DOUGLAS R. HOFSTADTER

WHO SHOVES WHOM AROUND INSIDE THE CAREENIUM?
or
WHAT IS THE MEANING OF THE WORD "I"?

The Achilles symbol and the Tortoise symbol encounter each other inside the author's cranium.

Achilles: Fancy meeting you here! I'd thought that our dialogue in Paris was the last one we'd ever have.

Tortoise: Your never can tell with this author. Just when you think he's done with you, he drags you out again to perform for his readers.

Achilles: I don't see why we should have to perform at his whim.

Tortoise: Just try resisting. Then you'll see why. You don't have any choice in the matter!

Achilles: I don't?

Tortoise: Look – to refuse to perform is tantamount to suicide. Let's face it, Achilles – you and I (at least in these Hofstadterian versions of ourselves) come to life only when Hofstadter writes dialogues about us. We had it good in *Gödel, Escher, Bach*, but now that that's over and done with, I have a feeling the pickings are going to be pretty slim. Hofstadter knows he can't live off of us forever! So we'd better take what we can get!

Achilles: Yes... I remember those good old days. Sometimes we had such wonderful lines. Like that one you had, something about how the "Achillean flash" swoops about my brain "in shapes stranger than the dash of a gnat-hungry swallow." Isn't that how it went?

Tortoise: Something like that. Hofstadter liked that one well enough that he had me say it in at least TWO dialogues! Pretty strange, eh?

Achilles: The way you talk about all this is so bizarre, to my mind. I mean, granted that we're figments of someone else's imagination; but still, you know how characters in a novel are supposed to "come alive" and have "wills of their own"... Surely it's not just a cliché?

Synthese **53** (1982) 189–218.

Tortoise: I wouldn't know. I'm not a novelist. Nor is Hofstadter.

Achilles: I mean, am *I* really just a tool of Hofstadter (however benevolent he is), or am I genuinely exerting my own free will here (as I feel I am doing)? What it comes down to is: Who pushes whom around inside this cranium?

Tortoise: Now THERE'S a planted line, if I ever heard one. That's a direct quote from *GEB*, page 710, where Hofstadter is quoting from Roger Sperry of split-brain fame. It's where Sperry's giving his mind-brain-free-will philosophy, which Mr. H. evidently espouses. But let's get on with the subject matter of THIS dialogue. I think we've done enough introduction. You must have something on your mind, Achilles, which Mr. H. wants to bring up through you.

Achilles: I wish you'd quit putting it in that upside-down way, Mr. T.

Tortiose: All right. But am I right? Isn't there something you're just itching to tell me?

Achilles: Come to mention it, yes. It's related to a book I saw in the bookstore the other day, called *Molecular Gods*: *How Molecules Determine Our Behavior*. It was the subtitle that intrigued me.

Tortiose: In what way?

Achilles: My first thought on reading it was, "Oh, that's interesting – I didn't know that the molecules inside me could affect me that much."

Tortiose: A classic reaction.

Achilles: I know it sounds dumb, but what's wrong with it?

Tortoise: How can you say that? Molecules is all you are, my friend! Read Francis Crick's *Of Molecules and Men* someday.

Achilles: Oh, yes – I know I'm made of molecules. Nobody could deny THAT. It just seems to me that my molecules are at MY beck and call – not individually, of course, but in large "chunks," such as my fingers, when I play my cello or sign a check. So that when *I* decide to do something, my molecules are forced to come along. So – haven't you really got it reversed? Isn't it REALLY the case that *I* shove those molecules around, and not vice versa?

Tortoise: (rather exasperated); What do you mean, "I"? What is this "you"?

Achilles: How I feel – let me put it that way. My free will determines what I do.

Tortoise: All right. Let me suggest a definition. Let me suggest that the term "free will," when you use it, is a shorthand for a complex

set of predispositions of your brain to act in certain ways. Just a moment ago, you used the word "fingers" as an abbreviation for a whole bunch of molecules. In a similar way, the phrase "free will" could be thought of as an abbreviation for a whole bunch of natural tendencies and constraints. So... Your free will – your set of preferred pathways for neural activity to flow along – constrains the motions of molecules inside your brain and those motions in turn are reflected in the patterns that your fingers will trace out.

Achilles: Are you saying that when I say "free will," I'm really using a shorthand for a kind of "hedge maze," like the ones on the grounds of Victorian palaces, a maze that allows some pathways and disallows others?

Tortoise: Yes, that's the idea – only of course this "hedge maze" is inside your skull, and is a bit more abstract. For instance, it's a little oversimplified to imagine that pathways are RIGIDLY allowed or disallowed. It would be more accurate to think of the set of predispositions in terms of a set of pins in a pinball machine. You know what I mean by "pins"?

Achilles: Those stationary round things with rubber "bumpers" that the shiny marbles bounce off of?

Tortoise: Correct. Were you to take an average of over a million marbles, you could find out how each pin statistically affects the way the marbles descend to the bottom. Pathways aren't just ALLOWED or DISALLOWED; rather, some are more likely, some are less likely, depending on how the pins are arrayed. But if you still like the image of the maze of hedges, that's not a bad one to hold in your head. The hedges make more rigid constraints – things are more black and white than with the pins. There are fewer degrees of freedom for the motions in a maze. But I can make the maze image richer. Suppose that in your maze, one of the effects of the people moving through the maze were that the hedges gradually shifted position. It's somehow as if the maze were formed of movable partitions constraining the maze runners, yet the maze runners' paths gradually move the partitions, thus changing the maze.

Achilles: You mean the maze runners could just decide – by free will – that they want to pick up a partition and plop it down somewhere else?

Tortoise: Not like that. It's got to be a deterministic outcome of the

act of maze running itself. Let me go back to the pinball analogy. It's more as if the pins, instead of being fastened on the board, were SLIDABLE objects like hockey pucks, objects that, as they get banged around from above and below and all sides, slightly slip and change positions. The pins need not be circular; they could be longish so that two or more located near each other could act like a channel or a funnel for marbles. In any case, they are jounced around by the rapidly moving pinballs.

Achilles: As in Brownian motion?

Tortoise: Exactly. There are really TWO SCALES in time and space operating here, each affecting the other. The heavy hockey-puck-like pins appear almost stationary to the light marbles. To a casual observer who's following the motions of the marbles, the massive pins would appear to be DETERMINING the light marbles' motions, to be telling the marbles where to go – or in Sperry's words, to be "shoving them around."

Achilles: I like that image. It agrees with my earlier view that *I* shove my molecules around.

Tortoise: True – provided you identify "yourself" with the configuration of the pins.

Achilles: That's a little strange, I admit.

Tortoise: Now imagine a second observer, who's watching a FILM of the whole thing speeded up by a factor of a thousand or more. To *her*, there is a smooth, interesting patterned motion of a bunch of large, variously-shaped pucks. She says to herself, "Wonder why they're moving that way – I can't see anything visible causing any of it."

Achilles: She doesn't see the marbles?

Tortoise: No – they are shooting around so fast in this time scale that their tracks all blur together into one uniform background color with no apparent motion.

Achilles: Ah, yes... Now the facts about Brownian motion begin to come back to me. I remember how people were mystified by the jostling motions of colloidal particles in solutions when they looked at them under a microscope. They couldn't figure out what was causing such motions. The molecules that were constantly battering them were too small to be visible, and besides, they were moving too quickly.

Tortoise: Exactly. An observer on this time scale might start to

develop a sense for the slow drifting patterns of the pucks, even without having any clear notion of what's CAUSING the pucks to move about.

Achilles: It's a natural human tendency. Why not?

Tortoise: The observer could anthropomorphize: "Oh, those two little ones don't like to be close together, and those two long thin ones are trying to be parallel" – and so on. So she develops a teleology, or a way of describing the heavy pucks' motions all on their own. She's quite unaware that they are being bombarded constantly by teeny objects, as in Brownian motion. (Let's pretend that the marbles are more like BB's – really small.) She doesn't know that something smaller is MAKING the pucks swim around in those patterns.

Achilles: So you can turn a knob on your movie projector and flip back and forth between the fast and slow views? Or even smoothly go between them? That's neat! At first, at the slowest setting, the immobile pucks seem to determine the paths of the many little bouncing marbles. As you speed up the film, the marbles become harder and harder to track, and pretty soon they become just a big blur. Meanwhile, you begin to notice that the pucks actually AREN'T immobile, after all. They're being shoved about by the marbles. So who's shoving whom around REALLY? Well, it's mutual, I now see.

Tortoise: Good. Now let me add some more richness to this whole metaphor. Let's say that marbles are constantly being shot in from all sides of the table, and also leaving on all sides. You can envision something like a pool table, with a lot of little marble-launching stations mounted on the walls, and a lot of pockets that act as exits for stray marbles that land in them. The inflow and outflow are equal, so there's no net gain or loss of marbles. And the bombardment is pretty uniform, but not exactly. The marbles are launched according to conditions OUTSIDE the table. For example, if there's a red light near a marble-launching station, that station slows down its firing rate; if a green light is near it, it speeds it up. So you have a set of TRANSDUCERS from EXTERNAL LIGHT to INTERNAL MARBLE-SHOOTING. Now if the puck observer watches both the lights AND the pucks, she'll be able to draw some causal connections between light patterns outside, and the puck-patterns inside. Using mentalistic language will become

quite natural. For instance, it would sound quite reasonable to say, "It saw the green light – its moving away from it – I guess it doesn't like green." And so on.

Achilles: Now you've got me thinking. I too want to add some strange features. I'll propose a physical linkage between one particular puck and an external "arm" that can move towards or away from the lights. So, when that puck moves a certain way on the table, the arm may push a light away or pull it closer. Of course this is very primitive – there are no fingers or anything, but at least there's now a two-way link between the pucks and the lights. Gosh! I'm almost completely forgetting about those marbles careering around down there! I'm just RELYING on the marble-shooters to keep on doing their job without much maintenance or attention needed... All I see now is the seemingly animate interplay – a sort of dance – among the pucks, the lights, and the arms...

Tortoise: We're really jumping from one metaphor to another, aren't we? And each time we escalate in complexity... Oh, well, that's fine with me. No matter how complex the scene gets, you can always slow down the projector, unblur the marbles and no longer see the pucks moving at all.

Achilles: Of course. But there's now something that bothers me. In the brain, there AREN'T these large- and small-sized units – everything's uniform, right? I mean, it's all just a dense packing of neurons. So where do the two scales come from? If we go back to the maze and partitions, there too we had two levels of objects (maze people and maze walls), each kind pushing the other around. But in the brain, this isn't so – or is it? What else is there besides neural activity?

Tortoise: Let's add, then, a level of detail to our picture that we didn't have before. Let's say there are no pucks at all. There are only marbles and a number of larger stiff yet malleable mobile metal strips, which I'll also describe as "stiff yet malleable membranes" (and you'll soon see why). They can be bent into U's or S's or circles....

Achilles: So these things are swimming in the soup of marbles, now, but there are no pucks, eh?

Tortoise: Right. Can you guess what might happen now?

Achilles: I can imagine that these strips...

Tortoise: Would you mind calling them "stiff yet malleable membranes," just to please me?

Achilles: Are you going to pull some acronymic trick off in a moment? Let's see – "SYMM" doesn't spell anything, does it? Is that really what you want me to call them, Mr. T.?

Tortoise: In fact you anticipated me, Achilles. Go ahead and do call them "SYMM"s.

Achilles: All right. So these SYMMs will now be knocked around along with the marbles which are bashing into one another. Will the SYMMs occasionally get wrapped around some group of marbles and form a circular membrane, separating out a group of marbles from the rest?

Tortoise: Just call the circular structure so formed a "SYMM-ball," if you please.

Achilles: Oh... I should have seen it coming. All right. Now I see that in this way, structures like pucks are emerging again, only this time as COMPOSITE STRUCTURES made up out of many, many marbles. So now, my old question of who pushes whom around in the cranium – er, should I say "in the CAREENIUM"? – becomes one of SYMMBALLS versus MARBLES. Do the marbles push the symmballs around, or vice versa? And I can twiddle the speed control on the projector and watch the film fast or slow.

Tortoise: I should mention that once a symmball is formed, it might have quite a bit of stability, because the marbles inside it get fairly densely packed together, and jostle each other around only a little bit when the symmball gets hit by a fast marble from the outside. The impact gets spread around and shared among the marbles inside, and the symmball won't tend to break up – at least not when you watch the film at either of the two speeds we've already mentioned. Perhaps the FISSION of a symmball would occur on a longer time scale than the MOTIONS of symmballs. And the same for the formation of a symmball.

Achilles: Would it be fair to liken a symmball's emergence to the solidifying of water into a cube of ice?

Tortoise: An excellent analogy. Symmballs are constantly forming and unforming, like blocks of ice melting down into chaotically bouncing water molecules – and then new ones can form, only to melt again. This kind of "phase transition" view of the activity is very apt. And it introduces yet a third time scale for the projector, one where it is running much faster and even the motions of the symmballs would start to blur. Symmballs have a dynamics, a way of forming, interacting, and splitting open and disintegrating, all

their own. Symmballs can be seen as reflecting, internally to the careenium, the patterns of lights outside of it. They can store "images" of light patterns long after the light patterns are gone – thus the configurations of symmballs can be interpreted as MEMORY, KNOWLEDGE, and IDEAS.

Achilles: It seems to me that although you got rid of the pucks, you added another structure – the SYMMs. So how is this new system any improvement, as a model of a brain, over the old one? Don't you still have two levels of basic physical constituents and activity?

Tortoise: The SYMMs are there only to provide a way for marbles to join up and form clusters. There are other conceivable ways I could have done this. I could have said, "Imagine that each marble is magnetic, or velcro-coated, so that they all attract each other and stick together (unless jostled too hard)." That suggestion would have had a similar effect – namely, of making much larger units grow out of smaller ones – and so you would have only ONE kind of basic physical constituent. Would that be more satisfying to you, Achilles?

Achilles: Yes, but then you'd have lost your pun on "symbols," which would be too bad.

Tortoise: Not at all! I'd cleverly rename the marbles themselves this time, as "small yellow magnetic marbles" – "SYMMs" – and a magnetically bound cluster of them would form a "SYMM-ball." No loss.

Achilles: That's a relief! I would hate to see a good metaphor go down the drain for lack of a pun to illustrate it.

Tortoise: Hofstadter would never let THAT happen! You can take it from me. Anyway, you can conceive of the larger units however you want, as long as you have it clear in your mind that starting with just ONE level, you wind up with TWO levels and TWO time scales – three time scales, in fact, when you take into account the slow formation, fission, fusion, and fizzling of the symmballs.

Achilles: Now can we go back and talk about whether *I* control my molecules, or my molecules control ME? That's where this all started, after all.

Tortoise: Certainly. Why don't you try to answer the question yourself?

Achilles: The problem is that in all those pulsations inside a careenium, I just don't see a "me." I see a lot of activity – I see a

lot of internalized representations of things "out there" – I mean of light patterns, in this case. And with fancier transducers, we could have a careenium in which symmball patterns reflected such things as sounds, touches, smells, temperatures, and so on.

Tortoise: Let your imagination run wild, Achilles!

Achilles: All right. If I stretch my imagination, I can even see a gigantic three-dimensional careenium, hundreds of feet on a side, filled with billions upon billions of marbles floating in zero gravity, shooting back and forth, and all over forming short-lived and long-lived symmballs, and with those symmballs in turn governing the marbles' paths. I can see all that, and yet I don't see free will or "I." I guess I can't see how *I* MYSELF could be a system like this inside my cranium. *I* feel alive! *I* have thoughts, feelings, desires, sensations!

Tortoise: Hold on, hold on! One at a time. These are all related, but let's try to talk about just one – say, thoughts. Let me propose that the word "thought" is a shorthand for the activity of the symmballs that you see when you run the movie fast: the way they interact and trigger patterns of motions among themselves (mediated, of course, by the constant background swarming of marbles, too fast to make out).

Achilles: But I FEEL myself thinking. There's no one INSIDE a careenium to FEEL those "thoughts." It's all just a bunch of silly yellow magnetic marbles bashing into each other! It's all impersonal and unalive. How can you call that "thought"?

Tortoise: Well, isn't it equally true of the molecules running around in YOUR brain? Where's the soul of Achilles that "shoves THEM around"?

Achilles: Oh, Mr. T., that's not a good enough answer. I've just heard it said too many times that we're made out of atoms, so there's no room for souls or other things – but I know I'm there, it's an undeniable FACT, so I need more insight than a mere reminder that my body obeys the laws of physics. WHERE does this feeling of "I" come from, a feeling that I have and you have but stones DON'T have?

Tortoise: You're calling my bluff, eh? All right. Let's see what I can do to turn you around. Let's add one more feature to the careenium – an artificial mouth and throat, just as we added an arm. Let various parameters of them be driven by various symmballs.

Now suppose we turn on a green light on the right-hand side of the careenium. New marble activity near that side begins immediately, and there follows a complex regrouping of symmballs. As it all settles down into a new steady configuration, the mouth-throat combination makes an audible sound: "There's a green light out there." Maybe it even says, "I saw a green light out there."

Achilles: You're trying to play on my weaknesses. You're trying to get me to identify with a careenium by making it more human-seeming, by making it simulate talking. But to me, this is merely "artificially signaling" (to borrow one of my favorite phrases from Professor Jefferson's Lister Oration). Do you expect me also to believe that somewhere out there, there is a conscious person reciting the time of day twenty-four hours a day, simply because I can dial a certain number and hear a human voice say, over the telephone, "At the tone, Pacific Daylight Time will be five forty-two"? A voice uttering sentence-like sounds doesn't necessarily signify the presence of a conscious being behind it.

Tortoise: Granted. But this careenium voice isn't merely uttering a mechanically repetitious sequence of sentences. It is giving a dynamic description of what is perceived in the vicinity.

Achilles: I have a question about that. Is the thing being perceived located OUTSIDE the careenium, or INSIDE it? Why does the mouth say, "I saw a green light OUT THERE" rather than say something such as, "Inside me, a new symmball just formed and exchanged places with an old one"? Isn't THAT a more accurate description of what it perceived?

Tortoise: In a way, yes, that is what it perceived, but in another way, no, it did not perceive its own activity. Think about what perception really involves. When you perceive something "out there," you cannot help but mirror that event inside you somehow. Without that internal mirroring event, there would be no perception. The trick is to know what kind of external event triggered it, and to describe what you felt out loud in public language that refers to something external. You subtract one layer of transduction. You omit, in your description of what happened, one step along the way. You omit mention of the step that converted the green light into internal symmball responses. You are not even aware of that step, unless you are something of a philosopher or psychologist.

Achilles: Why would I or anyone else omit a real level? What's the

origin of this socially conventional lie? *I* don't omit levels in MY speech!

Tortoise: Actually, you do. It's a universal phenomenon. If you live near a railroad track and hear a certain kind of loud noise coming from that direction – rumbling, bells dinging, and so on – do you say, "I hear a train," or do you say, "I hear the SOUND of a train"?

Achilles: I guess that ordinarily, I would tend to say, "I hear the train."

Tortoise: Do you see a train, or do you see LIGHT hitting your eyes? When you touch a chair, do you feel the chair, or do you feel your FEELING of the chair?

Achilles: I opt for the simpler alternative. I never would think those extra philosophical thoughts that go along with it. What point would it serve to say, "I hear the SOUND of the train"?

Tortoise: Exactly my point. The most convenient language, the least obfuscatory and pedantic, omits the heavy "extra" reference to the medium carrying the signals, omits mention of the transducers, and so on. It simply gets straight to the EXTERNAL SOURCE. This seems, somehow, the most HONEST way to look at things – and the least confusing. You hear and see a TRAIN, not an image of a train, not the light reflected off a train, not retinal cells firing – and most definitely not your perception of a train. We are constructed in such a way as to be unaware of our brain's internal activity underlying perception, and therefore we "map it outward."

Achilles: Yes, I see that pretty well. I think I see why a careenium with a voice might talk about a green light rather than talk about its own symmballs. But wait a minute. How would it know anything about green lights? It might PREFER to refer to things in the outside world – but nonetheless, all it knows about is its own internal state!

Tortoise: True, but its way of verbalizing its internal state employs words that you and I think refer to objects and facts OUTSIDE the careenium. In fact, it too thinks so. But you could very well argue that it is just making sounds mirror its internal state in some very complex way: it could be deluding itself. There might be nothing out there to refer to!

Achilles: True, but that's not exactly my question. What I want to know is, how come it uses the RIGHT WORDS to describe what's out there? Where did it learn to say "green light"? The same

question goes for people. How come we all say the same sounds for the same things?

Tortoise: Oh, that's not so hard. I had thought you were asking whether reality exists or not. I quickly tire of such pointless quibbling over solipsism. But let me answer the question you DID ask. When you were a tot, you saw things – say rattles – and heard certain sounds – namely, various pronunciations of the word "rattle" – at about the same time. Those sights and sounds were transduced from your retinas and eardrums into internal symbol states inside your cranium. Now, as a member of the human race, you were constructed in such a way as to enjoy mimicry, and so you made funny noises something like "wattle," which were then automatically picked up by your eardrums, and fed back into the interior of your cranium. You heard your own voice, to your great delight and thrill! You were then able to compare the sounds YOU'd just made with your memory of the sounds you'd heard. By playing this exciting new game, you were learning the English words for objects. Of course you started with the nouns for visible objects, but quickly you built on that most concrete level and over the next few years you developed a large vocabulary including such things as "ball", "pick up", "next to", "splash", "window", "seven", "sort of", "zebra", "tongue-twister", "remember", "maze", "stretch", "by accident", "of course", "blunder", "confetti", "equilibrium", "analogy", "vis-à-vis", "chortle", "Picasso", "double negation", "few and far between", "neutrino", "Weltanschauung", "n-dimensional vector space", "tRNA-amino-acyl synthetase", "solipsism", "careenium" –

Achilles: Wait a minute! What about "banana split"?

Tortoise: Now how did I overlook that? A shameful oversight. But I hope you get the point.

Achilles: I think I see what you mean. Gradually, I internalized a huge set of external, public, aural conventions – namely the English words attached with particular states of my own brain, states that were correlated with things "out there."

Tortoise: Not just things – actions and styles and relationships and so on.

Achilles: To be sure. But instead of conceiving that the words described my brain state, it was easier to conceive of them as describing things out there DIRECTLY. In this way, by omitting a level in my

interpretation of my own brain's state, I cast internal images outwards.

Tortoise: A careenium would do likewise – casting its internal symmball patterns outwards, attributing them to some properties of the external world. And if a large number of careenia happened to be located near some specific stimulus, they could all communicate back and forth by means of a set of publicly recognizable noises that are externalizations of their internal states! So it's actually very useful to subtract out the references to the transduction, perception, and representation levels.

Achilles: It all makes sense now. But unfortunately, something else is bothering me! If the system projects all its states OUTWARDS, talking about "green lights" and "red lights" and "traffic jams" and so forth, then how is there any room left for it to perceive its own INTERNAL state? Will it be able to say, "I'm annoyed" or "I forgot" or "I don't know" or "It's on the tip of my tongue" or "I'm in a blue funk"? Or will it project all those inner states outwards as well, attributing weird qualities to things outside of it? Could there be inward-directed transducers that focus on SYMMBALLS and come up with a REPRESENTATION of symmball activity? That would be a sort of sixth sense – an inward-directed sense.

Tortoise: You could call it the "inner eye."

Achilles: That's a perfect name for it.

Tortoise: The inner eye wouldn't need to do much transducing, would it? Symmball activity is the easiest thing in the world to monitor because it's right there inside you.

Achilles: Now, Mr. T., you always warn *me* about confusing use and mention; I think you yourself are committing that error here. To have a word such as "Tortoise" in a text is enough to make somebody conjure up the image of a Tortoise, but it is not at all the same as making that person start to think about the WORD "Tortoise," is it? They may not notice it at all.

Tortoise: Point well taken. There is a difference between HAVING your symmballs in certain states, and being *aware* of that fact. It's something like the difference between using grammar correctly and knowing the rules of grammar.

Achilles: Now I sense I could get really confused here – things could get very tangled. How can symmballs "watch" other symmballs? The ones that react to green lights, I can imagine and understand.

There are transducers – the marble shooters on the borders. But would there be some symmball that always reacts to, say, the fusion of two symmballs? How would it detect such a fusion? What would make it react that way? Would it be a sort of satellite or U-2 plane, with an overview of the whole terrain of the brain? And what purpose would it serve?

Tortoise: Imagine that you were watching an actual careenium, and at a very slow speed – so slow that you could reach down with your hand and pick up and remove an entire symmball before getting struck by any symms careening towards your hand. All of a sudden there would be a vacuum, where before there had been a dense mass of marbles. If you switched speeds now and watched the results in the SYMBALLS' time scale, you'd see a massive regrouping of symmballs all over the careenium, a kind of SHUDDER passing through the whole system as all the various symmballs come to occupy slightly different positions.

Achilles: You could call such a shudder a "mindquake."

Tortoise: An excellent suggestion. Various types of "mindquakes" would have characteristic qualities to them. They would have "signatures," so to speak. Now if YOU, Achilles, an observer from the outside, could learn to recognize such a signature, then why couldn't the system itself, from within, be even more able to do so? Such mindquakes would be, after all, just as tangible to the system as is an increase of marble-firings on any side. Both are simply INTERNAL EVENTS, even though the one is triggered by something external, while the other is set off by something internal.

Achilles: So would there be various "seismometer symmballs," each one sitting there waiting to feel a specific kind of mindquake, and when that happens it would react?

Tortoise: Sure. And for each type of mindquake, there is a special symmball just sitting there like a pencil on end – and when its type of mindquake comes along, it topples. Of course that "toppling" in itself is just MORE SYMMBALL ACTIVITY...

Achilles: Another mindquake?

Tortoise: Precisely – and it can set off further reactions inside the careenium. The whole thing is very circular – one shudder triggers another one and that one sets off more, and so on.

Achilles: It sounds like it would never stop. There would just be constant symmball activity rippling back and forth across the careenium.

Tortoise: Well, of course! That IS what happens with conscious systems, isn't it? We're constantly thinking thoughts – some fresh, some stale – constantly mentally alive and aware – partly of the external world, partly of our own state – for example, how confused or tired we are, what something reminds us of, how bored we are with this long monotonous dialogue. . ..

Achilles: Hey, wait a minute! The READERS may be bored, but I'm not!

Tortoise: Only kidding, Achilles. Just trying to liven things up a bit.

Achilles: All right. Well anyway, I admit that everything you've been saying is true, makes sense, but how is it USEFUL for us to monitor our own state?

Tortoise: Well, think first of a simple animal. What it needs most of all is food. Its brain – if it has one – is connected to its stomach by nerves and it transduces an emptiness in the stomach into a certain configuration of symbols in the brain. Actually, this animal might be so simple that the symmball level doesn't exist. There might just be marbles zipping around in its cranium, but no larger-scale agglomerations. In any case, the effect of this may then be a shuddering in its brain, which produces repercussions on the animal's peripheries. It may move. All this is very much at the reflex level. Mostly it involves monitoring the organism's hunger state and controlling its limbs. Every organism has to monitor itself in terms of hunger. But primitive organisms don't use much information about the external environment they're in: they just flap about and "hope" – if that isn't too strong a word! – to encounter some food. Pretty unconscious. On the other hand, take a more complex animal. It will have an elaborate representation of its environment inside itself, so it also has a lot of options when it detects internal hunger. The symbolic activity representing the empty stomach has to be dealt with in the context of all the other symbols, which might represent danger, priorities other than eating, choices of when and what to eat, and so on. The total interaction of symbols at that point we might call "consideration" or "deliberation" or "reflection" – as distinguished from "reflexes." Now after all this, let me ask you: does this help you to see why such a careenium might have a self?

Achilles: Well. . . I might grant that there's reflection going on in there, I might even grant that it's THINKING – but there's noBODY in there DOING the thinking!

Tortoise: Would you grant that there's FREE WILL inside there?
Achilles: Hardly!
Tortoise: Then I can see that you will need some more persuading. All right. Let me suggest that there IS free will, and that this notion of a careenium may help you understand more clearly what free will truly consists in. We began this discussion by talking about whether you can "shove your molecules around" or not. This is a central question – in truth, it is THE central question, I think. So I'd like to ask you, Achilles, can you freely decide to do anything?
Achilles: Of course I can! That's precisely what free will is about! I can decide to do whatever I want!
Tortoise: Really? Can you decide, say, to answer me in Sanskrit?
Achilles: Obviously not. But that has nothing to do with it. I don't speak Sanskrit. How could I answer you in it? Your question doesn't make sense.
Tortoise: Not so. You can only do what your brain will allow you to do, and that is very crucial. Let me ask you another question. Can you decide to kill me right now?
Achilles: Mr. T.! What a suggestion? How could you suggest such a thing, even in jest?
Tortoise: Could you nevertheless decide to do it?
Achilles: Sure! Why not? I can certainly IMAGINE myself deciding to do it.
Tortoise: That is beside the point, Achilles. Don't confuse hypothetical or fictitious worlds with reality. I'm asking you if you CAN decide to kill me.
Achilles: I guess that in this world, in the REAL world, I could not CARRY OUT such a decision, even had I "decided" – or claimed I'd decided – to do it. So I guess I COULDN'T decide to do it, actually.
Tortoise: That's right. That innocent-seeming trailer phrase that one tends to tack on – exactly as you did – is very telling, after all.
Achilles: What innocent phrase? What do you mean?
Tortoise: Don't you remember? You insisted vehemently to me, "I can decide to do WHATEVER I WANT." Now that phrase "whatever I want" may SOUND like a grand, universal, all-inclusive, sweeping phrase – but in fact, it represents quite the opposite: a severe constraint. It's not true that you are able to decide to do ANYTHING; you are limited to being able to decide to do only

things you WANT. Worse yet, you are in fact limited to doing, at any time, the ONE thing that you want MOST to do! Here, "want" is a complex function of the state of the entire system.

Achilles: Are you saying that choice is an illusion?

Tortoise: Only to the extent that "I" is an illusion. Let me explain. It's quite common for people to develop interests that begin to consume them – doing puzzles, doing music, thinking about philosophy... Sometimes such habits get so strong that they begin to interfere with the rest of their lives. A wife may pick up a bad habit – say, twiddling a cube or smoking cigars or constantly punning – and then TRY to get rid of it. Her exasperated husband may say to her, "What's this TRYING? Why can't you just DECIDE to stop cubing? It is driving a wedge between you and me. Why don't you just DECIDE to quit?" Yet the afflicted wife may, for all her good intentions, be unable to do so. Certainly having a modicum of desire is not enough. I would put it this way. The husband is appealing to what I would call his wife's "soul" – a coherent set of principles and tendencies and interests and personality traits and so forth that represent to him the person that he married. They have always before seemed to provide reasons or explanations for his wife's character, and he loves her for that aggregate of ways of being. So he appeals to this "soul" to put a clamp on its new obsession. But once the wife starts twiddling her cube, a PART of her takes over. She gets obsessed – or should I say "possessed"? – by one of her own subsystems!

Achilles: "Possessed" is the word for it. I myself find it very hard to stop practicing a piece on my cello once I have gotten into the swing of it. Before I start, I think, "Now, I'll just play this piece ONE TIME." (Or, "I'll just eat one potato chip," or "I'll just solve the cube one time.") But then, once I've let myself start, I'm no longer quite the same person – some things inside me have subtly shifted. And the NEW me thinks, "THAT guy said HE'D do it only once. That's what He thought. But *I* know better!" There is a kind of inner inertia that makes me want to continue, even when there are OTHER things I would also like to do. It's as if some part of you just "slips away" from a higher level of control – some subsystem gets "out of control" and won't obey the soul on top – like a bucking bronco unwilling to obey its rider.

Tortoise: A powerful image. In such cases the wife herself may be confused and torn. Her inner turmoil is like that of a country in

inner strife. There are factions battling each other – only in this case, the factions are neural firings, not people, of course. On some level, this woman may *FEEL* she *wants* to *be able to decide* to *give up* her habit – yet she may not have enough neurons on her side! And as in a country where the people won't support the government, so here – the "soul" has to have the support of its neurons! It can't just arbitrarily "shove them around," in reality.

Achilles: I'm all confused. Who IS in control, here?

Tortoise: We'd like to be able to say that the symmballs can DECIDE to do arbitrary things, but they are constrained. They are in a system that "wants" its parts to move in some ways but doesn't "want" them to move in others. We could come back to the hedge-maze metaphor, to make this more vivid.

Achilles: Yes, but that applied to the LOWER-LEVEL objects – marbles, symms, or neural firings.

Tortoise: Exactly. The "heavyweight" entities – hedges, pins, symmballs – constrain the "lightweight" entities – maze runners, pinballs, symms; but in revenge, the little things, acting together, control how the high-level ones are arrayed.

Achilles: So NOBODY's free here!

Tortoise: Well, from the outside, that's the way it seems. But on the inside, the system may feel, just as you did, that it can "decide to do whatever it wants to do." But, mind you, two symmballs in a careenium aren't free to DECIDE, arbitrarily, on their own, to move in (say) parallel – they have to have the cooperation of the marbles. The marbles have to do the work for them. Similarly, when the unhappy wife tries to "decide" to give up her cubing or punning habit, she can't do it without the agreement, the support, of her neurons.

Achilles: You make the wife's "soul" sound like a general trying to marshall unruly neurons, to force them into line when they have their OWN paths to follow. A military general has some degree of power over his soldiers, so he can coerce them to some extent – but only so far. Beyond that, they'll mutiny. So the general has to go along with the tide. He can't really dictate policy – he can only resonate with it.

Tortoise: It's true. However, sometimes an unexpected shift at a higher level can precipitate an abrupt "phase transition" at lower levels. A million tiny things suddenly find themselves swirling

around in unexpected ways, and realigning in totally novel higher-level patterns. Once in a while – just once in a while – the "general" CAN gain control of those unruly neurons – but only when they themselves don't know what they want, haven't reached any kind of consensus, and are instead in a malleable, leadable, chaotic state.

Achilles: It sounds like you're describing a "snap decision" – an exercise of pure will power, such as when I say to myself, "I'm going to quit cubing RIGHT NOW," or "I'm going to stop feeling sorry for myself and go out and get something useful done." But if I understand YOUR way of looking at this kind of thing, even a phrase like "snap decision" is really just a kind of shorthand for summarizing a lot of low-level activity. Is that so? It seems to me it would HAVE to be so, in your picture.

Tortoise: You're right, saying something like "snap decision" is really a coarse-grained manner of speaking about a huge cloud of neural activity, like a huge blurry cloud of symms in a careenium projected at high speed on the screen. And sometimes the activity of neurons inside a cranium, or of symms inside a careenium, lends itself admirably to such a high-level, coarse-grained, symbolic description – or in the case of a careenium, a "symm-ball-ic" description.

Achilles: Not always?

Tortoise: Are all ponds always frozen?

Achilles: Oh, I see what you mean. If the relevant portions of the carreenium are "frozen," so that they form macroscopic, higher-level structures like blocks of ice on a pond, then a symmball-level description can be made. One set of symmballs is seen to affect other sets of symmballs regularly and predictably. Whereas if there ARE no symmballs – just a lot of stray symms careening around like water molecules with nothing to constrain them except the careenium's boundary – then it's kind of chaotic, and no higher-level description applies. But when the whole careenium is "symm-ball-ic" – when the phase transitions have taken place – then the person – I mean the CAREENIUM! – feels very much in control of his or her thoughts.

Tortoise: ITS thoughts?

Achilles: Yeah, yeah – that's what I meant. *Its* thoughts. But when not enough phase transitions have taken place, then there's an indescribable hubbub: random symms careening all over the place without orderly constraints. But I wonder what it's like when the

brain is in sort of a halfway state – when there are lots of symmballs, but at the same time still a lot of stray symms that belong to no one. It reminds me of a half-frozen lake in early winter or early spring, when the molecules have only HALF-coalesced into large blocks of ice.

Tortoise: That's a wonderful state to be in. I find I'm most creative when I feel my brain consisting of such halfway-coalesced symbols – neurons acting somewhat independently, somewhat collectively. It's a happy medium where neural bubblings cooperate with symbolic channelings and yield the most creative, fulfilling, semi-chaotic sense of aliveness.

Achilles: You think some of that uncoalesced freedom is essential for creativity?

Tortoise: I was convinced of it by Hofstadter, who certainly feels that way. In *GEB*, writing about his plight as a writer, he portrayed himself as suffering from "helplessness" of the top level, for although he – or his symbol level – may in SOME sense have decided what to write, still he is entirely and utterly dependent on vast cooperating teams of unknown neurons to come up with imagery and ideas and choices of words and sentence structures. Those lower-level items feel to the top level as if they "bubble up" from nowhere. But in reality they are somehow formed from the churning, seething masses of interacting neural sparks – just as patterns of symmball motions emerge out of the chaotic Brownian motion of the many tiny symms. And a few of those ideas make it out through the narrow channel of verbalization, like grains of sand passing through the narrow neck of an hourglass. Yet most likely Hofstadter will INSIST that he himself is responsible for this dialogue, will desire the credit to accrue to HIM.

Achilles: Hmm... to the overall system that constrained the marbles to jounce in those ways... It is hard to assign "credit" or "blame," once you start analyzing thought mechanistically. I see that "decision" and "choice" are very subtle concepts that somehow have to do with the ways in which constraints on two different levels affect each other reciprocally, and at two different time scales, inside a cranium, or a careenium.

Tortoise: You're getting the idea.

Achilles: You know, now that I think about it, sometimes the decisions I make seem to be slow percolating processes, things that

are utterly out of my control. In fact, a rather gory image that illustrates this idea flashed before my mind's eye while we were talking about the difficulty of breaking out of mental ruts.

Tortoise: What was that?

Achilles: I imagined a grim scene where a man's young wife is in a car crash and is badly mangled. He will certainly REACT. Perhaps he will react with love and devotion, perhaps with pity. Perhaps he will even react with revulsion, to his own dismay. But it occurred to me that in such emotionally wrenching cases, you can hardly DECIDE what you will feel. Something just HAPPENS inside you. Subtle forces shift deep inside you, hidden, subterranean. It's quite scary, in a way, because in real crises like that, instead of being able to DECIDE how you'll act, you FIND OUT what sort of stuff you're made of. It's more passive than active – or more accurately put, the action is on levels of yourself that are far lower – far more microscopic – than you have direct control over.

Tortoise: Correct. You and your neurons are not on speaking terms, any more than a country could be on speaking terms with its citizens. There is, in both cases, a kind of collective action of a myriad tiny elements on low levels that swings the balance – exactly as in a country that "decides" to go to war or not. It will flip or not, depending on the polarization of its citizens. And they seem to align in larger and larger groups, aided by communication channels and rumors and so on. All of a sudden, a country that seemed undecided will just "swing" in a way that surprises everyone.

Achilles: Or, to shift imagery again, it's like an avalanche caused by the collective outcome of the way that billions upon billions of snow crystals are poised. One tiny event can get amplified into stupendous proportions – a chain reaction. But the crystals have to be poised in the right way, otherwise nothing will happen.

Tortoise: In cases of judgment, whether it be of one musical composer over another, one potential title or subtitle for a book over another, or whatever, the top level pretty much has to wait for decisions to percolate up from the bottom level. The masses down below are where the decision REALLY gets made, in a time of brooding and rumination. Then the top level may struggle to articulate the seething activity down below, but those verbalized reasons it comes up with are always a posteriori. Words alone are never rich enough to explain the subtlety of a difficult choice.

Reasons may sound plausible but they are never the essence of a decision. The verbalized reason is just the tip of an iceberg. Or, to change images, conflicts of ideas are like wars, in which EVERY REASON HAS ITS ARMY. When reasons collide, the real battleground is not at the verbal level (although some people would love to believe so); it's really a battle between opposing armies of neural firings, bringing in their heavy artillery of connotations, imagery, analogies, memories, residual atavistic fears and ancient biological realities.

Achilles: My goodness, it sounds terrifying! You make the battlefield of the mind sound like a vast mined battlefield! Or a treacherous ice field on a steep mountain face. I never realized that a mechanistic explanation of thinking could sound so organic and living. It's sort of awful and yet it's sort of awe-inspiring as well. But I am very confused now about the "soul," the free will.

Tortoise: I think that all these strangely evocative images have brought us back to your original perplexity, over the question of who pushes whom around in the cranium. Would you now be inclined to say, Achilles, that your molecules push YOU around, or that YOU push THEM around?

Achilles: Actually, I'm not sure how this "I" fits into a cranium – or a careenium. You've got my head so spinning now that I don't know what's up or down.

Tortoise: Wonderful! At least now your mind may be malleable. Do you see how "free will" in a careenium is actually constrained – PHYSICALLY constrained, I mean – by the "wants" of the system?

Achilles: Yes, I see that these seemingly intangible "wants" are actually physical attributes of the overall system – tendencies to shun certain modes of behavior, or to repeat certain patterns. So in a way I can see that a careenium has "free will" in this constrained sense of freedom. Maybe "free will" should be renamed "FREE WON'T."

Tortoise: Oh my, Achilles! Did you just make that clever one up?

Achilles: I don't know – it just came to me. I never thought about it. It just "bubbled up from nowhere." I don't know who deserves the credit. Maybe Hofstadter made it up. Or maybe it just bubbled up inside HIS brain – although I don't quite see the difference.

Tortoise: It sounds like the sort of thing Hofstadter's friend Scott Kim would say.

Achilles: Hmm... I still wonder, though – could a careenium's symmballs actually DECIDE to do anything on their own?

Tortoise: They certainly can't disobey the way the symms push them around – but on the other hand, the symms ARE always poised in just such a way that the ONE internal event that the symmballs MOST want to happen WILL happen. Isn't that a miraculous coincidence?

Achilles: Now that I understand how all this comes about, I can see that it's NOT AT ALL a coincidence. By the DEFINITION of "want," the symmballs will get shoved around the way they want to be (WHETHER THEY LIKE IT OR NOT)! I guess that the real conviction of having free will would arise when, repeatedly and reliably, a collection of symmballs wants something, and then watches its desire getting carried out. It must seem like magic!

Tortoise: It's what happens when you decide, say, to sign your name. Your fingers begin obeying you, and miraculously, you watch your name just appear before you, effortlessly! Is that magic?

Achilles: Aha! That brings back that ultimately confusing term, "I." We say "*I* decide to sign my name." But what does that mean? I can see everything in a careenium – wants, desires, beliefs – but I just can't seem to take that last step. I simply fail to see an "I" in there.

Tortoise: I've tried to explain that the word "I" is just a shorthand used by a system such as a careenium – a system that perceives itself in terms of symmballs and their predispositions to act in certain ways and not others – particularly a careenium that has NOT perceived that it is composed of small yellow magnetic marbles.

Achilles: Perceive, Shmerceive, Mr. T.! There's no one inside a careenium who COULD perceive such a thing. Perception requires AWARENESS, which no careenium has. There's no one inside a careenium to feel and experience and ENJOY its "free will," even if it's there, in your sense. Or maybe the best way to say it is that there's perception and free will there, but there's nobody there to have it.

Tortoise: You mean you seriously would grant that a physical system

could have FREE WILL but you wouldn't then feel forced to say there was SOMEONE EXERCISING that free will? Or that there was perception but no PERCEIVER? Perceiverless perception? Agentless, subjectless free will? Soul-less, inanimate free will? That's a real doozy!

Achilles: I know it sounds paradoxical. I could almost agree with you – except I'm still hung up on one thing. Just WHICH perceiver, WHICH agent, WHICH subject, WHICH soul would it be? Which person gets to BE that careenium? Or maybe I should turn the question around: Which careenium gets to have a given soul? Do you see what I mean?

Tortoise: I think so. You seem to be envisioning a corral of souls up in the sky, into which God (or some other Grand Agent) dips, whenever a new cranium or careenium comes into existence, and from which he pulls out a soul, imbuing that careenium or cranium with THAT identity forevermore – almost as if he were putting a cherry on top of a sundae.

Achilles: Are you mocking me?

Tortoise: I don't mean to be. If it sounds that way, it's only because I'm trying to take what I think your implicit notion of "soul" is, and to characterize it explicitly, by putting it into as graphic terms as possible, even if silly. But if you subtract out the imagery of a corral and God and cherries on sundaes, am I not putting into words the gist of your view?

Achilles: In a way, I suppose so – only you've made it sound so silly that I hesitate to adopt that view now.

Tortoise: It's so tempting to think that different I's are just "out there," dormant, waiting to be attached to structures, like saddles put on horses or cherries on sundaes. Then, once they are in place, suddenly there is a consciousness that "wakes up." As if the consciousness, and the identity, the "who-ness," were provided by the cherry, and without it there would be only a hollow "pseudo-I" – a thing possessing free will but with nobody to BE! Isn't that a little sad? Wouldn't you feel sorry for such a poor, deprived entity? Oh, no, of course you wouldn't – there would be no one to be sorry for, right?

Achilles: Well, it's hard to see where a sense of "who I am" could come to a bunch of marbles in a careenium, or even to a collection of firing neurons. It seems to me that the identity HAS to be

imposed on top of such a structure. A careenium is a complex pinball machine – a heap of metallic machinery – even if, unlike pinball machines, some of its states represent the world and its workings. But until you add some sort of living "flame" to that heap, it's empty – soulless. You need "flame" (although I admit I don't know quite what I mean by that term) to turn a physical object into a BEING, just as you need flame to turn a pile of wood into a fire. No matter how much lighter fluid you pour on it, without a flame, it's still inert.

Tortoise: Wait a minute! A pile of wood starts BURNING when you set flame to it – but does it acquire a SOUL at that moment? No – as you said, it simply becomes active instead of inert. Any old flame would do. The identity of a fire doesn't come from the flame that lit it, but from the combustible materials: It's the transition from inactivity to activity that makes the flame seem so critical. But a careenium doesn't need to become MORE active than it is. Yet for some reason, Achilles, you seem to balk at my suggestion that in that activity there is as much reason to see an individualized soul as in neural firing activity. But what's so special about neurons? You know what you remind me of?

Achilles: I don't know that I WANT to know, but tell me anyway.

Tortoise: You remind me of somebody who runs into a pile of metal that's merrily burning away, and who declares that although it LOOKS mighty like a fire, it surely can't be a fire (especially not a GENUINE fire!), because it's made of metal, and everyone knows that fires – especially GENUINE ones – are always made of burning wood or paper.

Achilles: That sounds pretty silly and narrow-minded – more so than I am, I should hope. I'm not insisting that no careenium could have a genuine soul so much as I am wondering, "IF a careenium had a soul, WHICH soul would it be? Who would be THIS careenium, who would be THAT one?" On what basis could a decision be made?

Tortoise: Boy, have you got things upside down! (Or backwards – I'm not sure which.) The same question goes for people as much as for careenia. Who gets to be which body? Do you also have the belief that ANY BODY could be ANYBODY? All it takes is the right flame inside? Could there be a "flame transplant," where someone else's flame – say mine – got implanted in your body,

leaving your brain and body intact? Then who would be you? Or, who would you be? Or WHERE would you be?

Achilles: And where would YOU be, Mr. T.? Something seems wrong in this picture, I admit. If a careenium is actually somebody, where does the decision as to WHO it is originate?

Tortoise: I think you've got things backwards. (Or upside down – I'm not sure which.) First of all, it's not a DECISION – it's an OUTCOME. Secondly, which "who" a careenium is is an outcome of its structure, particularly the way it represents its own structure in itself. The more it is able to see itself as an independent and coherent agent, the more of a "who" there is for it to be. Eventually, by building up enough of a sense of its unique self, it has built up a complete "who" for it to be: a soul, if you will. The continuity and strength of the feeling of "being someone" come from identification with past and future versions of the same system, from the way the system sees itself as a unitary thing moving and changing through time.

Achilles: That's a strange idea – a thing whose identity remains stable even though that thing changes in time. Is it like a country that changes and yet remains somehow the same country? I think of Poland, for instance. If ANY country has had its soul-flame tampered with, Poland is it – yet it seems to have maintained a continuous "Polish spirit" for hundreds of years.

Tortoise: A beautiful example. The sense of "one thing, extending through time" is very much at the root of our feeling of "being someone". And in a way it is nature's hoax: the illusion of soul-sameness. Or, if you prefer not to call it an illusion, you can say that the ability of an organism to abstract, to think it sees some constant thing, over time, that it considers its self even as it changes, makes that organism's soul NOT an illusion.

Achilles: You mean anything that can fool itself – I mean, SEE itself – as unchanging over time has a soul?

Tortoise: That's not such a silly notion – provided that the verb "see" has its usual abstract meaning, not some dilution of the term. If the organism is as perceptually powerful as living ones like you and me, then I would definitely say it has a soul, if it sees itself as essentially "the same organism" over time.

Achilles: But to see itself AS AN ORGANISM is not a trivial thing! It has to see itself as one coherent thing acting for REASONS, not just randomly.

Tortoise: Now you're talking! I couldn't agree more. Such a way of looking at something – namely, ascribing mental attributes to it – has been called by Daniel Dennett "adopting the intentional stance" toward that thing. In the case of you looking at a careenium, it would come down to your seeing it at the SYMMBALL level, and interpreting the symmball configurations and the patterns they go through over time as representing the system's beliefs, desires, needs, and so on, overlooking the underlying complexity of the marbles, either deliberately or out of ignorance.

Achilles: But you're not talking about ME looking at a system; you're talking about a situation where the system does that to ITSELF, right?

Tortoise: Exactly. It looks at its own behavior and, instead of seeing all the little marbles deep down there making it act as it does, it sees only its SYMMBALLS, acting in sensible, rational ways...

Achilles: The system sees itself just as observers of the fast film see it! It could say of itself, "It wants this, believes that, etc." – only now it is ascribing all these beliefs and penchants and preferences and desires and so forth to itself, so instead it says, "*I* want this, believe that," and so on. This seems peculiar to me. It makes up a bunch of hypothetical notions about itself simply out of convenience, then ascribes them to itself in all seriousness. For God's sake, though – if beliefs and desires and purposes and so on REALLY existed inside itself, wouldn't the blasted careenium itself have direct access to them?

Tortoise: What makes you think those beliefs AREN'T real? Aren't ice cubes and traffic jams and symmballs real? And what makes you think that this self-perception ISN'T direct access to its beliefs? After all, does your perception of your own feelings via your "inner eye" differ so wildly from this?

Achilles: I suppose not.

Tortoise: When an OUTSIDER ascribes beliefs and purposes to some organism or mechanical system, he or she is "adopting the intentional stance" toward that entity. But when the organism is so complicated that it is forced to do that with respect to ITSELF, you could say that the organism is "adopting the AUTO-intentional stance." This would imply that the organism's own best way of understanding itself is by attributing to itself desires, beliefs, and so on.

Achilles: That's a very strange sort of level-crossing feedback loop,

Mr. T. The system's self-image (as a collection of symmballs) is getting recycled back into the system, but of course this depends on the very concrete symms themselves to carry it out. It's like a television looking at its own screen, recycling a representation of itself over and over, building up a pattern of nested self-images on the screen.

Tortoise: And that stable pattern becomes a real thing in and of itself. If you were a careenium, merely by adopting the auto-intentional stance toward yourself, you would create a self-perpetuating delusion. As soon as you create this illustion that there is just one thing there – a unitary self with beliefs and desires rather than a mere bunch of goalless and soulless marbles – then that illusion reenters the system as one of its own beliefs. The more that illusion of unity is cycled through the system, the more established and hardened and locked-in the whole illusion becomes. It's like a crystal whose crystallization, once started, somehow has a catalyzing effect on its further crystallization. Some sort of vicious closed loop that self-reinforces, so that even if it starts out as a delusion, by the time it has locked in, it has so deeply permeated the system's structure that no one could possibly explain how or why the system works as it does without referring to its "silly, self-deluding" belief in itself AS A SELF.

Achilles: But by that time it isn't so silly any more, is it?

Tortoise: No, by then it has to be taken quite seriously, because it will have a lot of explanatory power. Once the self has become so locked-in, or "reified," in the system's own set of concepts, this fact determines much of the system's own future behavior – or at least if you are restricted to watching the fast projector, to looking at the SYMMBALL level, that is the easiest way to understand matters. And the curious thing is that this SAME level-crossing feedback loop (of adopting the auto-intentional stance) takes place in EVERY careenium of sufficient complexity. So that whichever careenium you take, the stable self-image pattern that it finally establishes in this loopy way is isomorphic to the stable self-image pattern in every OTHER careenium!

Achilles: Bizarre! The medium is different, but the abstract phenomenon it supports is the same. It's a universal. That's sort of hard to grasp.

Tortoise: Maybe so, but it's right. They all have isomorphic, identical

senses of "I." As Erwin Schrödinger suggested, there is just ONE sense of the word – just one referent – just one abstract pattern – yet each one seems to feel IT knows it UNIQUELY! There's a kind of fight for sole possession of something that everyone shares.

Achilles: "Sole possession," Mr. T.?

Tortoise: It was unintentional, Achilles – quite unintentional.

Achilles: Do you really believe there is just ONE "I", Mr. T.?

Tortoise: Not quite – an exaggeration for rhetorical purposes. The real point is, there's only ONE MECHANISM underlying "I-ness": namely, the circling-back of a complex representation of the system together with its representations of all the rest of the world. Which "I" you are is determined by the WAY you carry out that cycling, and the way you represent the world.

Achilles: So you mean that all that determines who "I" am is the set of experiences some organism has gone through?

Tortoise: Not at all. I said "The WAY things are cycled," not "WHICH things are so cycled and represented." You've got to distinguish between the SET of objects represented, and the overall STYLE or SPIRIT with which they are represented. It's that SPIRIT that determines how the loop will loop. THAT's what creates the uniqueness of each "I".

Achilles: Well, Mr. T., I think I am begining to see your point. It's just SO hard, emotionally, to acknowledge that a "soul" emerges from so physical a system as a careenium.

Tortoise: The trick is in seeing the curious bidirectional causality operating between the levels of the system, and in integrating that vision with a sense of how symbols have representational power, including the power to recognize certain qualities of their own activity, even though only approximately. This is the crux of the mental, and the source of the enigma of "I."

Indiana University

BIBLIOGRAPHY

Anderson, Alan Ross (ed.): 1964, *Minds and Machines*, Prentice-Hall, Englewood Cliffs.

Applewhite, Philip B.: 1981, *Molecular Gods*, Prentice-Hall, Englewood Cliffs.

Dennett, Daniel C.: 1981, *Brainstorms*, Bradford/MIT, Cambridge, Mass.

Hofstadter, Douglas R.: 1979, *Gödel, Escher, Bach: an Eternal Golden Braid*, Basic Books, New York.

Hofstadter, Douglas R.: 1982, *The Tumult of Inner Voices*, Utah State Board of Regents.
Hofstadter, Douglas R. and Daniel C. Dennett: 1981, *The Mind's I*, Basic Books, New York.
Schrödinger, Erwin: 1967, *Mind and Matter*, Cambridge Univ. Press, Cambridge.
Sperry, Roger: 1965, 'Mind, Brain, and Humanist Values' in John R. Platt (ed.), *New Views of the Nature of Man*, Univ. of Chicago Press, Chicago.