

SL #641: I don't think I said that! The sense of being alive and being a unique link in the infinite chain certainly is profound. It's just that it doesn't transcend physical law. To the contrary, it is a profound exploitation of physical law — hardly mundane! On the other hand, the all-too-common desire to mystify the pronoun "I", as if it concealed a deeper mystery than other words do, truly muddies up the picture. The sole root of all these strange phenomena is *perception*, bringing symbols and meanings into physical systems. To perceive is to make a fantastic jump from William James' "blooming, buzzing confusion" to an abstract, symbolic level. And then, when perception twists back and focuses on itself, as it inevitably will, you get rich, magical-seeming consequences. *Magical-seeming* mind you, but not *truly* magical. You get a level-crossing feedback loop whose apparent solidity dominates the reality of everything else in the world. This "I", this unreal but unutterably stubborn marble in the mind, this "Epi" phenomenon, simply takes over, anointing itself as Reality Number One, and from there on out it won't go away, no matter what words are spoken.

SL #642: So the "I" is all too marbelous — too marbelous for words?

SL #641: What?! I thought you thought my "I" idea was for the birds.

SL #642: It's true, I did, but I think I'm catching your drift. Perhaps I'm coming around a little bit. Your strange-loop view of an "I" is close to paradoxical, and yet not quite. It's like Escher's *Drawing Hands* — paradoxical when you're sucked *into* the drawing by its wondrous realism, yet the paradox dissolves when you step back and see it from *outside*. Then it's just another drawing! Most intriguing! It's all too much, and just too very Berry... to ever be in Russell's Dictionary.

SL #641: Ah, music to my ears! I'm so delighted you find a bit of merit in my ideas. As you know, they are only metaphors, but they help me to make some sense of the great puzzle of being alive and, as you kept on stressing, the great puzzle of being *here*. I thank you for the splendid opportunity of exchanging views on such subtle matters.

SL #642: The pleasure, I assure you, was all mine. And I shall await our next meeting with alacrity, celerity, assiduity, vim, vigor, vitality, savoir-faire, and undue velocity. Adieu till then, and cheerio!

[Exeunt.]



Hofstadter, I am a strange loop
Basic Books, 2009

CHAPTER 21

A Brief Brush with Cartesian Egos



Well-told Stories Pluck Powerful Chords

IN THE preceding dialogue, the query most insistently posed by Strange Loop #642 was, "What makes me housed in this particular brain, rather than in any other one?" However, even though Strange Loop #641 tried to provide an answer to this enigma in several different fashions, Strange Loop #642 always had the nagging feeling that Strange Loop #641 hadn't really gotten the question, and hadn't understood how profoundly central it is to human existence. Could it be that there is a fundamental breach of communication here, and that some people simply never will get the question because it is too subtle and elusive?

Well, if one is not averse to using a science-fiction scenario, this same question can be posed so vividly and starkly that hopefully no one could fail to understand and feel deeply troubled by the enigma. One way of doing this appears in the path-breaking book *Reasons and Persons* by the Oxford philosopher Derek Parfit. Here is how Parfit poses the riddle:

I enter the Teletransporter. I have been to Mars before, but only by the old method, a space-ship journey taking several weeks. This machine will send me at the speed of light. I merely have to press the green button. Like others, I am nervous. Will it work? I remind myself what I have been told to expect. When I press the button, I shall lose consciousness, and then wake up at what seems a moment later. In fact I shall have been unconscious for about an hour. The Scanner here on Earth will destroy my brain and body, while recording the exact states of all my cells. It will then transmit this information by radio.

Traveling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine. It will be in this body that I shall wake up.

Though I believe that this is what will happen, I still hesitate. But then I remember seeing my wife grin when, at breakfast today, I revealed my nervousness. As she reminded me, she has been often teletransported, and there is nothing wrong with *her*. I press the button. As predicted, I lose and seem at once to regain consciousness, but in a different cubicle. Examining my new body, I find no change at all. Even the cut on my upper lip, from this morning's shave, is still there.

*

Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant, "It's not working. What did I do wrong?"

"It's working," he replies, handing me a printed card. This reads: "The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities which this technical advance offers."

The attendant tells me that I am one of the first people to use the New Scanner. He adds that, if I stay an hour, I can use the Intercom to see and talk to myself on Mars.

"Wait a minute," I reply, "If I'm here I can't also be on Mars."

Someone politely coughs, a white-coated man who asks to speak to me in private. We go to his office, where he tells me to sit down, and pauses. Then he says: "I'm afraid that we're having problems with the New Scanner. It records your blueprint just as accurately, as you will see when you talk to yourself on Mars. But it seems to be damaging the cardiac systems which it scans. Judging from the results so far, though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days."

The attendant later calls me to the Intercom. On the screen I see myself just as I do in the mirror every morning. But there are two differences. On the screen I am not left-right reversed. And, while I stand here speechless, I can see and hear myself, in the studio on Mars, starting to speak.

Since my Replica knows that I am about to die, he tries to console me with the same thoughts with which I recently tried to console a dying friend. It is sad to learn, on the receiving end, how unconsoling these thoughts are. My Replica assures me that he will take up my life where I leave off. He loves my wife, and together they will care for my children. And he will finish the book that I am writing. Besides having all of my drafts, he has all of my intentions. I must admit that he can finish my book as well as I could. All these facts console me a little. Dying when I know that I shall have a Replica is not quite as bad as simply dying. Even so, I shall soon lose consciousness, forever.

What Pushovers We Are!

The concerns around which Parfit's two-part story revolves are clearly those that haunted Strange Loop #642. In the first part, we worry along with Parfit whether he will truly exist again after he is atomized on Earth and the signals carrying his ultra-detailed blueprint have reached Mars and directed the construction of a new body; we fear that the newly built person will merely be someone who looks precisely like and thinks precisely like Parfit, but is not Parfit. Soon, however, we are relieved to find out that our worries are unfounded: Parfit himself made it, down to the last tiny scratch. Great! And how do we know that he did? Because he told us so! But which "he" is it that gives us this good news? Is this Derek Parfit the philosopher-author, or is it Derek Parfit the intrepid space voyager?

It is Parfit the space voyager. As it happens, Parfit the philosopher is just spinning a good yarn, doing his best to make it sound teddibly realistic, but we soon find out that, in fact, he doesn't believe in several parts of his own story. The second episode in his fantasy starts out by contradicting the first one. When we find out that the New Scanner, in contrast to the old one, *doesn't* destroy the "original", we go right along with the tacit idea that Parfit the intrepid space voyager has not voyaged anywhere. We don't question his stepping out of the cubicle on Earth, because *he's still here*.

Oh, but what mindless pushovers we are! Whereas we bought right into the "teleportation equals travel" theme of Episode I, falling for it hook, line, and sinker, we seem in Episode II to have unthinkingly taken the path of least resistance, which runs something like this: "If there are two different things that look like, think like, and quack like Derek Parfit, and if one of those things is located where we last saw Parfit and the other one of them is farther away, then, by God, the close one is obviously the *real* one, and the far one is just a *copy* — a clone, a counterfeit, an impostor, a fake."

This already is plenty of food for thought. If the copy on Mars is a fake in Episode II, why wasn't it a fake in Episode I? Why were we such suckers when we read Episode I? We naïvely bought into his wife's reassuring smile at breakfast, and then, when he stepped out of the Martian cubicle, that telltale nick on his face convinced us beyond all doubt. We took his word for it that it was indeed *he* who was stepping out of the cubicle. But what else could we have expected? Was the newborn body going to step out of the cubicle and proclaim, "Oh, horrors, I'm not me! I'm someone else who merely *looks* like me, and who has all of my memories stretching all the way back to childhood, and even my memory of breakfast only a few moments ago with my wife! I'm just a sham, but oh, such a good one!"

Of course the newly built Martian is not going to utter something incoherent like that, because he would have no way of knowing that he is a fake. He would believe for all the world that he *is* the original Derek Parfit, only moments ago disintegrated in the scanner on Earth. After all, that's what his brain would tell him, since it's identical to Derek Parfit's brain! This shows that we have to treat claims of personal identity, even ones coming straight from the first person's mouth, with extreme caution.

Well then, given our new no-nonsense attitude, what should we think about Episode II? We have been told that Parfit the would-be space voyager instead stepped out of the cubicle *on Earth*, and with heart damage. But how do we know that *that one* is Parfit? Why didn't Parfit the storyteller tell us the story from the vantage point of the new Martian who also calls himself "Derek Parfit"? Suppose the story had been told this way: "The moment I stepped out of the Martian cubicle, I was told the terrible news that the *other* Parfit — that poor fellow way down on Earth — had suffered cardiac damage in bearing me up here. I was devastated to hear it. Soon he and I were talking on the phone, and I found myself in the odd position of trying to console him just as I had recently consoled a dying friend..."

If it had been recounted sufficiently smoothly, we might not have been able to resist the thought that *this* body, the Mars-borne one, is really Derek Parfit. Indeed, Derek Parfit the skilled philosopher-storyteller might even have gotten us to imagine that the earthbound body with the damaged heart was merely a pretender to the Unique Soul linked by birth and by divine decree to the name "Derek Parfit".

Teleportation of a Thought Experiment across the Atlantic

It seems that the way in which a science-fiction scenario is related is crucial in determining our intuitions about its credibility. This is a point that my old colleague and friend Dan Dennett has made many times in his

discussions of philosophers' crafty thought experiments. Indeed, Dan calls such carefully crafted fables *intuition pumps*, and he knows very well whereof intuition pumps in the field of philosophy of mind.

And I have to say that as I was typing Parfit's story from his 1984 book into this chapter, a little voice murmured softly to me, "Say, doesn't this remind you of Dan's foreword to *The Mind's I*, his ingenious teleportation fantasy that drew so many readers to our book when it came out in 1981?" And so after the Parfit story had been all typed in, I pulled a copy of *The Mind's I* off my shelf and reread its first few pages. I have to say that my jaw fairly dropped. It was exactly the same fantasy, only with planets reversed and sexes reversed, and told in a more American style. There was exactly the same bipartite structure, the first part featuring a "Teleclone Mark IV" that destroyed the original, and the second part featuring a new-and-improved version ("Mark V") that preserved the original.

What can I say? I love both of these stories, one from each side of the Atlantic, whether one is a "clone" of the other or their pedigrees are independent (though that seems unlikely, since *The Mind's I* is in Parfit's bibliography). In any case, now that I've got this little matter off my chest, I'll continue with my commentary on Parfit's provocative tale (and also, of course, on Dan's, thanks to the referential power of analogy).

The Murky Whereabouts of Cartesian Egos

The key question raised by Parfit's tale is this: "Where is space voyager Derek Parfit *really*, after the teletransportation has taken place in Episode II?" Put otherwise, which of the two claimants to being Parfit really is Parfit? In Episode I, Parfit the storyteller plants a most plausible-seeming answer, but then in Episode II he just as plausibly undermines that answer. At this point, you can probably almost hear Strange Loop #642 intensely identifying with the space traveler and screaming out, "Which of the two would I be?"

To my mind, one cannot claim to have said anything significant about the riddle of consciousness if one cannot propose (and defend) some sort of answer to this extremely natural-seeming and burning question. I think that by now you know my answer to the question, but maybe not. In any case, I'll let you ponder the issue for a moment, and meanwhile, I'll go on to tell you more or less how Parfit sees the matter.

This issue lies at the very core of Parfit's book, and the explanation of his position occupies about a hundred pages. The key notion to which he is opposed is what he dubs "Cartesian Pure Ego", or "Cartesian Ego", for

short. To put it in my words, a Cartesian Ego constitutes one exact quantum of pure soul (also known as “personal identity”), and it is 100 percent indivisible and undilutable. In short, it is what makes you be *you* and me be *me*. My Cartesian Ego is mine and no one else’s, has been from birth and will be to death, and that’s that. It’s my very own, completely private, unshared and unsharable, first-person world. It’s the subject of my experiences. It’s my totally unique inner light. You know what I mean!

I have to admit, parenthetically, that every time I see the phrase “Cartesian Ego”, although my eyes perceive only one “g” there, some part of me invariably hallucinates another “g”, and the image of an egg bubbles up in my brain — a “Cartesian Eggo”, if you’ll permit — a beautifully formed egg with a pristine white shell protecting a perfectly spherical and infinitely precious yolk at its core. In my strange distorted imagery, that yolk is the secret of human identity — and alas, Parfit’s central mission in his book is to mercilessly crush the whole egg, and with it, the sacred yolk!

There are two questions that Parfit does his best to answer. The first one is: When Parfit is teleported to Mars in Episode I, is his Cartesian Ego teleported along with him, or is it destroyed along with his body? The second question, seemingly even more urgent and confusing, is this: When Parfit is teleported to Mars in Episode II, where does his Cartesian Ego go? Could it possibly go to Mars, abandoning him on Earth? In that case, who is it that remains on Earth? Or conversely, does Parfit’s Cartesian Ego simply stay put on Earth? In that case, who, if anyone, is it that debarks from the cubicle on Mars? (Note that we are conflating the word “who” or the phrase “who it is” with the notion of a specific, uniquely identifiable Cartesian Ego.) The temptation to ask such questions (and to believe that these questions have objectively correct answers) is nearly irresistible, but nonetheless, the nearly universal intuitions that give rise to this temptation are what Parfit is out to crush in his book.

To be more specific, Parfit staunchly resists the idea that the concept of “personal identity” makes sense. To be sure, it makes sense in the everyday world that we inhabit — a world without telecloning or fanciful cut-and-paste operations on brains and minds. The fact is, we all more or less take for granted this notion of “Cartesian Ego” in our daily lives; it is built into our common sense, into our languages, and into our cultural backgrounds as profoundly, as tacitly, as seamlessly, and as invisibly as is the notion that time passes or the notion that things that move preserve their identity. But Parfit is concerned with investigating how well this primordial notion of Cartesian Egos stands up under extreme and unprecedented pressures. As a careful thinker, he is doing something analogous to what Einstein did

when he imagined himself moving at or near the speed of light — he is pushing the limits of classical notions — and, like Einstein, he finds that classical worldviews do not always work in worlds that are very different from those in which they were born and grew.

Am I on Venus, or Am I on Mars?

In his hundred or so pages of musings on this issue, Parfit analyzes many thought experiments, some dreamt up by himself and some by other contemporary philosophers, and his analysis is always keen and clear. I have no intention to reproduce here those thought experiments or his analyses, but I will summarize what his conclusions are. The essence of his position is that when pushed to its limits, personal identity becomes an indeterminate notion. In extreme circumstances such as Episode II, the question “Which one of them am I?” has no valid answer.

This will be extremely unsatisfying and unsettling to many readers of Parfit’s book, and to many readers of this book, as well. Our intuitions as we grew up on planet Earth have not prepared us for anything in the least like a nondestructive teleportation scenario, and so we clamor for a simple, straightforward answer, yet somehow we also intuit that none will be forthcoming. After all, we could invent Episode III, featuring a *destructive* teleportation scenario as in Episode I, but with signals simultaneously sent out to receiving stations on Venus and on Mars. In this scenario, shortly after the destruction of the original Parfit body and brain, two brand-new Parfifs (both complete with shaving nick) would be assembled more or less simultaneously on the two planets, and now there really doesn’t seem to be any valid claim of primacy for either one above the other (unless you argue that the *first one finished* should get to claim the honor of the Cartesian Ego, but in that case, we can simply posit that they are assembled in synchrony, thus barring that easy escape route).

To our everyday, downhome, SL #642-style minds, it’s very stark and very simple: one of the Parfifs is a fake. We cannot imagine being in two places at once, so we think (identifying ourselves with the intrepid voyager), “Either I’ve got to be the *Venus* one, or the *Mars* one, or *neither* one.” And yet none of these answers is in the least satisfying to our classical intuitions.

Parfit’s own answer is actually closer to the thought that I brusquely dismissed in the previous paragraph: that we are in two places at once! I say it’s *closer* to that answer rather than saying that it *is* that answer, because Parfit’s view, like mine in this book, is that these things that seem so black-and-white to us actually come in shades of gray — it’s just that in ordinary circumstances, things are always so close to being *pure* black or white that