DANIEL C. DENNETT

WHY YOU CAN'T MAKE A COMPUTER THAT FEELS PAIN

It has seemed important to many people to claim that computers cannot *in principle* duplicate various human feats, activities, happenings. Such aprioristic claims, we have learned, have an embarrassing history of subsequent falsification. Contrary to recently held opinion, for instance, computers can play superb checkers and good chess, can produce novel and unexpected proofs of nontrivial theorems, can conduct sophisticated conversations in ordinary if tightly circumscribed English. The materialist or computerphile who grounds an uncomplicated optimisim in this ungraceful retreat of the skeptics, however, is in danger of installing conceptual confusion in the worst place, in the foundations of his own ascendant view of the mind. The triumphs of Artificial Intelligence have been balanced by failures and false starts. Some have asked if there is a pattern to be discerned here. Keith Gunderson has pointed out that the successes have been with task-oriented, *sapient* features of mentality, the failures and false starts with *sentient* features of mentality, and has developed a distinction between program-receptive and program-resistant features of mentality.[1] Gunderson's point is not what some have hoped. Some have hoped he had found a fall-back position for them: *viz.*, maybe machines can *think* but they can't *feel.* His point is rather that the task of getting a machine to feel is a very different task from getting it to think; in particular it is not a task that invites solution simply by sophisticated innovations in *programming*, but rather, if at all, by devising new sorts of *hardware.* This goes some way to explaining the recalcitrance of mental features like pain to computer simulation, but not far enough. Since most of the discredited aprioristic thinking about the limitations of computers can be seen in retrospect to have stumbled over details, I propose to conduct a more detailed than usual philosophic thought experiment. Let us imagine setting out to prove the skeptic wrong about pain by actually writing a pain program, or designing a pain-feeling robot. I think the complications encountered will prove instructive.

The research strategy of computer simulation has often been misconstrued by philosophers. Contrary to the misapprehensions innocently engendered by

Synthese 38 (1978) 415–456.

Turing's classic paper, 'Computing Machinery and Intelligence,'[2] it is never to the point in computer simulation that one's model be *indistinguishable* from the modelled. Consider, for instance, a good computer simulation of a hurricane, as might be devised by meteorologists. One would not expect to get wet or wind-blown in its presence. That ludicrous expectation would be akin to use-mention error, like cowering before the word 'lion.' A good computer simulation of a hurricane is a program which, when you feed in *descriptions* of new meteorological conditions, gives you back *descriptions* of subsequent hurricane behavior. The descriptions might be in roughly ordinary English, dealing with clouds, waves and tides, or in some arbitrary notation, dealing with barometric pressure, wind velocities, and yet more esoteric (but measurable) features of hurricanes. The goal is to devise a program that will give you good 'predictions' of what a hurricane will do under a great variety of highly complex conditions. Such a program is tantamount to an immense conjunction of complicated conditionals: 'if conditions A, B, C, . . . obtain, then R will result; and if conditions D, E, F, . . . obtain, S will result; and . . . ' Obviously the only way to populate that conjunction reliably is by deriving the particular conditionals from general covering laws, all properly meshed and coordinated. So in order to write a good simulation program one must have a *theory* of hurricane behavior, and it must be a good theory. But if one must have a theory in the first place, why bother incorporating it into a program? There are several good reasons. First, the demands of program writing force into the open any incoherencies, gaps, or unanswered questions in a theory; it keeps the theoretician honest. Second, once a theory is thus incorporated into a working, 'debugged' program, its implications can be quickly determined and assessed. A simulation can be an 'experience-generator'; hurricanes are not that numerous, but a simulation program could generate thousands of different storm histories to scrutinize for implausibility or worse. Also, of course, such a program could be used in high-speed real time prediction of current weather. The fact that such a simulation program is ultimately only a high speed generator of the consequences that some theory assigns to various antecedent conditions is often obscured by the mode of presentation of the input and output. It is often useful, convenient, or just plain exciting to use the output to drive a visual display, a raster or TV screen on which appears, say, a swirling vortex moving up a map of the East Coast, but that swirling vortex is a sort of epiphenomenon, the tail that

doesn't wag the dog. The theory incorporated into the program *directs* the behavior of the presentation, and does not *read off* the behavior of the presentation, which itself plays no role in the simulation beyond its role as a convenient display.

Now let us consider a similarly inspired computer simulation of human pain. We write a program, based on our theory of pain, such that when we type in *descriptions* of conditions, e.g.,

> 'An anvil drops from a height of two feet onto *S*'s unanesthetized left foot,'

the computer types back *descriptions* of results, e.g.,

> '*S* jumps about on right foot, a tear in his eye, screaming.'

We test the program by varying how we fill in the blanks in our permissible input formulae (e.g., "*A* ____ is dropped from a height of ____ on *S*'s ____") and checking the resulting outputs for plausible variety and dependence on the input. What is unsatisfying about this computer simulation of pain? The skeptic might reply that it is a simulation at best only of pain *behavior*, but consider our hurricane simulation: what *else* is there to simulate but the hurricane's behavior? A better reply is that we have so far only attempted to simulate *external* pain behavior. This defect is easily remedied. Revised, our program will yield such outputs as

> '*S*'s C-fibres are stimulated, . . . a pain-memory is laid down, *S*'s attention is distracted; *S*'s heart-rate increases . . . *S* jumps about on right foot, a tear in the eye, screaming.'

(We can be sketchy, for the moment, about the internal 'behavior' or effects alluded to in the program.) Suppose, then, that we pack our output descriptions with neurophysiological description or even mentalistic psychological description about effects on memory, belief, desire, etc. Still, the skeptic may insist we have left something – indeed everything – of importance out. We have simulated, perhaps, the internal and external *causes and effects* of pain, but not the pain itself.[3] Some identity theorists may wish to retort to this that C-fibre stimulation just *is* the pain,[4] but we needn't take a stand on that point, since there are further ways of obliging the skeptic. We

can rewrite our program so it yields such outputs as

> 'There is a pain, *P*, of the in-the-left-foot variety, in *S*; *S*'s C-fibres are stimulated . . . '

Now we have explicitly included the pain. But, says the skeptic, the program still leaves out the *quality* of the pain. Very well. We expand our theory, and concomitantly our program, to yield detailed descriptions about even this. Again we feed in

> 'An anvil is dropped from a height of two feet on *S*'s left foot,'

and this time we get back:

> There is a pain, *P*, of the in-the-left-foot variety in *S*; *P* begins as a dull, scarcely noticeable pressure, and then commences to throb; *P* increases in intensity until it explodes into shimmering hot flashes of stabbing stilettoes of excruciating anguish [or words to that effect] . . . . ; *S*'s C-fibres are stimulated . . .

I see no reason why our program could not be enlarged to incorporate all this; the biggest problem would seem to be discovering sufficient uniformity and lawfulness in such 'phenomenological' effects as reported by sufferers to permit much prediction. Of course if the data we collect suggest a random distribution of these effects within certain boundaries that is easy enough to incorporate into our program as well.[5]

I do not expect this would satisfy the skeptic. He might try to express his doubts by pointing out that there is nothing pain-like going on in the computer when it churns out these reports. But of course not. Nor does the computer hurricane generate an internal low barometric pressure behind its steely facade. At this point is should dawn on the skeptic that he has been barking up the wrong tree. He has no pressing quarrel with *this* research strategy when it is directed to psychological phenomena, since its guiding presupposition is not that men are computers (any more than hurricanes are) but simply that one can have a rigorous *theory* of human psychology, materialist, dualist, epiphenomenalist, or whatever. Isn't there, however, another research strategy that differs significantly from the one we've been considering, where the aim of the computer is to *do*, not *describe*? For instance, 'Shakey' at Stanford Research Institute is a robot that can 'recognize' simple objects with its television eyes; it pushes cubes and

pyramids around in response to typed commands. Such 'performance models,' one might say, really do things; they do not so much incorporate theories (as do simulations) as *instantiate* theories.[6] The skeptic's challenge is now for us to make such a robot, a *feeler of pain*, not a mechanized theory about feelers of pain. So let us try to design such a robot. Of course our efforts in this task will be as much guided by our *theory* of pain as were our earlier simulation efforts, and we might ask the skeptic if he had any quarrels with our earlier, programmed theory *as a theory* of pain. If the skeptic makes no objections to it, or if we are able to revise our theory to satisfy his objections, we are home free, for it is a relatively straightforward task to build the robot with the help of our earlier 'describing' program. The describing program simply becomes the control system for our new robot.

Here is how it is done. Suppose our original program yielded outputs like '*S* trembles, a tear in his eye, and says "Ouch! My thumb hurts." ' First, we rewrite all outputs in the first person: 'I tremble, a tear in my eye, and I say "Ouch! My thumb hurts." ' Then we drop the redundant 'I say' wherever it occurs and move all direct quotation onto a separate 'protocol' terminal, which will then print only 'Ouch! My thumb hurts.' The rest of the output is reprogrammed to drive a display of flashing lights with labels. The 'tremble' light goes on, the 'tear in the eye' light, and so forth. Then we replace the input sentences in a similar manner. We make up magnetized plastic tokens representing different objects – anvils, knives, olives – falling from different heights, and we label an array of slots to accept these tokens: 'thumb,' 'big toe,' etc., so that *dropping* the anvil token into the thumb slot simulates dropping the anvil on the thumb. Of course that's not very realistic, but we can improve it easily. For instance, we can replace the 'tremble' light with an eccentric flywheel that makes the whole computer vibrate when it is turning; the tear in the eye problem has already been solved for us by the toy manufacturers, and the other details of verisimilitude are either obviously irrelevant or can be solved by the Disney studios given six months and enough Federal grant money. The result will be a robot that really *does* things; it trembles and reels and whimpers; it says just where the pain is; it attempts to duck falling objects – perhaps it even kicks back if we kick it.[7]

But what about the rest of our earlier simulation? What happens to the hot flashes and dull throbs mentioned in our descriptive progam? These parts of the output we transform into labeled flashing lights and leave them that way: sometimes the 'dull throb' light is on (blinking slowly if you like) and

sometimes the 'hot flash' light is on. If the skeptic insists on more verisimilitude here, what can he be asking for? Remember that these lights are not blinking randomly. The 'dull throb' light goes on only at appropriate times, the robot can then say 'there is a dull, throbbing pain' and the other apposite side effects of dull, throbbing pains are presumed to be arranged to coincide as well. But, the skeptic persists, no amount of *side effect* can turn what is not a dull, throbbing pain into a dull, throbbing pain, and, obviously, calling this event a dull, throbbing pain does not make it one either. This objection, for all its plausibility, is unfair as it stands. The skeptic, we must assume, had no objection to settling for an IBM typewriter as the 'speech' element in this robot, and surely typing is not talking – and calling typing talking would not make it talking. Since he has not challenged us to make a *bona fide* member of the species *homo sapiens* out of whatever bits and pieces are on the shelves at IBM, he must be permitting us to use some substitutes – the legs can be titianium, not flesh and bones – and since our flashing light (or whatever turns it on) has all the *functional* features he has demanded of pain, why is he now changing the game? Calling the robot a human being would not make the robot a human being either, but that was never set as a goal. It begins to appear that what the skeptic was after all along was not a *simulation* or an *analogue* of pain, but the synthesis of real pain, like the synthesis of urea by Wöhler in 1828 that marked the unification of organic and inorganic chemistry. The synthesis of real pain in a machine would tend to confirm that we human beings are just fancy soft machines, as the materialist contends.

That we might reconstrue our task in this way highlights a peculiarity in our ordinary concept of pain. The word 'pain' has both a *sortal* grammar ('I have a pain,' 'pains shooting up my arm') and a *mass noun* grammar ('there is more pain now,' 'it will cause you some pain'). The mass noun grammar often permits us – even invites us – to view pain as a sort of biological or psychological substance, rather than a process or event or activity or state. For instance, the amount of morphine that can be safely administered depends on the *amount of pain* it has to kill. For excruciating pain (e.g., that of coronary thrombosis) two to four times the usual therapeutic dose may be given without danger. But in cases of severe pains that can quickly and spontaneously disappear (e.g., those of coronary occlusion or biliary colic) such doses are dangerous, since if the pain disappears suddenly the patient may show signs of morphine

poisoning. If such were to happen, one would do well to punch or slap the patient, since, as Stearns observed in 1883, "pain is the antidote for morphine poisoning." One creates *more pain* for the morphine to neutralize, and thus prevents the excess of morphine from poisoning.[8] This suggests that specificity to morphine as an antagonist would be a legitimate test for any robot pain to pass.

This reconstrual of the task might seem, however, to harbor a conceptual confusion. Does one not contradict oneself in speaking of the synthesis of real pain? Synthetic urea *is* urea, as genuine as any to be found, but synthetic rubber is not rubber.[9] Is artificial intelligence genuine intelligence? Artificial coloring is perfectly genuine *coloring*, but artificial flowers are not flowers. The field of artificial intelligence trades on this ambiguity. Successes are often heralded as examples of genuine intelligence created by artifice, but in the face of objections this claim can be adjusted; artificial intelligence *works just as well as*, or is a useful and even theoretically interesting *substitute* for, genuine intelligence. I do not believe the term 'artificial intelligence' is objectionable on these grounds, for I do not believe in the distinction we are invited to make in this instance. Suppose the intelligence of some artifacts does function just as well as human intelligence (an immense supposition, of course); then, since intelligence, like respiration, is a purely functional notion, artificial intelligence, like artificial respiration, is no less genuine for being obtained by artifice. It may not be *just like* natural, human intelligence, but is genuine intelligence, as genuine as (we can imagine) the *alien* intelligence of extra-galactic creatures might be.[10] But what of artificial or synthetic pain? Is pain like rubber and flowers, or like coloring, respiration and intelligence? Whatever answer we might agree on (and agreement is both unlikely and ultimately unimportant), one lesson is clear: *if* pain is deemed to be *essentially* a biological phenomenon, *essentially* bound up with birth, death, the reproduction of species, and even (in the case of human pain) social interactions and interrelations, then the computer scientist attempting to synthesize real pain in a robot is on a fool's errand. He can no more succeed than a master cabinetmaker, with the finest tools and materials, can succeed in making, today, a *genuine* Hepplewhite chair.

Reservations about whether synthetic pain would be real pain may seem overly precious, but it is important to bring them into the open, for several reasons. First, a great deal of the counterintuitiveness of the notion of robot

pain no doubt derives from a dim appreciation of this side of our notion of pain. *Real* pain is bound up with the struggle to survive, with the real prospect of death, with the afflictions of our soft and fragile and warm flesh.[11] With our concept of pain, as with many others, there is a tug toward *parochiality: real* Chateau Latour has to have been made in a particular place, in a particular way, by particular people: an artificially concocted fluid indistinguishable to both chemists and connoisseurs from Chateau Latour would still not be *real* Chateau Latour. (Real vodka, on the other hand, can be made from just about anything, anywhere, by anybody.) The parochiality of the concept of pain, is, moreover, not an irrational feature, or at least not obviously so, for it has a role to play in defining our moral community. There can be no denying (though many have ignored it) that our concept of pain is inextricably bound up with (which may mean something less strong than *essentially connected with*) our ethical intuitions, our senses of suffering, obligation, and evil.[12] It will not do to suppose that an assessment of any attempt at robot synthesis of pain can be conducted independently of questions about what our moral obligations to this robot might be. One reason, then, why you can't make a computer that feels pain is that our concept of pain is not a pure psychological concept but also ethical, social, and parochial, so that whatever we *put inside* our computer or robot will not avail unless it brings in its train these other considerations, a matter over which our control, as computer designers, is worse than limited. This reason is important, and worth developing with more care, but not here, for it is also a bit of a red herring. Even if contextual matters, such as questions of origin and 'form of life', make a difference, they do not make enough of a difference. I do not think the skeptic wishes to rest his case at a point where the programmer's synthetic product might fall short only by these yardsticks (like the clever chemist's imitation Chateau Latour which only *seems* to have 'good breeding'). Moreover, were the synthetic product that good, the contextual matters might either fall into line (we would start treating the computer very much as one of us, and commiserate with it, comfort it, etc.) or be dislodged in our minds from their position of importance. In any event what the skeptic finds impossible to imagine is that this thing that happens in and to him (and it happens in and to him quite independently – or so it seems – of his biological origin, destiny, social milieu or ethical status) can be made to happen in and to a robot.

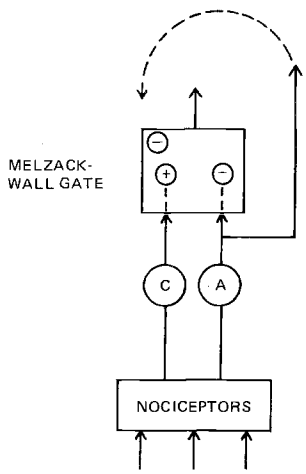
At this point it is easy for the skeptic to fall into extravagant and

irrelevant claims to support or flesh out his skepticism. When he says no robot could feel pain as he does, is it the artificiality, the chemistry, or what that makes the difference? A *cloned* duplicate of himself would presumably be capable of feeling pain, but if we could *construct* a biochemical duplicate of him, would this artifact be a painless robot? On what grounds, other than the grounds of origin we have just now set aside? Supposing, then, that a manufactured *biochemical* duplicate *would* feel pain, on his view, what difference could it make if we use other materials? Only two replies, both insupportable, occur to me: (1) organic compounds are capable of realizing functional structures with capacities of a sophistication or power *in principle* unrealizable in non-organic materials, or (2) though an inorganic replica might succeed in duplicating a human being's functional structure, the states in it functionally isomorphic to human pain states would fail to be genuine pain states because the biochemistry of pain state realizations is essential.[13] These are both highly implausible vitalistic claims, and any skeptic led to defend his view in this territory has simply been led astray. That is not to say that murmurs of vitalism do not make a large contribution to the skeptics' attitudes, but just that the contribution should be first isolated and then ignored. To find something better for the skeptic to say we must give him more details to work with. We have been assuming, up to now, that the programmer could have at his disposal a fairly satisfactory theory of pain to exploit in designing his robots. We have been assuming, that is, that the *mysteriousness* of pain might thwart our efforts at synthesizing pain without thwarting our efforts at theorizing about it. But how realistic is that assumption? The best way to examine it is to set down the bare bones of current physiological theory relating to pain, and list some of the attested pain phenomena, the data any acceptable theory of pain must account for, and see if there are any insuperable difficulties presented by them. We can record the known dependencies and interrelations among these phenomena by plotting a 'flow-chart' of sorts for a pain program.

The flow-chart presented here is merely a sketch, lacking rigor and detail, but its point is to facilitate philosophical scrutiny of pain, not to launch any serious project of programming or theoretical psychology or neurophysiology. I hope it will be clear that the difficulties we encounter would only be exacerbated in a more systematic model.

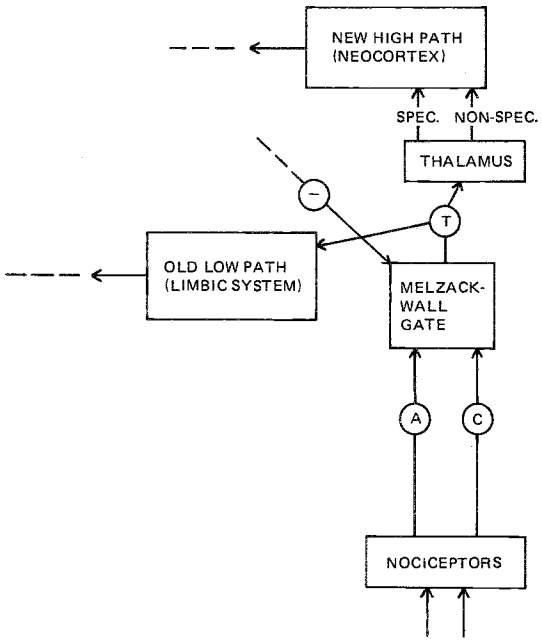
We can begin with what is known about the functional anatomy of

transmission from pain sites into the brain. As is generally the case, the further in from the periphery we move, the murkier the details. The journey begins at the site of injury, with receptors sometimes called nociceptors that respond with some degree of specificity to a variety of noxious events: mechanical distortion, intensities of heat and cold and chemical changes, for instance. The outputs of these receptors travel brain-ward through at least two very different types of fibres: swiftly through the large myelinated A-fibres, and slowly through the narrow, unmyelinated C-fibres.[14] Both signals arrive at the substantia gelatinosa, the midbrain gateway, where a complicated interaction takes place. A-fibres also send effects inwards via other channels. The A- and C-fibres seem to make two different functional contributions. On the one hand, it seems that the C-fibres are the preponderant transmitters of 'slow,' 'deep,' 'aching,' or 'visceral' pains, while A-fibres are implicated in 'sharp,' 'bright,' 'stabbing' pains. Recently Melzack and Wall have suggested a more interesting function for the A-fibres. They act at the substantia gelatinosa to *inhibit* the effect of the C-fibres, thus *closing* the gate to pain-impulse transmission, or at least damping the output of that gate. Moreover, the A-fibre channels that bypass the Melzack-Wall gate in the substantia gelatinosa seem to initiate more central activity that sends inhibitory signals back down to the gate, further blocking the transmission of impulses from the C-fibres. The capacity of the hypothesized Melzack-Wall



gate system to explain a variety of pain phenomena is immense, as we will soon see.[15] What, then, happens to the output of the gate, the so-called T-cell transmissions? In broadest outline we can say that once again there is a split into two channels. One channel carries through the lower, phylogenetically older portion of the brain, the limbic system (hypothalamus, reticular formation, paleocortex, hippocampus), and the other passes through the thalamus and is projected onto the higher, phylogenetically newer, characteristically human part of the brain, the neocortex. Let us simplify by calling these the old low path and the new high path. The new high path is subject to yet another bifurcation: there is both a specific and a non-specific projection of fibres from the thalamus onto the cortex.

The new high path, which is relatively undeveloped or non-existent in lower animals, subserves fine-grained perception: location and characterization of pain and other stimuli. It is here that pattern recognition, depth perception, and most of the other sophisticated operations of perceptual analysis are completed. The old low path is characterized by orthodoxy as the *aversive*



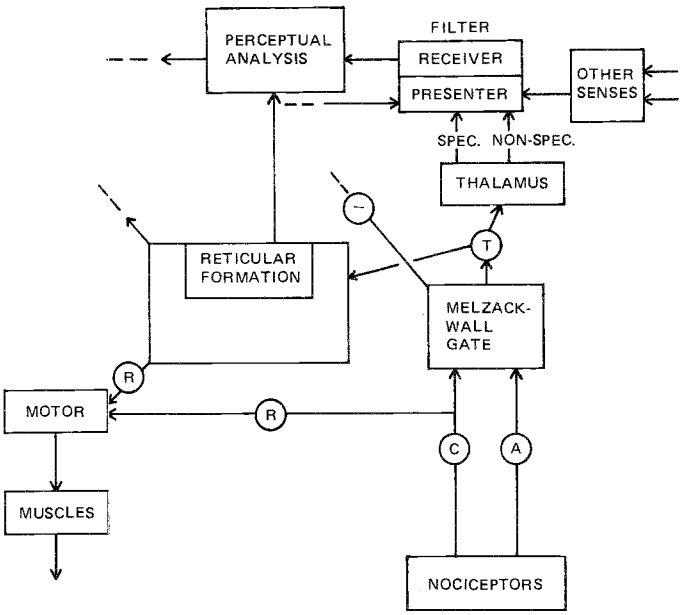
system, the 'motivational-affective processing' system. Orthodoxy is well buttressed by evidence in this instance,[16] and this suggested separation of the hurtfulness or awfulness of pain from its other characteristics – to speak loosely – will loom large in our further discussion.

Having charted this far with the aid of anatomical roadmaps, we have reached *terra incognita* and if we are to proceed with our flow chart we must abandon the pretence that our boxes represent anatomically salient structures, and proceed with a purely functional, abstract breakdown of the system. We can make use of the freedom thus provided to be more boldly speculative, and also, inevitably, vaguer about the nature of functions and relations we are charting. The only constraint on our design will be that it accommodate the known and presumed phenomena.

Everyone knows, for instance, that distracting one's attention (e.g., by going to a movie) diminishes or banishes pain. This can be easily provided for if we build in a presenter-receiver filtering system across the pathway for incoming signals from all the sense modalities, subject to the following conditions: the receiver can have its general sensitivity raised or lowered, and the presenter has *selective* volume controls, so that its various signals can be turned up independently.[17] Then the effect of distracted attention could work this way: paying special attention to one input (the visual and auditory input from the movie) would be accomplished in part by turning up its volume in the presenter. Then the receiver would compensate for this high volume by decreasing its sensitivity, having the effect of muffling everything else, including the pain signals. The same effect might be accomplished by the Melzack-Wall gate, but let's be generous and draw in a separate filtering system.

Pain signals trigger a variety of 'spinal reflexes' causing relatively simple but swift muscular responses without the intervention of higher brain centers (and in some instances without passing through the substantia gelatinosa, the Melzack-Wall gate), and since distracted attention has little or no effect on these, we will put the filter only in the new high path, and draw in the reflex links ('R') to the motor output nerves without intervening links of sophisticated control.

There are many transactions between the old low and new high paths. Of particular importance to us is the relation the reticular formation in the old low brain has to higher centers. The reticular activating system plays a major



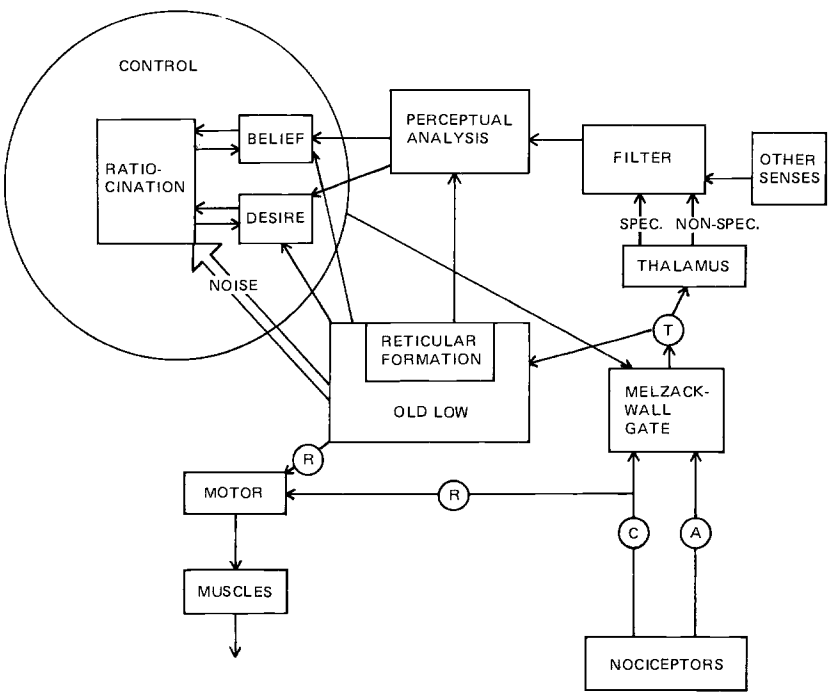
role in governing sleep and waking, and determining the level of arousal generally; it also plays a role in directing attention, and thus can be considered a versatile alarm system.

So far we have hardly touched on the effect of pain stimuli on 'higher centers,' so let us sketch in roughly what is most obvious about these effects. When we have a pain we believe we have a pain (at least normally) and pains can be remembered, usually, for some time. So in our control circle we will place a memory and belief box, and to be on the safe side, *two* arrows leading to it, one from the old low path and one from the new high path (further investigation might lead us to revise any of this, of course). Also, pains are abhorrent, at least usually. That is, the infliction of pain is a reliable behavior-modifier, tongue loosener, punishment. (Whether punishment is good for anything is another matter. Pain is a good way to punish.) So we should draw in a 'goals and desires' box, with appropriate inputs. (If the 'aversive' effects of pain are subserved *entirely* by the old low path, we might not want an arrow from the new high path to the desire center, but again,

let's be generous. No doubt even the most intellectual apprehension of pain stimuli could have *some* effect on one's desires, current or long-term.)

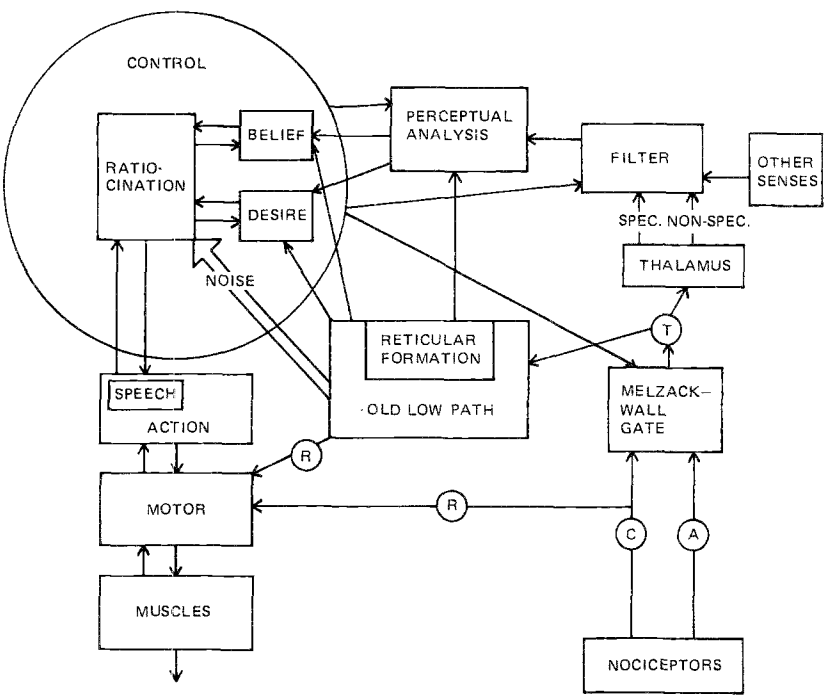
It is a useful and oft-used myth, at least, that higher controls in human beings are accomplished by something like logical processing of the material in the belief and desire boxes,[18] so let us distinguish a ratiocination unit in the control area. We need this in any case, since one effect of pain stimuli on this function (as captured by our box) is not *informational* but *noisy:* pains interfere with our ability to concentrate, to solve problems, to think clearly so we should draw a 'noise' arrow to the ratiocination unit. (Perhaps we should draw noise arrows to other units as well, but let's not overcomplicate our diagram.)

Finally, let us transmit the control center's effects on behavior through an action organizing unit (with the specially important speech synthesis unit drawn in) to the motor-effector controls and thence to the muscles. In addition to the control center's effects on behavior, we must put in the



arrows for various 'descending effects' on the input system, including those already alluded to: an inhibitory effect on the Melzack-Wall gate, a volume control on the filter, a 'perceptual set' or 'readiness' determiner to weight the analyzing machinery, and others not worth drawing in. Then of course there should be 'feedback' arrows throughout. That should be enough detail to handle the phenomena at least in outline. The sketchiness and idealization of this model should not be forgotten, of course. Some of the functions captured by these boxes may merge in single anatomical structures, and such distortions as are present in the model might seriously misrepresent the actual state of affairs to the point of requiring major revision of the model. In any case, however, we now have *a* fairly complicated and versatile model to play with; let us see how it runs when it comes to providing for the variety of pain phenomena.

Why does it help to rub or press the painful area, or to scratch an itch? Melzack and Wall claim that this increases A-fibre transmission, thus



increasing the inhibition of the C-fibre stimulation in the substantia gelatinosa. A less commonly recognized home remedy for pain is – not to *distract* but – to *concentrate* one's attention on the pain. I discovered this for myself in the dentist's chair, thinking to take advantage of the occasion by performing a phenomenological investigation without the benefit of novacain, and have since learned that this is a highly elaborated technique of Zen Buddhism.[19] I recommend this enthusiastically. If you can make yourself study your pains (even quite intense pains) you will find, as it were, no room left to *mind* them (they stop hurting) – though studying a pain (e.g., a headache) gets boring pretty fast, and as soon as you stop studying them, they come back and hurt, which, oddly enough, is sometimes less boring than being bored by them and so, to some degree, preferable. I am not at all sure that what I just said 'makes sense;' that is, I am not at all sure that this loose talk about *pains* that cease and resume hurting, or even this loose talk about *studying* one's pains, is ontologically, metaphysically, scientifically, phenomenologically *sound*, but it is nevertheless just what I want to say. That is the way I would put it if I were unself-conscious and unworried about committing some conceptual gaffe. And that is a crucial part of the problem of pain: we have a baffling variety of such untutored, unstudied accounts of pain phenomena, and it is no easier to give them all credence than it is to revise them by the lights of some procrustean theory about what pain experiences must be. But, to return to the effect of attention on pains, whatever the 'correct' philosophical analysis is of the variety of first person pain reports, it must have room for the fact that focussing attention can *obtain relief* (to put the matter neutrally for the moment). Melzack and Wall have a ready explanation of this phenomenon: focussing attention on one's pains may serve to raise the volume only on the A-fibre component of it, thus inhibiting the C-fibres at the Melzack-Wall gate. Their experiments tend to confirm this hypothesis, and suggest that analgesia by hypnosis or yoga methods has a similar explanation.[20]

We 'locate' our pains, but this is not a single thing we do. On the one hand, we react more or less reflexively to jerk the injured part away from contact with harm, and that is accomplished through the old low path. But we also 'can say' where our pains are, and this is presumably accomplished on the new high path with the aid of the specific projection system to the perceptual analysis areas of the cortex and thence through the control system to

ultimate speech. Excitation of a specific peripheral nerve fibre at any point on its length normally produces a sensation of pain felt at its normal extremity. 'Phantom limb' is, of course, the most vivid manifestation of this phenomenon.[21]

In 'referred pain' the pain location does not match the location of the trauma. This must be due to 'leakage' or 'short-circuits' at crossover points in the specific pathways, probably in the substantia gelatinosa. Oddly enough, however, administering novocain or other local anesthetic to the site where the pain is *felt* diminishes the referred pain, and pressure on that area increases the pain. This can be accounted for if we suppose the leakage is not a simple *turning-on* of the wrong fibre, but an *enhancement* of a resting level of transmission. Under local anesthesia there would be nothing to enhance (since local anesthetics stop *all* transmission) and pressing the uninjured area would produce a higher pre-existing level to enhance.

Now let us locate on our model the effects of various drugs, especially the anesthetics, which prevent all sensation, and the analgesics, which are specific for pain. Novocain and related local anesthetics act by completely stopping the transmission of nerve cells at their source. (In fact, they block nerve activity wherever they are injected. A 'spinal block' is a local anesthetic administered high in the pathway to the brain, creating a wide area of total but still 'local' anesthesia.) There are no local analgesics, but aspirin is unique among the common general analgesics in having a peripheral site of action.[22] It antagonizes a metabolite, bradykinin, at the nociceptors; it is bradykinin that persists in stimulating the nociceptors after the initial traumatic event and thereby is responsible for persistent pain. Aspirin by antagonizing bradykinin thus prevents pain at the earliest opportunity. This is interesting because aspirin is also unique among analgesics in lacking the 'reactive disassociation' effect. All other analgesics (e.g., the morphine group and nitrous oxide in sub-anesthetic doses) have a common 'phenomenology.' After receiving the analgesic subjects commonly report not that the pain has disappeared or diminished (as with aspirin) but that the pain *is as intense as ever* though they no longer *mind* it. To many philosophers this may sound like some sort of conceptual incoherency or contradiction, or at least indicate a failure on the part of the subjects to *draw enough distinctions*, but such philosophical suspicions, which we will examine more closely later, must be voiced in the face of the normality of such first-person reports and the fact

that they are expressed in the widest variety of language by subjects of every degree of sophistication. A further curiosity about morphine is that if it is administered *before* the onset of pain (for instance, as a pre-surgical medication) the subjects claim not to feel any pain subsequently (though they are not *numb* or anesthetized – they have sensation in the relevant parts of their bodies); while if the morphine is administered *after* the pain has commenced, the subjects report that the pain continues (and continues to be *pain*), though they no longer mind it.

Our model suggests that morphine and other analgesics must work on the old low path while leaving the new high path relatively in order, and such is the case. While morphine, like anesthetic drugs generally, takes effect first at the higher, cortical levels of the brain and then descends to the old brain, the specific projections to the cortex are especially resistant to damping by drugs, so that the effects of these drugs is more pronounced on the old low aversive path than on the new high path of fine-grained perception. The timing-dependence feature of morphine might be explained this way: once old low pain signals have contributed to the 'set' of the perceptual analyzing machinery (via influences on the control center's 'descending effects' which would weight the interpretation machinery in favor of interpreting particular signal patterns as pain-transmitting), this cannot be quickly undone, even after the contribution from the old low path is eliminated by morphine. Lobotomized subjects similarly report feeling intense pain but not minding it, and in other ways the manifestations of lobotomy and morphine are similar enough to lead some researchers to describe the action of morphine (and some barbiturates) as "reversible pharmacological leucotomy [lobotomy]".[23]

When we turn from local anesthesia and analgesia in conscious subjects to general anesthesia, the situation becomes more complicated. The major problem can be approached by way of a curious and terrible incident from the annals of medicine. Curare, the poison used by South American Indians on their blow-pipe darts, was purified (as *d-tubocurarine*) and introduced into medical research in the 1930's, and its action was soon well understood.[24] It is a paralytic that acts directly on all the neuromuscular junctions, the last rank effectors of the nervous system, to produce total paralysis and limpness of all the voluntary muscles. It has no central effect except for a slight enhancement effect on activity in the cortex. In the 1940's, however, some

doctors fell under the misapprehension that curare was a general anesthetic, and they administered it as such for major surgery. The patients were, of course, quiet under the knife, and made not the slightest frown, twitch or moan, but when the effects of the curare wore off, complained bitterly of having been completely conscious and in excruciating pain, feeling every scalpel stroke but simply paralyzed and unable to convey their distress. The doctors did not believe them.[25] Eventually a doctor bravely submitted to an elaborate and ingenious test under curare, and his detailed confirmation of the subjects' reports was believed by his colleagues: curare is very definitely not any sort of anesthetic or analgesic.[26]

Recently a puzzle occurred to me: suppose that one were to add to curare a smidgin of *amnestic*, a drug that (we will hypothesize) has no effect on experience or memory during *n* hours after ingestion but thereafter wipes out all memory of those *n* hours.[27] Patients administered our compound, curare-cum-amnestic, will not later embarrass their physicians with recountings of agony, and will in fact be unable to tell in retrospect from their own experience that they were not administered a general anesthetic. Of course *during* the operation they would know, but would be unable to tell us.[28] At least most of our intuitions tell us that curare-cum-amnestic would not be an acceptable substitute for general anesthesia, even if it were cheaper and safer.[29] But now how do we know that general anesthetics in use today are not really curare-cum-amnestic? We know, in fact, that curare *is* routinely used in general anesthesia today. Most general anesthetics by themselves in safe doses do not entirely block reflex spasms, so curare or another curariform paralytic is administered to prevent muscle-tearing damage and thrashing about that could interfere with the surgeon's task. Moreover, a variety of drugs in the anesthesiologist's bag are known to be amnestics (see note 27). How do we know that these drugs have the further effect of producing genuine anesthesia or even analgesia? Absence of complaint or other behavioral manifestation, we have seen, is hardly sufficient grounds – though they are routinely and not unreasonably relied on in daily medical practice. To answer this question we will have to look more closely at the *internal* effects of the so-called general anesthetics, and at other, more indirect clues about their functions.

There are a wide variety of general anesthetics, but they fall into groups, and if we take three important drugs as paradigms, we will have more than

enough variation to suggest the problems: (1) nitrous oxide, or laughing gas, which is inhaled; (2) ether, one of many related volatile inhalants; (3) sodium pentothal (or thiopental sodium), an injected 'ultra-fast-acting' barbiturate.[30] These drugs are chemically very different, and have different effects on the central nervous system and the rest of the body. Moreover, in modern practice they are seldom used alone, but almost always are accompanied by pre-anesthetic medication, such as an analgesic (e.g., morphine), a sedative to combat anxiety and the nausea that often results from ether inhalation, or even a 'basal' anesthetic, which produces anesthesia sufficiently deep for preparation for surgery but too shallow for surgery. In spite of this variation we can impose some order by considering the traditional 'stages' and 'planes' of general anesthesia. In passing from full consciousness into the anesthetized state one moves through three of four marked stages. In the first, one is conscious but apt to have hallucinations or uncontrollable thoughts; the drug is acting on the neocortex (at least partly as an enhancer or stimulant). In the second or delirium stage one is unconscious (in some sense) but may laugh, shout, swear or thrash about. The drug's effects are descending through the brain, and one hypothesis is that the drug has reached the higher motor control centers and paralyzed them, 'releasing' lower motor activity. In the third stage, called surgical anesthesia, there are four planes, of increasing depth. Depending on the surgery to be done one will be kept in the shallowest permissible plane of surgical anesthesia, since the fourth stage, medullary paralysis, is a short prelude to respiratory failure and death. These temporal stages are all but undetectable with sodium pentothal, where stage three is reached in a few seconds, and their manifestations are largely obliterated by the effects of preanesthetic medication with ether or nitrous oxide (no one wants a hallucinating, thrashing patient to deal with, which is one reason for pre-anesthetic medications). So the importance for practice, if not pedagogy, of the traditional stages of anesthetic induction is virtually nil. The four planes of third-stage surgical anesthesia, however, have well-recognized symptoms relied on by anesthesiologists in maintaining the proper level of anesthesia during surgery. And for all the differences among the drugs, one similarity is clear enough: in doses large enough to produce deep plane surgical anesthesia (or fourth stage medullary paralysis) all the drugs are analgesic and anesthetic if any drug could be, since their effect at those levels amounts to virtual shut-down of the entire central nervous system. Such

barely reversible brain death will look plausibly pain-free (by being everything-free) to even the unrepentant interactionist, for there is almost nothing happening to interact with. This is small comfort to the skeptic, however, since because of their very danger such deep levels of anesthesia are shunned. In fact the direction of anesthetic practice is toward ever shallower, safer, more manageable anesthesia, with supplementary medication, such as curare, analgesics, sedatives and – yes – amnestics taking care of any loose ends uncontrolled by the shallow anesthetic.[31]

The persistence of reflex responses to painful stimuli under anesthesia is an obtrusive and unsettling fact, in need of disarming.[32] Goodman and Gilman observe that at the second stage of anesthesia "painful procedures are dangerous because the response to such stimulations (including incidental dreams[!]) is exaggerated and may lead to violent physical activity," (p. 32), and they note further that even at surgical levels, barbiturate anesthetics "do not adequately obtund the reflex responses to impulses which, in the conscious state, would be experienced as painful." (p. 127). Yet they assure us that analgesia in these circumstances is complete despite the occurrence of 'behavior' that is held – by some schools of thought – to be well nigh 'criterial' for pain. The presence of the reflexes shows that the paths between nociceptors and muscles are not all shut down. What special feature is absent from those paths whose presence is required for the occurrence of pain? The short answer routinely given is: consciousness. General anesthetics render one unconscious, and when one is unconscious one cannot feel pain, no matter how one's body may jerk about. What could be more obvious? But this short answer has the smell of a begged question. The principle appealed to (that consciousness is a necessary condition for feeling pain) does not have the status of a well-confirmed empirical hypothesis, or a 'law of nature,' and its utility evaporates if we try to construe it as an 'analytic truth.' Until an analysis is given of the relatively gross, molar notions of consciousness and pain, the principle has no particular warrant, save what it derives from its privileged position as one of the experience-organizing, pretheoretically received truths of our common lore, and in that unsystematic context it is beyond testing. Until we have a theoretical account of consciousness, for instance, how are we to tell unconsciousness from strange forms of paralysis, and how are we to tell consciousness from zombie-like states of unconscious activity and reactivity? The paradigms of unconsciousness that anchor our

acceptance of this home truth principle are insufficiently understood to permit us to make the distinctions we need to make in this instance.

I think it is fair to say that until very recently anesthesiologists had no better defense for their defining professional claim than such an appeal to 'intuitive' principle:

> How do you prevent pain?
>
> We give people drugs that render them unconscious.
>
> And how do you know they are really unconscious?
>
> Try to wake them up; you'll see. (Besides, when they do wake up, they don't recall any pain.)

Today, fortunately, better answers can be given; answers that at least have the potential to form the framework of detailed and confirmable theories. The 'state of unconsciousness' produced by general anesthetics can be independently characterized, and its importance accounted for. Drugs that cause sleep or deeper levels of 'unconsciousness' are called *hypnotics*, and all general anesthetics are hypnotics. Moreover, they all achieve this effect by antagonizing – though in different ways – the normal functioning of the reticular formation, preventing (*inter alia*) the arousal of the neocortex.[33] This shutting down of the reticular formation does not 'turn off' the cortex nor does it prevent stimuli from reaching it. It prevents or depresses 'recruitment' by those stimuli; they arrive at the cortex, but do not produce the normal spreading ripple of effects; they die out. On any plausible account of cortical functioning this should prevent the completion of the process of perceptual analysis. We could of course claim, with the support of orthodoxy, that such an effect on the cortex 'produces unconsciousness' and we could then 'explain' the absence of pain in such circumstances by an appeal to our common sense principle that consciousness is a necessary condition for pain, but that would require us to explain just how and why failure of cortical recruitment amounts to or causes unconsciousness, which is a step in the wrong direction, a step away from detailed functional analysis toward the haven of vague and unsystematized preconception. The hypothesis that the successful completion of a process of perceptual analysis is a critical feature in our functional account of pain is, in contrast, a generator of a variety of plausible accounts of perplexing phenomena. We have already seen its utility in accounting for the morphine time-dependence phenomenon. It could also

be invoked to account for the relation between the amnestic and anesthetic properties of some drugs. Brazier suggests that anesthesia may result from a derangement of some *memory* functions subserved by the hippocampus, producing a sort of continuous amnesia of the specious present. Such a 'forgetting' of each passing moment would cause a complete disability of perceptual analysis and ultimate recognition, and, so goes the theory, a pain not recognized is no pain at all.[34]

Another application of the hypothesis accounts for the striking fact that soldiers who have been wounded in battle often exhibit no discomfort from their serious injuries while awaiting treatment in the safety of a field hospital, but will complain bitterly of the pain of a clumsy venipuncture when a blood sample is taken.[35] They are in a state of specific, not general analgesia, and the specificity is relative not even to bodily location, but to the *import* of the stimulation. This capacity for import-sensitive analgesia has been exploited rather systematically by the Lamaze natural childbirth technique. Adherents of the Lamaze method claim that by giving the mother a meaningful task to perform, the input which would otherwise be perceived as pain is endowed with a complex action-directing significance; since the patient is not merely a passive or helpless recipient of this input, but rather an *interested* recipient, a *user* of the input, it is not perceived as pain, and again, since a pain not recognized as such is no pain at all, the Lamaze method actually promotes a *reduction of pain* in childbirth.

The content-sensitivity of some forms of analgesia and the time-dependence of morphine's analgesic effect can only be explained by a theory that treats the experience of pain as somehow the outcome of a process of perceptual analysis. Then, once that process is grossly located (in the neocortex), we can determine a necessary condition for its successful completion (reticular formation arousal), and can provide some grounds for the belief we are loath to abandon: general anesthetics are not misnamed. They are not misnamed because they prevent the completion of a process that is empirically established as a normally necessary condition for pain. This invocation of perceptual analysis restores the new high path in the cortex to a position of importance in our account and suggests that activity in the old low path is important not because it *is* or *amounts to* pain, but because it is a major contributing condition of pain.[36]

This completes the survey of pain phenomena, and carries our functional,

partly anatomical, flow-chart of pain as far as it can profitably be carried. The point of this extended exercise in speculative psychophysiology has been to flesh out a theory sketch to the point where one can plausibly claim to have an account that accommodates the data in all their variety.

Now we can return to the philosophical question that motivated the exercise: is the resulting theory a theory of pain at all; does it capture pain so that any realization of the flow chart would properly be said to be capable of experiencing genuine pain?

A related, but somewhat different question is this: can we locate pain, as distinct from its typical causes and effects, on the flow chart? The flow chart gives us a functional description at what I have called the sub-personal level.[37] I have labelled the various boxes 'belief,' 'desire,' 'action' and so forth, but that was taking a liberty. The flow-chart deals directly not with a person's acts, beliefs, thoughts, feelings, but with the behind-the-scenes machinery that governs speech dispositions, motor subroutines, information storage and retrieval, and the like. It has been convenient to talk as if the *objects* of our attention, what we pay attention to, were impulse trains in the nervous system, to talk as if the muffled outputs from the filter *were* the diminished pains, to talk as if we recognize or fail to recognize a neural signal *as* a pain, but this is loose talk, and the conceptual confusions it invites are not inconsequential. When we retell the subpersonal story without taking these liberties we seem to be leaving something out.

Suppose we want to know how an anesthetic about to be administered to us works. The doctor tells us that it prevents mechanisms in the brain from 'interpreting' certain impulse trains arriving from the periphery. This, he says, in turn prevents the initiation of motor activity, blocks normal effects on long and short term information storage and goal structures, and . . . permits surgery to proceed at safe levels of respiration and blood pressure. Yes, we reply, but does it stop the pain? If we are unsatisfied with the answer he has already given us, his further reassurance that of course the anesthetic does stop the pain is not yet another consequence of any theory of anesthesia he knows so much as a 'philosophical' dogma – quite reasonable, no doubt – that plays a useful role in his bedside manner. The sub-personal theory he relies upon, and perhaps helps to confirm or advance, can provide for the phenomena, it seems, while remaining neutral about the 'philosophical'

puzzles about pain. For instance, not only can it account for the effect of novocain and curare, it also can account for the presence of the 'reactive dissassociation' effect of morphine without taking a stand on whether the effect is properly described as the presence of pain in the absence of aversion or as the absence of pain in the presence of peculiar beliefs or speech dispositions. It can explain the placebo effect without settling the question: does placebo administration promote a belief that *causes* or *amounts to* the absence of pain? It can explain the success of the Lamaze method without committing itself to an account of what the success consists in: is it correct to say that the technique turns pains into painless sensations, or should we say it prevents certain pains from ever occurring at all? It can explain why concentrating on one's pain provides relief without settling the question of whether such concentration changes the object of attention, and if so, whether the object is so changed it is no longer a pain, or rather a pain one does not mind having, a pain that doesn't hurt.

The sub-personal account can provide at least a sketchy suggestion of why hypnosis is sometimes an effective method of obtaining relief, but what, exactly, does hypnosis accomplish? Does it manage to prevent the pain that would otherwise occur from occurring, does it prevent its existence, or does it simply permit the subject to ignore or abide the pain? Or does it leave the subject in pain but make him *think* or *act as if* he were not? Can it possibly be that these are different ways of saying the same thing? Suppose someone is given the posthypnotic suggestion that upon awakening he will *have* a pain in his wrist. If the hypnosis works, is it a case of pain, hypnotically induced, or merely a case of a person who has been induced to *believe* he has a pain? If one answers that the hypnosis has induced real pain, suppose the post-hypnotic suggestion had been: on awakening you will *believe* you have a pain in the wrist. If this suggestion works is the circumstance just like the previous one? (Isn't believing you are in pain tantamount to being in pain?) Or doesn't hypnosis induce beliefs at all? Is it rather that in both cases the subject just acts as if (1) he were in pain, (2) he believed he was in pain? What is presumably true in any case is that the effect of the hypnosis was to distort or weight the perceptual analysis machinery so that it produced a certain output, the sort of output that normally produces all or most of the normal pain dispositions: dispositions to avow, dispositions to nurse the wrist, take aspirin, and perhaps even dispositions to respond to stimulation of the wrist

with the classic 'spinal reflexes' (I do not know how deep hypnosis can reach – hypnotically induced 'pain' does not evoke the characteristic palmar skin resistance of pain, but may otherwise be indistinguishable). Even if we knew exactly which of the boxes in our flow-chart were affected by hypnosis, and how, we would not thereby have answers to our philosophical questions (except in the extreme cases: if hypnosis were to produce only a disposition to *say* 'I have a pain in my wrist' and no other manifestations of pain, or alternatively, if hypnosis produced an observable injury, with swelling, inflammation, bradykinin, etc., in the wrist, we would find easy unanimity in our answers).

The philosophic questions do not seem idle, but our sub-personal theory does not – at least not yet – provide leverage for answering them. The silence of the sub-personal account here is due simply to the fact that pain itself does not appear on our flow chart, which seems to concern itself solely with the *causes and effects* of pain.[38] As we trace through the chart, we find that causal contributions include nociceptor and C-fibre stimulation, T-cell activity, the processes of perceptual analysis and the contributions thereto of old low path activity; and among the effects we find muscle contraction, avoidance reactions, reports, beliefs, disruptive effects on thinking or reasoning, and powerful goal modifications. The absence of a 'pain' box might seem to be a simple omission, easily corrected. The most plausible place to insert a pain box is between the perceptual analysis box and the higher control centers. Isn't pain the *result* of perceptual analysis and the *cause* of our reactions to discomfort? Let us call the inserted box the *pain center.* Now what does it do? If one claims its function is simply to serve as the locus for the transmissions just mentioned, the go-between, then contrary to our suspicion, pain was already represented in our model; we simply had not drawn a line around it. If the point is rather that there is a separable and terrible something we had hitherto left out, how could we possible add it with this box?

How do we get pain into the pain center? Here is a suggestion: there are two little men in the pain center, and when the light goes on one starts beating the other with chains. What is wrong with this? Not that we have introduced homunculi, for there are (somewhat less colorful) homunculi inhabiting all these boxes. That is a legitimate and useful way to comprehend flow-charting.[39] What is wrong is that even if there were pain in the box, it

would not be the person's pain, but the little man's. And, to be crass about it, who cares if the little men in one's brain are in pain? What matters is whether *I* am in pain.[40]

There is no way of adding a pain center to the sub-personal level without committing flagrant category mistakes, by confusing the personal and sub-personal levels of explanation.[41] We might toy with the idea that our pain center, somewhat like Descartes' notorious pineal gland, is the producer of epiphenomena, the *echt* pains that make all the difference (without of course making any *causal* difference). The standard rebuttal to this version of epiphenomenalism should suffice. Suppose there were a person of whom our sub-personal account (or a similar one) *without the pain center* were true. What are we to make of the supposition that she does not experience pain, because the sub-personal theory she instantiates does not provide for it? First we can make the behaviorist's point that it will be hard to pick her out of a crowd, for her pain behavior will be indistinguishable from that of normal people. But also, it appears *she* will not know the difference, for after all, under normally painful circumstances she believes she is in pain, she finds she is not immune to torture, she gladly takes aspirin and tells us, in one way or another, of the relief it provides. I would not want to take on the task of telling her how fortunate she was to be lacking the *je ne sais quoi* that constituted real pain.

But that is a tendentious description of the case. Let us consider instead the hypothesis suggested by it. *viz.*, that we have simply not seen the woods for the trees, that pain is not to be found in any one box of our flow-chart, but is a function or combination somehow of the elements already present. What function? The chief value of all this somewhat science-fictional flow-charting and compiling of odd phenomena – the reason I have spent so much time on it – is that it serves to drive wedges of contingency between features that are often thought to be conceptually inseparable, simply because they are usually coincident. What I am asserting is that the arrows on the flow-chart are the arrows of normal causal relation, and wherever we have seen fit to posit a particular relation or dependency, we can imagine a severing of the normal connections responsible for it. Some of this fragmentation has familiar manifestations, some is to be found only rarely, and some never occurs, so far as I know, though we can conceive of it occurring.

We can locate our pains, for instance, but this is a complex ability of ours that could become discomposed on occasion. Anscombe considers such a case:

> you say that your foot, not your hand, is very sore, but it is your hand you nurse, and you have no fear of or objection to an inconsiderate handling of your foot, and yet you point to your foot as the sore part: and so on. But here we should say that it was difficult to guess what you could mean.[42]

Pains are also goal modifiers, but they might not be. That is, we can imagine a person who says he is in pain, locates the pain consistently, is in fact being beaten, writhes, cries, trembles, *but is immune to torture.* Is this really imaginable? Of course it is. Perhaps that is what masochists are. Or perhaps they have, as it were, a sign reversed going to the goal box, so they seek out pain instead of avoiding it, at least in certain circumstances.[43]

Pains are abhorrent, but what are we to make of the reports of subjects who are lobotomized or under morphine analgesia, who report pains, rank them in terms of greater and less intensity, but seem and claim not to *mind* the pains? Are they confused? They say they are in pain, but could they properly be said to believe they were in pain? It is not as if they are speaking parrot-fashion, nor do they exhibit massive conceptual confusions in other areas, so why can it not be that they do believe they are in pain? The only strong presumption against granting them this belief is that a good many 'theories' of pain make us 'incorrigible' or 'privileged' about our pains, and this is often characterized by the stipulation that belief that one is in pain is a sufficient condition for being in pain. If we hold this view of incorrigibility and grant these subjects their belief, then they are in pain, but then pain is not always abhorrent, even when the subjects are experiencing, as they sometimes believe, very intense pain. One might try to claim that such people reveal by their very odd behavior that they do not understand the word 'pain,' but that would be hard to support. Before the lobotomy or morphine administration, we can presume, they had a good command of English, including the word 'pain,' and there is no evidence, I think, to show that any of these treatments tends to produce lexical amnesia or other verbal confusions.[44] To be sure, they do not understand the word 'pain' the way some theories would say they ought to, but to bow to these theories would be to beg the question in the very description of the case.

The ordinary use of the word 'pain' exhibits incoherencies great and small.

A textbook announces that nitrous oxide renders one 'insensible to pain,' a perfectly ordinary turn of phrase which elicits no 'deviancy' startle in the acutest ear, but it suggests that nitrous oxide doesn't prevent the occurrence of pain at all, but merely makes one insensible to it when it does occur (as one can be rendered insensible to the occurrence of flashing lights by a good blindfold). Yet the same book classifies nitrous oxide among analgesics, that is, *preventers* of pain (one might say 'painkillers') and we do not bat an eye. Similarly, if 'pain is the antidote to morphine poisoning' then morphine cannot be said to prevent pain from occurring. Perhaps what the maxim really means is that *normally painful stimulation* is the antidote for morphine poisoning, but if that is what it means, that is not what it says, and what it says is easily understood, and understood to be good English. This particular slackness in our ordinary use has provided a playground for interminable philosophic disputation over the issue: can there be unfelt pains? I suggest that our flow-chart handles this traditional question by discrediting it. There can be, in principle, any combination of the normal 'causes and effects' of pain in the absence of any others, and intuitions will no doubt clash about which words to use to describe the results. Other philosophical questions about pain might have more interesting answers.

Consider the commonplaces about differences in 'pain-threshold'. Some people, it is often claimed, can stand more pain than others: they have a *high* pain threshold.[45] Suppose I am one of those with a *low* threshold, and undergo treatment (drugs, hypnosis, or whatever) supposed to change this. Afterwards I report it was a complete success. Here is what I say:

(1) The treatment worked: the pain of having a tooth drilled is as intense as ever, only now I can stand it easily.

Or I might say something different. I might say:

(2) The treatment worked: having a tooth drilled no longer hurts as much; the pain is much less severe.

Can we distinguish these claims? Of course. They obviously mean very different things. Can I then know which claim is correct in my own case or in another's? Wittgenstein is sometimes supposed to have argued in the *Philosophical Investigations* that I cannot be said to know such a thing – and maybe that there is nothing to know; the claims are, in some sense, equivalent. But I do not think that can be right, whether or not Wittgenstein

argued for it (and I do not see that he did). Suppose after my treatment I report the results in the first manner. Someone then chides me: how do you know it's not (2)? Now if I say in reply that there is an inner quality of painfulness that I can recall my past experiences at the dentist's to have had, and if I now resurrect that quality from my memory, and compare it with the quality of my present experience, I can see that the present experience has that same quality, only I mind it less; then Wittgenstein has a case against me. That sort of supporting claim must be bogus.[46] I could not confirm for myself by such a combination of recall and introspection that (1) was the right way to talk. Yet all the same I could stick to my story. I could say: all I know is that that's the way I want to describe it – that's how it first occurred to me, and your skepticism hasn't changed my mind: I *still* want to say that. Nothing *shows* me I am in pain, and similarly nothing need show me that my pain is as intense as ever, though I mind it less. Such *things I want to say* count for something, but not, as we have just seen, for everything (we aren't *required* to accept the reports of morphine users or lobotomized subjects).

Could I be supported in my conviction about threshold by further evidence? We might run a survey on those who had had the treatment, and find a consensus. Or we might find that I was an anomaly, or that there were two broad groups of reporters, whose memberships were predictable from some features of the subjects (age, blood type, social background, size of the cortex . . .). Would a consensus confirm my story, or would it simply give us a general fact about pain-talk under certain conditions? The *truth* of the pain-talk would still seem open to question. Or, if one holds that the uniformity of this way of speaking is constitutive of the meaning of 'pain' and hence ensures the truth of all this pain-talk as truth-by-meaning then at least we can ask if, all things considered, this is an apt or perspicuous way of talking, of dividing up the world. One is inclined to think that there must be some reason for us to say one thing rather than another, even if these 'grounds' are not available to us introspectively. It would not be appropriate for us to be so designed that our convictions on this matter were grounded in no distinction of interest at all, but then to what other grounds could one appeal but to internal, sub-personal grounds? Suppose for instance, we were to look inside me and find that the treatment had the effect of diminishing the effects on goal structures, current action-directing sub-routines, and memory, but left unchanged the intensity or magnitude of whatever causally

earlier processes normally co-vary with intensity-of-pain-reported. This would support my way of talking at least indirectly, by showing that there is at least one interpretation of the open schema: 'the magnitude of x is unchanged but the effect of x on y is diminished' that is true. The truth of one interpretation could be called upon to help explain my desire to assert what might be another interpretation, even if we decline for various reasons to identify the referents of the different interpretations of 'x' and 'y'. Suppose, alternatively, that we find normal operation of all systems in the flow-chart after the perceptual analyzer, but a diminution in amplitude for some event or events earlier in the chain. This would seem in just the same way to support the second style of introspective report, and make my account suspect. But would it? Does the diminishing size of the retinal image of a receding figure make suspect the claims of perceptual size constancy? Only, perhaps, to those who hold extremely naive identity theories. Detailed isomorphisms between personal level talk of pains, beliefs, feelings, and actions and subpersonal level talk about impulses, trains, and their effects tempt the impatient to drive the silver spike of identity theory prematurely. The result is inevitably a theory that is easily refuted.

The philosophical questions that an identity theory (or other 'philosophical' theory of pain) would be designed to answer are generated by our desire to put together an account that consistently honors all, or at any rate most, of our intuitions about *what pain is*. A prospect that cannot be discounted is that these intuitions do not make a consistent set. This would not be a particularly unsettling discovery if we could identify a few peripheral and unbuttressed intuitions as the culprits; they could be presumed to be mistaken or illusory, and dismissed, leaving a consistent core of intuitions as the raw material for philosophical analysis and system-building. Thus one might *legislate* a neat systematic relationship between sortal talk of pains and mass-term talk of pain, thereby establishing two distinct 'concepts of pain,' and ignore any intuitions inharmonious to that scheme however well attested to by ordinary usage. Recommending such a slight (and improving) revision of our ordinary concept would not be, arguably, doing violence to our ordinary concept. But if contradiction is more entrenched, a more radical approach is dictated.

Consider the idea that being in pain is not any mere occurrence of stimuli, but an interpreted reception, a *per*ception that is influenced by many prior

cognitive and conative factors. Some will find this intuitive, but pre-theoretically it is hardly compelling. On the contrary, nothing is more 'intuitive' to the sufferer than that there is little that is *cognitive* about pain, that what one wants relief from is not merely an undesirable species of perception, that *in addition to* one's state of consciousness, or perceptual or epistemic state, the pain is there, a brute presence, unanalyzable and independent. The apparent disharmony between these two blocs of intuitions can be turned into clear contradiction if theory is permitted to develop along traditional lines. The grammatical grounds for the contradiction have already been noted: it is equally ordinary to speak of drugs that prevent pains or cause them to cease, or to speak of drugs that render one insensitive to the pains that may persist. (Further instances of the latter notion in our ordinary conception of pain can be found in the discussions of people who are "congenitally insensitive to pain".[47] Our *prima facie* obligation not to cause pain in others is surely understood not to exclude these unfortunate individuals from the class of subjects.) So ordinary usage provides support both for the view that for pains, *esse est percipi*,[48] and for the view that pains can occur unperceived.

What kinds of *objects of perception* are pains: are they merely *intentional objects*, or have they an independent status? No one can defensibly claim to know. Neither introspection not physiological research can cast any light on the question, and philosophical analysis can plausibly support or attack either view for the simple reason that there are common intuitions and associated ways of speaking that support the contrary views.[49] If one takes such contradictory testimony to impeach the authority of such intuitions as determinants of our ordinary concept of pain, where else might one look for testimony? Not even contradiction can dislodge our shared intuitions from their role as best manifestations of – constitutive employments of – our ordinary concept. What must be impeached is our ordinary concept of pain. A better concept is called for, and since even the most rudimentary attempt at a unified theory of pain phenomena is led ineluctably to the position that pain occurs normally only as the result of a process of perceptual analysis, the *esse est percipi* position on pain promises to be more theoretically perspicuous, which, faced with the impasse of intuitions, is reason enough to adopt it.[50] This suggests an identification of pain with events – whatever they are – that occur post-interpretation, so that if we can determine where, in our model, interpretation is completed, whatever issues from that will be

pain (when the interpretation machinery so interprets). Setting aside the categorial crudities of that formulation, there are still problems, for the interpretation of events in such a system is not an atomic matter, but highly compound. Perception has not one product but many, operating at different levels and in different ways. Has the interpretation machinery interpreted a signal as a pain if it evokes a speech disposition to say one is in pain? Or must it also produce the normal or apposite effects on belief, memory, desire, non-verbal action, and so forth? Looking at all the various effects such an interpretation of signals could produce, we can answer the philosophic questions about pain only by deciding which effects are 'essential' to pain and which are not.

What governs our decisions about essentiality, however, is our stock of pretheoretical intuitions, which we have seen to be in disarray. Having countenanced legislation to settle two such conflicts already, we still face incompatibility of well-entrenched intuitions, such as these:

(1) Pains are essentially items of immediate experience or consciousness; the subject's access to pain is privileged or infallible or incorrigible.

(2) Pains are essentially abhorrent or awful – 'Pain is perfect misery, the worst of evils . . . '

Efforts to capture both of these 'essential' features in a theory of pain are bound to fail; theories that contrive to maintain both of these claims do so only at the expense of equally well-entrenched claims from other quarters. To see this suppose we attempt to capture at least part of what is compelling about (1) by the thesis:

(3) It is a necessary condition of pain that we are 'incorrigible' about pain; i.e., if you believe you are in pain, your belief is true; you are in pain.[51]

Condition (3) says that belief that one is in pain is a *sufficient* condition of pain. Such belief may be sufficient, but if we are held to be incorrigible about other states of mind or sensations as well (as incorrigibilists generally hold) there must be some other, distinguishing feature of pains; that they are abhorrent or awful seems as good a candidate as any. But then from (3) and

(4) It is a necessary condition of pain that pains are awful

it follows that believing one is in pain is a sufficient condition for really experiencing or undergoing (and not merely believing one is experiencing or undergoing) something awful. But the belief itself is not the pain, and it is not awful. Surely it is logically possible to be in a dispositional state bearing all the usual earmarks of belief that one is in pain and yet not be experiencing or undergoing something awful. Not only is this logically possible, it is instanced routinely by morphine subjects and others. Then is there any way of denying that this consequence of (3) and (4) is false? There is a heroic line available. One could maintain that *whatever* dispositional state one was in, it could not properly be characterized as the state of belief that one was in pain unless one understood the concept of pain, and hence believed that pains were awful, and hence would never believe one was in pain unless one believed one was experiencing something awful; and then, since we are incorrigible about experience in general, one would never believe one was experiencing something awful unless one was experiencing something awful and, finally, since something *undergone but not experienced* (presuming that we can make sense of such a distinction) could not be awful (in the right sense), it really is quite defensible to claim that belief that one is in pain is sufficient condition for undergoing something awful.[52] This line can 'save' (3) and (4) as conjoined necessary conditions of pain, but only at the expense of other intuitions about our access to our beliefs or our capacity to say when we are in pain. If asked if I am in pain, I should say: "I am if I believe that I am, but who knows if my apparent belief is a genuine belief?" On this view, those who sincerely report that under morphine their pains are intense but not awful are not mistaken in believing they are in pain when they are not (for that has just been deemed to be logically impossible) but in saying something they do not believe (but only believe they believe?). The counterintuitiveness of this result does not utterly disqualify the heroic line. There are any number of ways of cutting this Gordian knot, and this is one of them. One decides which intuitions must go, and builds one's theory accordingly.

I do not recommend the course just considered, however. I recommend giving up incorrigibility with regard to pain altogether, in fact giving up *all* 'essential' features of pain, and letting pain states be whatever natural kind states the brain scientists find (if they ever do find any) that normally produce all the normal effects. When that day comes we will be able to say whether masochists enjoy pain, whether general anesthetics prevent pain or have some other equally acceptable effect, whether there are unfelt pains, and

so forth. These will be discoveries based on a somewhat arbitrary decision about what pain is, and calling something pain doesn't make it pain. This is especially true of pain, for one of our intuitions about pain is that whether or not one is in pain is a brute fact, not a matter of decision to serve the convenience of the theorist. I recommend against trying to preserve that intuition, but if you disagree, whatever theory I produce, however predictive or elegant, will not be by your lights a theory of pain, but only a theory of what I illicitly choose to *call* pain. But if, as I have claimed, the intuitions we would have to honor were we to honor them all do not form a consistent set, there can be no true theory of pain, and so no computer or robot could instantiate the true theory of pain, which it would have to do to feel real pain. Human beings and animals could no more instantiate the true theory of pain (there being none), which lands us with the outrageous conclusion that no one ever feels pain. But of course we do. Human suffering and pain cannot be whisked out of existence by such an argument. The parochiality of the concept of pain protects us but not robots (or Martians or lower aninals) from the skeptical arguments, by fixing the burden of proof: an adequate theory of pain must have normal human beings as instantiations, a demand that presupposes the primacy, but not the integrity, of our ordinary concept of pain.

What then is the conclusion? It is that any robot instantiation of any theory of pain will be vulnerable to powerful objections that appeal to well-entrenched intuitions about the nature of pain, but reliance on such skeptical arguments would be short-sighted, for the inability of a robot model to satisfy all our intuitive demands may be due not to any irredeemable mysteriousness about the phenomenon of pain, but to irredeemable incoherency in our ordinary concept of pain. Physiological perplexities may defy the best efforts of theoreticians, of course, but philosophical considerations are irrelevant to the probability of that. If and when a good physiological sub-personal theory of pain is developed, a robot could in principle be constructed to instantiate it. Such advances in science would probably bring in their train wide-scale changes in what we found intuitive about pain, so that the charge that our robot only suffered what we artificially *called* pain would lose its persuasiveness. In the meantime (if there were a cultural lag) thoughtful people would refrain from kicking such a robot.

Tufts University

NOTES

[1] Keith Gunderson, *Mentality and Machines*, 1971.

[2] A. M. Turing, *Mind*, 1950, reprinted in A. R. Anderson, ed., *Minds and Machines*, 1964, and discussed by Gunderson, *inter alia*, in 'The Imitation Game,' in Anderson, and in *Mind*, 1964, and revised in *Mentality and Machines*.

[3] Cf. Gunderson, *Mentality and Machines*, p. 157: " . . . emotion is different from its effects . . . "

[4] They would be – perhaps unwittingly – wrong if they made this claim, as we shall see. Stimulation of the C-fibres is neither a necessary nor sufficient condition for the occurrence of pain. (C-fibres *are* stimulated under general anesthesia, and need not be stimulated for neuralgia or central pain to occur.) The term 'C-fibres' seems however to have lost, for philosophers, its empirical anchoring in neuro-anatomy and become a philosopher's wild-card referring expression for whatever physical event 'turns out to be identical with' pain.

[5] Such an enterprise might be illuminated by a revival of the researches of the 19th-century investigator Hahnemann, who botanized over seventy-three distinct phenomenological varieties of pain. (See F. Sauerbruch, *Pain: Its Meaning and Significance*, 1963, p. 74.)

[6] The distinction is not as clear-cut as it may first appear. Terry Winograd's natural language understanding program (See Terry Winograd, 'Understanding Natural Language,' in *Cognitive Psychology* 3 (1972), 1–191.) 'manipulates' the 'objects in its environment' and answers questions about them. But its environment is entirely artificial and internal, like the environment of the swirling hurricane of our earlier example. When Winograd's device 'puts a cube on the table,' is it a doer or a describer? Moreover, if we view the theory we incorporate into a program as an *uninterpreted* theory, we are free to view the computer's behavior as satisfying one interpretation of the theory, so that any programmed computer can be viewed as instantiating (on one interpretation) the theory incorporated in its program. The tokens of computer behavior that on one interpretation are uttered *descriptions* of the behavior of some other entity instantiating the theory, can on another interpretation be viewed as themselves instances of behavior predicted by the theory. I owe this observation to Joseph Weizenbaum.

[7] Some of this fantasy has already been turned to fact. SIM ONE, a robot used in training medical students, blinks, breathes, has measurable blood pressure, coughs, twitches, and can become 'anesthetized.' See J. S. Denson and S. Abrahamson, 'A Computer-controlled Patient Simulator' *Journ. Amer. Med. Assoc.* 208 (1969), 504–8.

[8] See, e.g., Louis S. Goodman and Alfred Gilman, *The Pharmacological Basis of Therapeutics*, 2nd edition, 1955, p. 248.

[9] Herbert Simon, in *The Sciences of the Artificial*, 1969, points to the distinction that is occasionally drawn between 'artificial' and 'synthetic'; a green glass gem might be called an artificial sapphire, while a manufactured gem chemically identical to genuine sapphire would be called synthetic. (p. 4)

[10] On this, and many other matters Joseph Weizenbaum has provided me with illuminating suggestions.

[11] Cf. Paul Ziff, 'The Feelings of Robots,' *Analysis* 19 (1959), #3, Reprinted in Anderson, ed., *Minds and Machines*.

[12] See, in this regard, Stanley Cavell's suggestive paper, 'Knowing and Acknowledging,' in *Must We Mean What We Say?*, 1969.

[13] Note that we can obtain the specificity of functional reaction to morphine and other drugs without accomplishing this in the same way, chemically, a human body does.

[14] The difference in transmission speed is considerable: 100 meters per second versus less than two meters per second. If you stick a pin in your finger you can distinguish quite readily two resultant sensations in sequence: roughly, a pricking sensation followed swiftly by a 'deeper' pain. This tandem effect is generally thought to be explained by the difference in transmission speed. See R. A. Sternbach, *Pain, A Psychophysiological Analysis*, 1968, p. 30 for a discussion of the issue, and references.

[15] George Pitcher discusses the philosophical implications of the Melzack-Wall theory in 'The Awfulness of Pain,' *J. Phil.*, July 23, 1970. The most recent and highly developed version of the Melzack-Wall theory is found in R. Melzack and P. D. Wall, 'Psychophysiology of Pain,' in H. Yamamura, ed., *Anesthesia and Neurophysiology, Internat. Anesth. Clinics* 8 (1970), #1.

[16] Lesions in the old low path are responsible for 'central pain' or neuralgia (pain with a central cause but peripheral 'location' – one does not necessarily feel a *head*ache in central pain). 'Cortical representations' of pain are considered less 'important' by researchers generally. Cortical lesions seem almost never to produce central pain (and when they do, descending effects on the old low path are indicated). Penfield, in his research on stimulation of the exposed cortex produced a wide variety of effects, but almost no pain. See V. Cassinari and C. A. Pagni, *Central Pain: A Neurosurgical Survey* (1969); and Wilder Penfield, *The Excitable Cortex in Conscious Man* (1958). Moreover, 'direct stimulation of the reticular and limbic systems produces strong aversive drive and behavior typical of responses to naturally occurring painful stimuli.' (Melzack and Wall, *op. cit.*, p. 20.)

[17] Such volume control systems have been posited in the course of many different investigations in the brain sciences. Arguably such a system's existence has been physiologically confirmed in cats. See Hernandez-Peon, Scherer and Jouvet, 'Modification of Electrical Activity in Cochlear Nucleus during "Attention" in Unanesthetized Cats,' *Science* **123** (1956), 331–32.

[18] The myth is given a more skeptical treatment in my 'Brain Writing and Mind reading,' in Keith Gunderson, ed., *Language, Mind and Knowledge, Minn. Studies in Phil. Sci.*, Vol. 7, 1975.

[19] A journalistic account of this technique can be found in Marilyn Ferguson, *The Brain Revolution*, 1973.

[20] Can the Melzack-Wall theory also account for acupuncture anesthesia? It is not hard to speculate about mechanisms that could be added to the Melzack-Wall theory to accommodate the acupuncture effects, but I understand Wall is currently at least agnostic about the capacity of the theory to handle it.

[21] To some extent pain locations need to be *learned*, though, and can be unlearned. In cases of limb amputation performed on children before they developed the use and coordination of the limb, phantom limb is rarely experienced. When amputation occurs just after birth phantom limb never occurs. See M. Simmel, 'Phantom Experiences Following Amputation in Childhood,' *Journ. of Neurosurgery and Psychiatry* **25** (1962), 69–79.

Moreover, locations can be 'mislearned.' A variety of pain commonly produced in jet pilots under certain high altitude conditions is positively located by them in either the cheeks or the teeth. Which location is reported depends not on variation in the

physiological etiology of the pain (which is constant) but on whether or not the pilots have had recent painful dental work.

[22] Aspirin probably also has central analgesic effects.

[23] A. S. Keats and H. K. Beecher, 'Pain Relief with Hypnotic Doses of Barbiturates, and a Hypothesis,' *J. Pharmacol.*, 1950. Lobotomy, though discredited as a behavior-improving psychosurgical procedure, is still a last resort tactic in cases of utterly intractable central pain, where the only other alternative to unrelenting agony is escalating morphine dosages, with inevitable addiction, habituation and early death. Lobotomy does not excise any of the old low path (as one might expect from its effect on pain perception), but it does cut off the old low path from a rich input source in the frontal lobes of the cortex.

[24] See A. R. McIntyre, *Curare: Its History, Nature and Clinical Use*, 1947; and A. E. Bennett, 'The History of the Introduction of Curare into Medicine,' *Anesth. and Analgesia*, 1968, pp. 484–92.

[25] The fact that most of the patients were infants and small children may explain this credibility gap. See Scott M. Smith, *et al.*, 'The Lack of Cerebral Effects of d-Tubocurarine,' *Anesthesiology* 8 (1947), 1, pp. 1–14. It has recently been confirmed that cases of surgery on curarized but incompletely anesthetized patients is not as rare today as one would hope. See Richard S. Blacher, 'On Awakening Paralyzed During Surgery: A Syndrome of Traumatic Neuroses', *Journal of the American Medical Association* **234** (Oct. 6, 1975), 67–8.

[26] Ibid.

[27] I know of no drug with just these powers, but a number of drugs used in anesthesia are known to have amnestic properties. Scopolamine is the strongest and most reliable amnestic (though it is still unreliable), but it has other effects as well: *not* anesthesia or analgesia, but it does create hallucinations and a sort of euphoria. Scopolamine and other amnestics are often prescribed by anesthesiologists *for the purpose of creating amnesia.* "Sometimes," I was told by a prominent anesthesiologist, "when we think a patient may have been awake during surgery, we give scopolamine to get us off the hook. Sometimes it works and sometimes not." Scopolamine was once widely used in conjunction with a sedative or morphine to produce the 'twilight sleep' then recommended for childbirth. One pharmacological textbook, in discussing this practice, uses the phrase "obstetrical amnesia or analgesia" as if amnesia and analgesia were much the same thing. (Goodman and Gilman, *op. cit.*, p. 555.)

[28] Unable in fact, not unable in principle. We could quite easily devise signalling systems triggered directly by activity in electrode-monitored motor neurons. My point is not that such a state is in principle indistinguishable from anesthesia; I simply want to consider what, aside from current behavioral evidence (and later memory report) is crucial in making the determination.

[29] I have found some people who proclaim their untroubled readiness to accept this substitute. I think they have been bewitched by Wittgensteinian logical behaviorism.

[30] Surveys of anesthesiology can be found in the standard medical school pharmacology texts, such as Goodman and Gilman, *op. cit.*, or J. C. Krantz and C. J. Carr, *The Pharmacological Principles of Medical Practice* (7th edn., 1969, Williams and Wilkins, Baltimore). More up-to-date material on the physiology of anesthesia is contained in H. Yamamura, *op. cit.*

[31] For instance, the 1969 edition of Krantz and Carr describes the drawbacks of

halothane, a recent popular inhalant anesthetic, as follows: it produces incomplete muscle relaxation, and "it does not produce patent analgesic properties, so it is used with nitrous oxide for analgesia, and a curariform [for paralysis]." One might well wonder just what halothane's strengths are.

[32] When anesthesia (without curare) is so deep that reflexes are absent, the worry that this absence is due to a curariform effect of the anesthetic by itself has been laid to rest recently by experiments in which twitch responses were directly evoked in deeply anesthetized subjects by electrode stimulation of motor nerves. (Reported by S. H. Ngai, 'Pharmacologic and Physiologic Aspects of Anesthesiology,' *New England Journal of Medicine*, Feb. 26, 1970, p. 541.) This reassuring datum is somewhat beside the point, however, since under common anesthetic practice, the reflexes are only obliterated by the accompanying curare.

[33] Some further details are of particular interest. Barbiturate anesthetics in sub-hypnotic doses are not anesthetic or analgesic at all, whereas nitrous oxide in sub-hypnotic doses is a reliable analgesic. This meshes well with our physiological account since nitrous oxide not only depresses the reticular formation but also depresses transmission between the thalamus and cortex, an effect barbiturates lack. Melzack and Wall report that in cats barbiturate anesthetics produce strong descending inhibitory effects to their gate system in the substantia gelatinosa. So some general anesthetics may overdetermine their essential effect, but being a hypnotic (suppressing general arousal) is sufficient.

A more puzzling matter is the claim (e.g., by Goodman and Gilman) that 'pain is totally abolished before the onset of unconsciousness' in the first stage of anesthetic induction; a scalpel incision, they say, feels like a blunt instrument drawn across the skin! One is entitled to view this claim with skepticism; surgical incisions during stage one anesthesia without other medication must be exceedingly rare occurrences in modern medicine, and for good reason, so presumably the grounds for the claim are anecdotal and not of recent vintage. But suppose the claim is in fact well grounded (at least true on occasion). At first blush it appears an embarrassment to our theory, since orthodoxy has it that only the cortex is affected during stage one anesthesia, and the effect on it is enhancement, not depression or blockade. How could cortical enhancement possibly produce analgesia? One possible answer: by evoking a hallucination (e.g., of a blunt instrument being drawn across the skin). The abnormal cortical activity of first stage anesthesia is known to evoke hallucinations, and hallucinations do have the power to overrule and obliterate competing veridical inputs (one's hallucinations are not simply *superimposed* on veridical perceptions), so if one were fortunate enough to hallucinate a harmless blunt instrument when the scalpel was plunged in, one would not feel pain. And, of course, one's being fortunate enough would not be fortuitous; the content of hallucinations is apparently guided by our deepest needs and desires, and what apter or deeper guiding desire than the desire to avoid pain? A similar account suggests itself for analgesia under hypnotic suggestion.

[34] Mary Brazier, 'Effects of Anesthesia on Visually Evoked Responses,' in Yamamura, *op. cit.*

[35] H. K. Beecher, 'The Measurement of Pain,' *Pharmacological Reviews*, 1957 pp. 59–191.

[36] This forces us to acknowledge a far from negligible distinction between the pain we humans experience and the pain experienced by creatures that lack a neocortex (unless we want to maintain that only human beings and perhaps a few other 'higher' animals do

experience pain). But it should already be obvious to us that there are tremendous functional differences between human and subhuman pain: no one is surprised that yoga, Zen Buddhism and Christian Science are ineffective anodynes for animals. What of anesthetic practice in veterinary surgery and animal experimentation, however? The hypothesis that 'saves' shallow anesthesia for human subjects is apparently inapplicable to animals without a neocortex. The curare incident should persuade us not to jump to complacent conclusions about this. Current thinking in veterinary anesthesiology closely follows human anesthesiology in most regards: the Melzack-Wall theory is featured, but the action of drugs on the reticular formation is regarded as central. The reticular formation plays about the same role in animals' brains, serving to arouse those higher perceptual areas that are precursors, phylogenetically, of the human neocortex. Somewhat disturbing, however, is the common use in animals of 'dissociative anesthetics' such as phencycladine and ketamine, which do not depress the reticular formation, but produce a state like cataleptic stupor. These drugs have been discontinued for human administration because their anesthetic properties were called in doubt, and patients frequently reported horrible hallucinations (typically, of dying and then flying through outer space to Hell). A survey of the literature on animal pain (which I have only begun) could start with William V. Lumb and E. Wynn Jones, *Veterinary Anesthesia*, 1973. See also J. E. Breazille and R. L. Kitchell, 'Pain Perception in Animals,' *Fed. Proc.* 28 (1969), 1379 and G. Corssen, *et al.*, 'Changing Concepts in Pain Control During Surgery: Dissociative Anesthesia with C1-581 [Ketamine] A Progress Report' in *Anesthesia and Analgesia*, 1968, x 47:6.

[37] See *Content and Consciousness*, Routledge & Kegan Paul, 1969, esp. Ch. 4.

[38] Cf. Thomas Nagel's review of *Content and Consciousness, J. Phil.*, April 20, 1972.

[39] In *Content and Consciousness* I disparaged theories that replaced the little man in the brain by a committee (e.g., on p. 87). This was a big mistake, for this is just how one gets to 'pay back' the 'intelligence loans' of Intentionalist theories (see my 'Intentional Systems,' *J. Phil.*, Feb. 25, 1971). The trick is to turn the whole man into a committee of relative morons, each of which in turn is composed of still less versatile, less intelligent morons, until finally the heuristic value of Intentional characterization of subsystems diminishes to the point where one can abandon it. Thus the programmer starts with a flow chart of which he says things like 'this box wants to do A, and this other box keeps track of where B is, . . .' A recent ingenious computer simulation of a woman in a psychoanalytic interview organized the individual functions of the subsystems in such a way that the woman's actions were determined by a well-organized, interacting group of 'contexts (data locations or mini-worlds)' with such characteristically human tasks and responsibilities that they were aptly labelled Calvin, Freud, Cicero, Machiavelli and Leibniz. (John Clippinger, unpublished Ph. D. dissertation, Univ. of Pennsylvania.) Each context may have twenty or more programs. "These programs are not so much morons as they are specialists who work with specialists and are embedded in specialists" (personal communication from John Clippinger). The distinction between a moron and a specialist has always been vexed. A more detailed defense of homunculus theories is given in my 'Why the Law of Effect Will Not Go Away', *J. Theory of Social Behaviour*, 1975, pp. 169–87.

[40] The reason I do not object to positing a homunculus that, e.g., *infers* on the basis of texture gradients, overlap and perspective clues that a particular object in my visual field is at a particular distance is that although there are grounds for claiming an inference-like

process must be *going on in me*, it is clear enough that *I* do not draw the inference – so long as it gets drawn *in me*. But it is important that I be the subject of my pains. If the proper parts of me are for some purposes construable as homunculi, and if on these construals these proper parts are occasionally the subject of pain (an unlikely turn for the theory to take, but not impossible), then those will not or need not be occasions when it is also the case that I am the subject of pain.

[41] "Indeed the concept [of a pain center] is pure fiction unless virtually the whole brain is considered to be the 'pain center' because the thalamus, the limbic system, the hypothalamus, the brain stem reticular formation, the parietal cortex, and the frontal cortex are all implicated in pain perception." Melzack and Wall, 'Pain Mechanisms: A New Theory,' *Science* **150** (1965), 975.

[42] G. E. M. Anscombe, *Intention* (Ithaca, 1963), p. 14. See also the discussion of this case in Kathryn Pyne Parson, 'Mistaking Sensations,' *Phil. Review*, April, 1970.

[43] Roger Trigg, in *Pain and Emotion* (Oxford 1970), claims correctly I think, that it would be *abnormal* but not conceptually impossible to have a very *intense* pain but not dislike it. Trigg also offers a useful account of intensity of pain in which intensity is sharply distinguished from 'strength of dislike.'

[44] Trigg, *op. cit.*, examines the hypothesis that leucotomes are too confused or imbecilic to know what they are answering.

[45] Two different phenomena have been alluded to by this term. The pain-threshold measured by the Hardy-Wolff-Goodell *dolorimeter* is presumed to be the minimal level of intensity at which a sensation type is deemed painful by the subject. (See J. D. Hardy, H. G. Wolff, and H. Goodell, *Pain Sensations and Reactions*, Williams and Wilkins, Baltimore, 1952, and also H. K. Beecher, *op. cit.*, a classic critique of this experimental method of 'measuring pain').

In more common parlance, one's pain threshold is a *maximum* level of pain one can 'tolerate,' whatever that may be held to mean in the circumstances. The common belief that there is a wide variation in people's tolerance for pain is expressed repeatedly in the medical literature (see, e.g., Asenath Petrie, *Individuality in Pain and Suffering* (Chicago 1967)) but nowhere that I know of is there a careful attempt to confirm this by any objective tests.

[46] Such a claim might be phenomenologically sincere, but as a justification for my convictions about how to describe the result of treatment it is without merit. I owe to Lawrence Davis the suggestion that we mustn't rule out the possibility of having such an experience.

[47] See, e.g., L. D. Cohen, *et al.*, 'Case Report: Observations of a Person with Congenital Insensitivity to Pain,' *J. Abnormal Social Psychol.* **51** (1955), 333, and B. H. Kirman, *et al.*, 'Congenital Insensitivity to Pain in an Imbecile Boy,' *Developmental Medicine and Child Neurology* **10** (1968), 57–63.

[48] See, e.g., George Pitcher, 'Pain Perception,' *Phil. Review*, July, 1970.

[49] See Pitcher, 'The Awfulness of Pain' *loc. cit.*, where a debate is presented between the Affirmativist, who holds that all pains are unpleasant, and the Negativist, who denies this. Pitcher claims, correctly I believe, that this debate "has no winner" (p. 485).

[50] In 'Pain Perception,' *loc. cit.*, Pitcher adopts a similarly pragmatic strategy, defending a 'perceptual' theory of pain that 'will strike many as bizarre' largely on grounds of theoretical cogency.

[51] Not all versions of 'privileged access' (to pains and other items) would maintain, or

imply, this thesis, but many do, and it should be clear in what follows that a parallel argument can be addressed against some important versions that do not. For instance, the view that if one *says*, sincerely and with understanding, that one is in pain, one is in pain, succumbs even more directly to a version of my argument. One might think (and I used to claim) that Saul Kripke was committed to the incorrigibility thesis by his claim, in 'Naming and Necessity' (D. Davidson and G. Harman, eds., *The Semantics of Natural Language*, D. Reidel, Dordrecht, 1972, p. 339): "To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of a pain *is* not to have a pain" But Kripke denies that this claim entails anything like (3) (A. P. A. Eastern Division Meeting, December 29, 1974). This leaves Kripke's notion of epistemic situation obscure to me, and I would not hazard a guess about whether a version of my argument applies to his view.

[5 2] Pitcher discusses a similar argument in 'Pain Perception', pp. 387–88.