of an utterance develop? Consider the following nesting of commands, leading from grand overall strategy through detailed tactics to basic operations:

- (1) Go on the offensive!
- (2) Do something nasty but not too dangerous to him!
- (3) Insult him!
- (4) Cast aspersions on some aspect of his body!
- (5) Tell him his feet are too big!

- (6) Say: "Your feet are too big!"
 (7) Utter: *yər FĪT ər tū bĭg!*

Surely something like this zeroing-in on the final act must happen. Human speech is purposive activity; there are ends and means, and we somehow do a passable job of traversing the various options. We could have shoved him instead of insulting him, or belittled his intelligence instead of enlarging on his feet, or said, quoting Fats Waller, "Your pedal extremities are obnoxious!"

But does this zeroing-in all get accomplished by a bureaucratic hierarchy of commanders giving orders to subordinates? In this cascade of commands there appears to be a lot of decision-making — "moments" at which options are "selected" over their rivals, and this invites a model in which there is delegation of responsibility for finer details, and in which subordinate agents with their own intentions appreciate reasons for the various selections they make. (If they didn't have to understand at all why they were doing what they were doing, they wouldn't really be agents, but just passive, rubber-stamping functionaries letting whatever happened to cross their desks control them.)

Levelt's blueprint exhibits fossil traces of one of its sources: the von Neumann architecture that was inspired by Turing's reflections on his own stream of consciousness and that has inspired in turn many models in cognitive science. In chapter 7, I attempted to overcome resistance to the idea that human consciousness is rather like a von Neumann machine, a serial processor with a succession of definite contents reeling through the bottleneck of the accumulator. Now I must put on the brakes and emphasize some ways in which the functional architecture of human consciousness is not like that of a von Neumann machine. If we compare Levelt's blueprint to the way von Neumann machines standardly emit words, we can see that Levelt's model may borrow slightly too much.

When a von Neumann machine says what is written in its heart, it outputs the contents of its single central workplace, the accumulator, which at each instant has entirely specific contents in the fixed language of binary arithmetic. The rudimentary "preverbal messages" of a von Neumann machine look like this: 10110101 00010101 11101101. One of the primitive instructions in any machine language is an OUTPUT instruction, which can take the current contents of the accumulator (e.g., the binary number 01100001) and write it on the screen or the printer, so that an outside user can gain access to the results accomplished in the CPU. In a slightly more user-friendly variation, a routine

operation composed of a series of primitive instructions can first translate the binary number into decimal notation (e.g., binary 00000110 = decimal 6) or into a letter of the alphabet via ASCII code (e.g., binary 01100001 = "a" and 01000001 = "A") and then output the result. These subroutines are at the heart of the fancier output instructions found in higher-level programming languages, like Fortran or Pascal or Lisp. These permit the programmer to create further subroutines for building larger messages, fetching long series of numbers from memory and running them through the accumulator, translating them and writing the results on the screen or printer. For instance, a subroutine can make several trips to the accumulator for values to plug into the blanks in

You have overdrawn your account, Mr. _____, by \$ _____. Have a nice day, Mr. _____!

— a "canned" sentence formula that itself is kept stored as a series of binary numbers in the memory until some subroutine determines that it is time to open the can. In this manner, a strict hierarchy of fixed routines can turn sequences of specific contents in the accumulator into expressions that a human being can read on a screen or printer: "Do you want to save this document?" or "6 files copied" or "Hello, Billy, do you want to play tic-tac-toe?"

There are two features of this process that are shared by Levelt's model: (1) the process takes an already determinate content as its input, and (2) the bureaucracy — the "flow of control" in computer-science jargon — has to have been carefully designed: all "decision-making" flows hierarchically by a delegation of responsibility to subagents whose job descriptions dictate which bit of means/ends analysis they are authorized to perform. Interestingly enough, the first of these features — the determinate content — seems to be endorsed by Otto's view of his own processes: There is a determinate "thought" somewhere in the Center, waiting to be "put into words." The second shared feature, however, seems alien: The hierarchy of routines that slavishly render *that very thought in natural language* have been predesigned by someone else — by the programmer, in the case of the von Neumann machine, and presumably by the combination of evolution and individual development in the case of the activities in Levelt's Formulator. The creative, judgmental role that the thinker of the thought should play in getting the thought into words does not appear in the model; it is either usurped by the Conceptualizer, who does all the creative work before sending an order to the Formulator, or it is implicit in the design of the Formulator, a *fait accompli* of some earlier design process.

How else could ends and means be organized? Let's consider an opposing caricature: a pandemonium of word-demons. Here is how we talk: First we go into vocal noise-making mode — we turn on the horn:

Beeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeep. . . .

We do this for no good reason, but just because no good reason not to do it occurs to us. The internal "noise" excites various demons in us who begin trying to modulate the horn in all sorts of random ways by interfering with its stream. The result is gibberish, but at least it's English gibberish (in English speakers):

Yabba-dabba-doo-fiddledy-dee-tiddly-pom-fi-fi-fo-fum. . . .

But before any of this embarrassing stuff actually hits the outside world, further demons, sensitive to patterns in the chaos, start shaping it up into words, phrases, clichés . . .

And so, how about that?, baseball, don't you know, in point of fact, strawberries, happenstance, okay? That's the ticket. Well, then . . .

which incites demons to make further serendipitous discoveries, augmented by opportunistic shaping, yielding longer bits of more acceptable verbiage, until finally a whole sentence emerges:

I'm going to knock your teeth down your throat!

Fortunately, however, this gets set aside, unspoken, since at the same time (in parallel) other candidates have been brewing and are now in the offing, including a few obvious losers, such as

You big meany!

and

Read any good books lately?

and a winner by default, which gets spoken:

Your feet are too big!

The muse has failed our speaker on this occasion; no witty retort made it to the finals, but at least something halfway appropriate to the speaker's current "mind-set" got blurted out. As the speaker walks away after the encounter, he will probably resume the chaotic tournament, muttering and musing about what he should have said. The muse may

then descend with something better, and the speaker will savor it, turning it over and over in his mind, imagining the stricken look it would have provoked on the face of his interlocutor. By the time the speaker gets home, he may vividly “remember” having skewered his interlocutor with a razor-sharp witticism.

We can suppose that all of this happens in swift generations of “wasteful” parallel processing, with hordes of anonymous demons and their hopeful constructions never seeing the light of day — either as options that are consciously considered and rejected, or as ultimately executed speech acts for outsiders to hear. If given enough time, more than one of these may be silently tried out in a conscious rehearsal, but such a formal audition is a relatively rare event, reserved for occasions where the stakes are high and misspeaking carries heavy penalties. In the normal case, the speaker gets no preview; he and his audience learn what the speaker’s utterance is at the same time.

But how is this tournament of words judged? When one word or phrase or whole sentence beats out its competitors, how does its suitability or appropriateness to the current mind-set get discriminated and valued? What is a mind-set (if not an explicit communicative intention), and how does its influence get conveyed to the tournament? For after all, even if there isn’t a Central Meaner, there has to be some way for the content to get from deep inside the system — from perceptual processes, for instance — to verbal reports.

Let’s review the issues. The problem with the bureaucratic extreme is that the Conceptualizer seems ominously powerful, a homunculus with too much knowledge and responsibility. This excess of power is manifested in the awkward problem of how to couch its output, the *preverbal message*. If it already specifies a speech act — if it already is a sort of speech act in Mentalese, a specific command to the Formulator — most of the hard work of composition has happened before our model kicks in. The problem with the Pandemonium alternative is that we need to find a way in which sources of content can influence or constrain the creative energies of the word-demons without dictating to them.

What about the process described in chapter 1, the rounds of question-answering that generated hallucinations on the model of the game of Psychoanalysis? Recall that we eliminated the wise Freudian dream-playwright and hallucination-producer by replacing him with a process from which content *emerged* under the incessant questioning

of a questioner. The residual problem was how to get rid of the clever questioner, a problem we postponed. Here we have the complementary problem: how to get answers to an eager flock of contestants asking questions like "Why don't we say, 'Your mother wears army boots!'" or (in another context) "Why don't we say, 'I seem to see a red spot moving and turning green as it moves?'" *Two complementary problems — could they perhaps solve each other by being mated?* What if the word-demons are, in parallel, the questioners/contestants, and the content-demons are the answerers/judges? Fully fledged and executed communicative intentions — Meanings — could emerge from a quasi-evolutionary process of speech act design that involves the collaboration, partly serial, partly in parallel, of various subsystems none of which is capable on its own of performing — or ordering — a speech act.

Is such a process really possible? There are a variety of models of such "constraint satisfaction" processes, and they do indeed have striking powers. In addition to the various "connectionist" architectures of neuronlike elements (see, e.g., McClelland and Rumelhart, 1986), there are other more abstract models. Douglas Hofstadter's (1983) Jumbo architecture, which hunts for solutions to Jumbles or anagrams, has the right sorts of features, and so do Marvin Minsky's (1985) ideas about the Agents making up the "society of mind" — which will be discussed further in chapter 9. But we must reserve judgment until models that are more detailed, explicit, and directly aimed at language production are created and put through their paces. There may be surprises and disappointments.

We know, however, that somewhere in any successful model of language production we must avail ourselves of an evolutionary process of message generation, since otherwise we will be stuck with a miracle ("And then a miracle occurs") or an infinite regress of Meaners to set the task.¹ We also know — from the research Levelt surveys — that

1. Dan Sperber and Deirdre Wilson (1986) open up a new perspective on how we compose our communications by insisting on models of how things actually work. in the speaker and hearer, contrary to recent practice among philosophers and linguists, who have tended to wave their hands about the mechanisms while appealing to rational reconstructions of the supposed tasks and their demands. This permits Sperber and Wilson to raise considerations of practicality and efficiency: least-effort principles, and concerns about timing and probability. They then show from this new perspective how certain traditional "problems" disappear — in particular, the problem of how the hearer finds the "right" interpretation of what the speaker intended. Although they do not pitch

there are quite rigid and automatic processes that take over eventually and determine the grammatical-to-phonological transformations that compose the final muscular recipe for speech. The two caricatures define extremes along a continuum, from hyperbureaucratic to hyperchaotic. Levelt's actual model — in contrast to the caricature I have used in order to make the contrast vivid — incorporates (or can readily be made to incorporate) some of the nonbureaucratic features of the opposing caricature: for example, there is nothing deep or structural preventing Levelt's Formulator from engaging in more or less spontaneous (unrequested, undirected) language generation, and, given the monitoring loop through the Speech-Comprehension System back to the Conceptualizer (see Figure 8.1), this spontaneous activity could play the sort of generating role envisaged for the multiple word-demons. Between the two caricatures there is an intervening spectrum of more realistic ways alternative models could be developed. The main question is how much interaction is there between the specialists who determine the content and style of what is to be said and the specialists who "know the words and the grammar"?

At one extreme, the answer is: None. We could keep Levelt's model intact, and simply supplement it with a pandemonium model of what happens inside the Conceptualizer to fix the "preverbal message." In Levelt's model, there is nearly complete separation between the processes of message generation (specs-setting) and linguistic production (specs-meeting). When the first bit of preverbal message arrives at the Formulator, it triggers the production of the beginning of an utterance, and as the words get chosen by the Formulator, this constrains how the utterance can continue, but there is minimal *collaboration* on revision of the specs. The subordinate language-carpenters in the Formulator are, in Jerry Fodor's terms, "encapsulated"; in their automatic way, they do the best they can with the orders they receive, with no ifs, ands, or buts.

At the other extreme are the models in which words and phrases from the Lexicon, together with their sounds, meanings, and associations, jostle with grammatical constructions in a pandemonium, all "trying" to be part of the message, and some of them thereby make a substantial contribution to the very communicative intentions that still fewer of them end up executing. At this extreme, the communicative

their model at the level of evolutionary processes of the sorts we have just been considering, it certainly invites just such an elaboration.

intentions that exist are as much an effect of the process as a cause — they emerge as a product, and once they emerge, they are available as standards against which to measure further implementation of the intentions. There is not one source of meaning, but many shifting sources, opportunistically developed out of the search for the right words. Instead of a determinate content in a particular functional place, waiting to be Englished by subroutines, there is a still-incompletely-determined mind-set distributed around in the brain and constraining a composition process which in the course of time can actually feed back to make adjustments or revisions, further determining the expressive task that set the composition process in motion in the first place. There still is an overall pattern of serial passage, with concentration on one topic at a time, but the boundaries are not sharp lines.

In the Pandemonium model, control is usurped rather than delegated, in a process that is largely undesigned and opportunistic; there are multiple sources for the design “decisions” that yield the final utterance, and no strict division is possible between the marching orders of content flowing from within and the volunteered suggestions for implementation posed by the word-demons. What this brand of model suggests is that in order to preserve the creative role of the thought-expresser (something that mattered a good deal to Otto), we have to abandon the idea that the thought-thinker begins with a determinate thought to be expressed. This idea of determinate content also mattered a good deal to Otto, but something has to give (and section 4 will explore the alternatives more fully).

Where on the spectrum does the truth reside? This is an empirical question to which we do not yet know the answer.² There are some phenomena, however, that strongly suggest (to me) that language-

2. As Levelt notes, “If one could show, for instance, that message generation is directly affected by the accessibility of lemmas or word forms, one would have evidence for direct feedback from the Formulator to the Conceptualizer. This is an empirical question, and it is possible to put it to the test. . . . So far, the evidence for such feedback is negative” (p. 16). The evidence he reviews is from tightly controlled experiments in which a very specific task was given to the speaker: such as *describe the picture on the screen as fast as you can* (pp. 276–282). This is excellent negative evidence — I for one was surprised at the lack of effect in these experiments — but, as he recognizes, it is not at all conclusive. It is not really *ad hoc* to claim that the artificiality of these experimental situations successfully drowned out the opportunistic/creative dimension of language use. But perhaps Levelt is right; perhaps the only feedback from Formulator to Conceptualizer is indirect: the sort of feedback that a person can produce only by explicitly talking to himself and then framing an opinion about what he finds himself saying.

generation will turn out to involve Pandemonium — opportunistic, parallel, evolutionary processes — almost all the way down. The next section will review some of them briefly.

3. WHEN WORDS WANT TO GET THEMSELVES SAID

Whatever we may want to say, we probably won't say exactly *that*.

MARVIN MINSKY (1985), p. 236

The AI researchers Lawrence Birnbaum and Gregg Collins (1984) have noted a peculiarity about Freudian slips. Freud famously drew our attention to slips of the tongue that were not random or meaningless, he insisted, but deeply meaningful: unconsciously *intended* insertions into the fabric of discourse, insertions that indirectly or partially satisfied suppressed communicative goals of the speaker. This standard Freudian claim has often been vehemently rejected by skeptics, but there is something puzzling about its application to particular cases that has nothing to do with one's opinion about Freud's darker themes of sexuality, the Oedipus complex, or death wishes. Freud discussed an example in which a man said

Gentlemen, I call upon you to hiccup to the health of our Chief.

(In German — the language actually spoken in the example — the word for "hiccup," *aufzustossen*, was slipped in for the word for "drink," *anzustossen*.)

In his explanation, Freud argues that this slip is a manifestation of an unconscious goal on the part of the speaker to ridicule or insult his superior, suppressed by the social and political duty to do him honor. However, . . . one cannot reasonably expect that the speaker's intention to ridicule his superior gave rise originally to a plan involving the use of the word "hiccup": *A priori*, there are hundreds of words and phrases that can more plausibly be used to insult or ridicule someone. . . . There is no way that a planner could have reasonably anticipated that the goal of ridiculing or insulting its superior would be satisfied by uttering the word "hiccup," for exactly the same reason that it is implausible that the

planner would have chosen to use the word as an insult in the first place.

The only process that could explain the frequency of serendipitous Freudian slips, they argue, is one of "opportunistic planning."

. . . What examples like the above seem to indicate, therefore, is that the goals *themselves* are active cognitive agents, capable of commanding the cognitive resources needed to recognize opportunities to satisfy themselves, and the behavioral resources needed to take advantage of the opportunities. [Birnbaum and Collins, 1984, p. 125]

Freudian slips draw attention to themselves by seeming to be mistakes and not mistakes at the same time, but the fact (if it is one) that they satisfy unconscious goals does not make them any harder to explain than other word choices that fulfill several functions (or goals) at once. It is about as hard to imagine how puns and other forms of intended verbal humor could be the result of nonopportunistic, encapsulated planning and production. If anyone has a plan for designing witticisms — a detailed plan that actually works — there are more than a few comedians who would pay good money for it.³

If Birnbaum and Collins are right, creative language use can be accomplished only by a parallel process in which multiple goals are simultaneously on the alert for materials. But what if the materials themselves were at the same time on the alert for opportunities to get incorporated? We pick up our vocabulary from our culture; words and phrases are the most salient phenotypic features — the visible bodies — of the memes that invade us, and there could hardly be a more congenial medium in which memes might replicate than a language-production system in which the supervisory bureaucrats had partially abdicated, ceding a large measure of control to the words themselves, who in effect fight it out among themselves for a chance in the limelight of public expression.

3. Levelt tells me that he himself is an inveterate pun-hunter (in his native Dutch), and he knows just how he does it: "By lifelong training I turn around just about every word I hear. I then (quite consciously) check the result for its meaning. In 99.9 percent of the cases there is nothing funny coming out. But one per thousand is fine, and those I express right away" (personal communication). This is a perfect example of von Neumanesque problem-solving: serial, controlled — and conscious! The question is whether there are other, more pandemonic, ways of generating wit unconsciously.

It is no news that some of what we say we say primarily because we like the way it sounds, not because we like what it means. New slang sweeps through subcommunities, worming its way into almost everybody's speech, even those who try to resist it. Few of those who use a new word are deliberately or consciously following the school-teacher's maxim "Use a new word three times and it's yours!" And at larger levels of aggregation, whole sentences appeal to us for the way they ring in our ears or trip off our tongues, quite independently of whether they meet some propositional specs we have already decided upon. One of the most quotable lines Abraham Lincoln ever came up with is:

You can fool all the people some of the time, and some of the people all the time, but you can not fool all the people all of the time.⁴

What did Lincoln mean? Logic teachers are fond of pointing out that there is a "scope ambiguity" in the sentence. Did Lincoln mean to assert that there are some dunces who can always be fooled, or that on every occasion, someone or other is bound to be fooled — but not always the same people? Logically, these are entirely different propositions.

Compare:

"Someone always wins the lottery."

"It must be rigged!"

"That's not what I meant."

Which reading did Lincoln intend? Maybe neither! What are the odds that Lincoln never noticed the scope ambiguity and never actually got around to having one communicative intention rather than "the other"? Perhaps it just sounded so good to him when he first formulated it that he never picked up the ambiguity, *and never had any prior communicative intention* — except the intention to say something pithy and well cadenced on the general topic of fooling people. People do talk that way, even great Meaners like Lincoln.

The fiction writer Patricia Hampl, in a thoughtful essay, "The Lax

4. According to *The Oxford Dictionary of Quotations* (second edition, 1953), this famous sentence is also attributed to Phineas T. Barnum. Since Barnum is an illustrious alumnus and generous benefactor of my university, I feel duty-bound to draw attention to the possibility that Lincoln may not be the originator of this highly replicative meme.

Habits of the Free Imagination," writes about her own process of composing short stories.

Every story has a story. This secret story, which has little chance of getting told, is the history of its creation. Maybe the "story of the story" can never be told, for a finished work consumes its own history, renders it obsolete, a husk. [Hampl, 1989, p. 37]

The finished work, she notes, is readily interpretable by critics as an artifact cunningly contrived to fulfill a host of sophisticated authorial intentions. But when she encounters these hypotheses about her own work, she is embarrassed:

"Hampl" had precious few intentions, except, like the charlatan I suddenly felt myself to be, to filch whatever was loose on the table that suited my immediate purposes. Worse, the "purposes" were vague, inconsistent, reversible, under pressure. And who — or what — was applying the pressure? I couldn't say. [p. 37]

How then does she do it? She suggests a maxim: "Just keep talking — mumbling is fine." Eventually, the mumbling takes on shapes that meet with the approval of the author. Could it be that the process Hampl detects on a grand scale in her creative writing is just an enlargement of the more submerged and swift process that produces the creative speaking of everyday life?

The tempting similarity does not involve just a process but also a subsequent attitude or reaction. Hampl's confessional zeal contrasts with a more normal — and not really dishonest — reaction of authors to friendly interpretations by readers: these authors defer gracefully to the imputations of intent, and even willingly elaborate on them, in the spirit of "Hey, I guess that is what I was up to, all along!" And why not? Is there anything self-contradictory in the reflection that a certain move one has just made (in chess, in life, in writing) is actually cleverer than one at first realized? (For further reflections on this topic, see Eco, 1990.)

As E. M. Forster put it, "How do I know what I think until I see what I say?" We often do discover what we think (and hence what we mean) by reflecting on what we find ourselves saying — and not correcting. So we are, at least on those occasions, in the same boat as our external critics and interpreters, encountering a bit of text and putting the best reading on it that we can find. The fact that we said it gives it a certain personal persuasiveness or at least a presumption of authen-

ticity. Probably, if I said it (and I heard myself say it, and I didn't hear myself rushing in with any amendments), I meant it, and it probably means what it seems to mean — to me.

Bertrand Russell's life provides an example:

It was late before the two guests left and Russell was alone with Lady Ottoline. They sat talking over the fire until four in the morning. Russell, recording the event a few days later, wrote, "I did not know I loved you till I heard myself telling you so — for one instant I thought 'Good God, what have I said?' and then I knew it was the truth." [Clark, 1975, p. 176]

What about the other occasions, though, where we have no such sense of a *discovery* of self-interpretation? We might suppose that in these, the normal, cases, we have some intimate and privileged advance insight into what we mean, just because we ourselves are the Meaners, the *fons et origo* of the meaning of the words we say, but such a supposition requires a supporting argument, not just an appeal to tradition. For it could as well be the case that we have no sense of discovery in these cases just because it is so obvious to us what we mean. It doesn't take "privileged access" to intuit that when I say, "Please pass the salt" at the dinner table, I'm asking for the salt.

I used to believe there was no alternative to a Central Meaner, but I thought I had found a safe haven for it. In *Content and Consciousness* I argued that there had to be a functionally salient line (which I called the awareness line) separating the preconscious fixation of communicative intentions from their subsequent execution. The location of this line in the brain might be horrendously gerrymandered, anatomically, but it had to exist, logically, as the watershed dividing malfunctions into two varieties. Errors could occur anywhere in the whole system, but every error had to fall — by geometric necessity — on one side of the line or the other. If they fell on the execution side of the line, they were (correctable) errors of expression, such as slips of the tongue, malapropisms, mispronunciations. If they fell on the inner or higher side of the line, they changed that which was to be expressed (the "preverbal message" in Levelt's model). Meaning was fixed at this watershed; that's where meaning came from. There had to be such a place where meaning came from. I thought, since something has to set the standard against which "feedback" can register failure to execute.

My mistake was falling for the very same scope ambiguity that bedevils the interpretation of Abe Lincoln's dictum. There does indeed have to be something on each occasion that is, for the nonce, the stan-

dard against which any corrected “error” gets corrected, but there doesn’t have to be the same single thing each time — even within the duration of a single speech act. There doesn’t have to be a *fixed* (if gerrymandered) line that marks this distinction. In fact, as we saw in chapter 5, the distinction between pre-experiential revisions that *change that which was experienced* and post-experiential revisions that have the effect of *misreporting or misrecording what was experienced* is indeterminate in the limit. Sometimes subjects are moved to revise or amend their assertions, and sometimes they aren’t. Sometimes when they do make revisions, the edited narrative is no closer to “the truth” or to “what they *really* meant” than the superseded version. As we noted earlier, where prepublication editing leaves off and postpublication errata-insertion cuts in is a distinction that can be drawn only arbitrarily. When we put a question to a subject about whether or not a particular public avowal adequately captures the ultimate inner truth about what he was just experiencing, the subject is in no better position to judge than we outsiders are. (See also Dennett, 1990d.)

Here is another way of looking at the same phenomenon. Whenever the process of creating a verbal expression occurs, there is at the outset a distance that must be eliminated: the “mismatch distance in semantic space,” we might call it, between the content that is in position to be expressed and the various candidates for verbal expression that are initially nominated. (In my old view, I treated this as a problem of simple “feedback correction,” with a *fixed point* for a standard against which verbal candidates were to be measured, discarded, improved.) The back-and-forth process that narrows the distance is a feedback process of sorts, but it is just as possible for the content-to-be-expressed to be adjusted in the direction of some candidate expression, as for the candidate expression to be replaced or edited so better to accommodate the content-to-be-expressed. In this way, the most accessible or available words and phrases could actually *change the content of the experience* (if we understand the experience to be what is ultimately reported — the settled event in the heterophenomenological world of the subject).⁵

If our unity as Meaners is no better guaranteed than this, then in

5. This is reminiscent of Freud’s view of how the “preconscious” works: “The question, ‘How does a thing become conscious?’ would be more advantageously stated: ‘How does a thing become preconscious?’, and the answer would be: ‘Through becoming connected with the word-presentations corresponding to it’ ” (*The Ego and the Id*, English edition, 1962, p. 10).

principle it ought to be possible for it to become shattered on rare occasions. Here are two cases in which that seems to have happened.

I was once importuned to be the first base umpire in a baseball game — a novel duty for me. At the crucial moment in the game (bottom of the ninth, two outs, the tying run on third base), it fell to me to decide the status of the batter running to first. It was a close call, and I found myself emphatically jerking my thumb up — the signal for OUT — while yelling "SAFE!" In the ensuing tumult I was called upon to say what I had meant. I honestly couldn't say, at least not from any privileged position. I finally decided (to myself) that since I was an unpracticed hand-signaler but competent word-speaker, my vocal act should be given the nod, but anyone else could have made just the same judgment. (I would be happy to learn of other anecdotes in which people have not known which of two very different speech acts they had meant to perform.)

In an experimental setting, the psychologist Tony Marcel (in press) has found an even more dramatic case. The subject, who suffers from blindsight (about which I will say more in chapter 11), was asked to say whenever he thought there was a flash of light, but he was given peculiar instructions about how he was to do this. He was instructed to perform this single speech act by three distinct acts at once (not in sequence, but not necessarily "in unison" either):

- (1) saying "Yes"
- (2) pressing a button (the YES button)
- (3) blinking YES

What is startling is that the subject didn't always perform all three together. Occasionally he blinked YES but didn't say YES or button-push YES, and so forth. There was no straightforward way of ordering the three different responses, either for fidelity to intention or for accuracy. That is, when there were disagreements among the three actions, the subject had no pattern to follow about which act to accept and which to count as a slip of the tongue, finger, or eyelid.

Whether similar findings can be provoked under other conditions with other subjects, normal or otherwise, remains to be seen, but other pathological conditions also suggest a model of speech production in which verbalization can be set in motion without any marching orders from a Central Meander. If you suffer from one of these pathologies, "your mind is on vacation, but your mouth is working overtime," as the Mose Allison song puts it.

Aphasia is loss or damage of the ability to speak, and several

different varieties of aphasia are quite common and have been extensively studied by neurologists and linguists. In the most common variety, Broca's aphasia, the patient is acutely aware of the problem, and struggles, with mounting frustration, to find the words that are just beyond the tip of her tongue. In Broca's aphasia, the existence of thwarted communicative intentions is painfully clear to the patient. But in a relatively rare variety of aphasia, jargon aphasia, patients seem to have no anxiety at all about their verbal deficit.⁶ Even though they are of normal intelligence, and not at all psychotic or demented, they seem entirely content with such verbal performances as these (drawn from two cases described by Kinsbourne and Warrington, 1963):

Case 1:

How are you today?

"Gossiping O.K. and Lords and cricket and England and Scotland battles. I don't know. Hypertension and two won cricket, bowling, batting, and catch, poor old things, cancellations maybe gossiping, cancellations, arm and argument, finishing bowling."

What is the meaning of "safety first"?

"To look and see and the Richmond Road particularly, and look traffic and hesitation right and strolling, very good cause, maybe, zebras maybe these, motor-car and the traffic light."

Case 2:

Did you work in an office?

"I did work in an office."

And what kind of firm was it?

"Oh, as an executive of this, and the complaint was to discuss the tonations as to what type they were, as to how they were typed, and kept from the different . . . tricu . . . tricula, to get me from the attribute convenshments . . . sorry . . ."

"She wants to give one the subjective vocation to maintain the vocation of perfect impregnation simbling."

"Her normal corrucation would be a dot."

6. Levelt tells me that research underway at the Max Planck Institute for Psycholinguistics in Nijmegen casts doubt on this, the received view. Work by Heeschen suggests that at some level jargon or Wernicke's aphasics do have anxiety about their deficit, and seem to adopt a strategy of repetition, in hopes of achieving communication.

asked to identify a nail file:

"That is a knife, a knife tail, a knife, stale, stale knife."

and scissors:

"Groves — it's a groves — it's not really a groves — two groves containing a comb — no, not a comb — two groves providing that the commandant is not now — "

A strangely similar condition, and much more common, is *confabulation*. In chapter 4, I suggested that normal people may often confabulate about details of their own experience, since they are prone to guess without realizing it, and mistake theorizing for observing. Pathological confabulation is unwitting fiction of an entirely different order. Often in cases of brain damage, especially when people have terrible memory loss — as in Korsakoff's syndrome (a typical sequel of severe alcoholism) — they nevertheless proceed to prattle on with utter falsehoods about their lives and their past histories, and even about events of the last few minutes, if their amnesia is severe.

The resulting verbiage sounds virtually normal. It often sounds, in fact, just like the low-yield, formulaic chitchat that passes for conversation in a bar: "Oh, yes, my wife and I — we've lived in the same house for thirty years — used to go out to Coney Island, and, you know, sit on the beach — loved to sit on the beach, just watching the young people, and, but that was before the accident . . ." — except that it is made up out of whole cloth. The man's wife may have died years ago, never been within a hundred miles of Coney Island, and they may have moved from apartment to apartment. An uninitiated listener can often be entirely unaware that he is encountering a confabulator, so natural and "sincere" are the reminiscences and the ready answers to questions.

Confabulators have no idea they are making it all up, and jargon aphasics are oblivious to the fact that they are spouting word-salad. These stunning anomalies are instances of *anosognosia*, or inability to acknowledge or recognize a deficit. Other varieties of this absence of self-monitoring exist, and in chapter 11 we will consider what they have to tell us about the functional architecture of consciousness. In the meantime, we can note that the brain's machinery is quite able to construct apparent speech acts in the absence of any coherent direction from on high.⁷

7. Another anomalous linguistic phenomenon is the familiar symptom of schizophrenia: "hearing voices." It is now quite firmly established that the voice the schizophrenic "hears" is his own: he is talking to himself silently without realizing it. As simple an obstacle as having the patient hold his mouth wide open is sufficient to stop

Pathology, either the temporary strain induced by clever experiments, or the more permanent breakdowns caused by illness or mechanical damage to the brain, provides an abundance of clues about how the machinery is organized. These phenomena suggest to me that our second caricature, Pandemonium, is closer to the truth than a more dignified, bureaucratic model would be, but this has yet to be put to the proper empirical test. I am not claiming that it is impossible for a largely bureaucratic model to do justice to these pathologies, but just that they would not seem to be the natural failings of such a system. In Appendix B, for scientists, I will mention some research directions that could help confirm or disconfirm my hunch.

What I have sketched in this chapter — but certainly not proven — is a way in which a torrent of verbal products emerging from thousands of word-making demons in temporary coalitions could exhibit a unity, the unity of an evolving best-fit interpretation, that makes them appear as if they were the executed intentions of a Conceptualizer — and indeed they are, but not of an inner Conceptualizer that is a proper part of the language-producing system, but of the global Conceptualizer, the person, of which the language-producing system is itself a proper part.

This idea may seem alien at first, but it should not surprise us. In biology, we have learned to resist the temptation to explain design in organisms by positing a single great Intelligence that does all the work. In psychology, we have learned to resist the temptation to explain seeing by saying it is just as if there were an internal screen-watcher, for the internal screen-watcher does all the work — the only thing between such a homunculus and the eyes is a sort of TV cable. We must build up the same resistance to the temptation to explain action as arising from the imperatives of an internal action-orderer who does too much of the specification work. As usual, the way to discharge an intelligence that is too big for our theory is to replace it with an ultimately mechanical fabric of semi-independent semi-intelligences acting in concert.

This point applies not just to speech act generation; it applies to intentional action across the board. (See Pears, 1984, for a development of similar ideas.) And contrary to some first appearances, phenomenology actually assists us in seeing that this is so. Although we are occasionally conscious of performing elaborate practical reasoning, leading to a conclusion about what, all things considered, we ought to

the voices (Bick and Kinsbourne, 1987). See also Hoffman (1986), and the commentary by Akins and Dennett, "Who May I Say Is Calling?" (1986).

do, followed by a conscious decision to do that very thing, and culminating finally in actually doing it, these are relatively rare experiences. Most of our intentional actions are performed without any such preamble, and a good thing, too, since there wouldn't be time. The standard trap is to suppose that the relatively rare cases of conscious practical reasoning are a good model for the rest, the cases in which our intentional actions emerge from processes into which we have no access. Our actions generally satisfy us; we recognize that they are in the main coherent, and that they make appropriate, well-timed contributions to our projects as we understand them. So we safely assume them to be the product of processes that are reliably sensitive to ends and means. That is, they are rational, in one sense of that word (Dennett, 1987a, 1991a). But that does not mean they are rational in a narrower sense: the product of serial reasoning. We don't have to explain the underlying processes on the model of an internal reasoner, concluder, decider who methodically matches means to ends and then orders the specified action; we have seen in outline how a different sort of process could control speaking, and our other intentional actions as well.

Slowly but surely, we are shedding our bad habits of thought, and replacing them with other habits. The demise of the Central Meaner is more generally the demise of the Central Intender, but the Boss still lives on in other disguises. In chapter 10 we will encounter him in the roles of the Observer and Reporter, and will have to find other ways of thinking about what is going on, but first we must secure the foundations of our new habits of thought by tying them more closely to some scientific details.