CRITICAL STUDY

Peter Unger, *Identity, Consciousness and Value* (New York: Oxford University Press, 1990), xiv + 341 pp.

PETER VAN INWAGEN
Syracuse University

What are the conditions of a (human) person's survival? That is, under what conditions do we have a single person who exists at two different times? What are the things that might happen to me such that, after they had happened, a certain entity that existed thereafter would be strictly and literally *myself*? In this impressive book, Unger proposes not so much to answer these questions as to uncover our deepest beliefs about their answers—and to uncover what we most deeply value in matters connected with survival. (But he does not propose to uncover these beliefs in order to refute or criticize them. Since he is one of "us," they are *his* beliefs, and, apparently, nothing that has come to light in the course of his inquiry has suggested to him that they need revision.) The data Unger considers relevant to his investigation are these: Our reactions when we are presented with various possible cases and thought-experiments ("Yes, that would still be me"); our "second-order intuitions," the intuitions that are revealed when we compare and contrast cases and thought-experiments ("Well, I think I'd survive in both cases, but I'm a lot more sure about the first than the second"); our tendency to assent to or reject general philosophical principles ("I really don't see how anyone I meet in the future could be the person I met this morning unless there was some sort of physical continuity between them"); our willingness (supposing that we are concerned only about ourselves) to be subjected to torture now to save a carefully described future person from undergoing a greater torture, and our willingness to make a sacrifice in order to be subjected to one "treatment" rather than another. These data, carefully interpreted, can be used to test proposals concerning our deepest beliefs about survival. As one might expect, a body of "raw" data of these sorts has to be taken with a grain of salt. It might be, for example, that an informant's response was "tainted" by his favorite theory of personal identity. Or it might be that a case was described in a tendentious or otherwise misleading way. Or it might be that a case was described so abstractly that the informant was forced to fill in a good many details before he could decide what he thought, the consequence being that the reaction would not be to the case as presented but to the case as privately "filled in" by the informant. Or it might be that various features of the case

were deeply at odds with the informant's world-view ("Suppose a wizard were to change you into a bronze statue . . ."), the informant's reaction being then partly a product of his belief in the sheer impossibility of the case. Or it might be that the informant has assented to a general philosophical principle only because he had not considered a particular class of cases. And there might be many other factors that would distort the evidential value of one's responses in particular circumstances. (Unger makes the very important point that one's reactions to cases described in works of fiction are influenced by many factors other than one's deepest beliefs about personal identity—the desire to let the author get on with the story, for example. Thus, our willingness, while caught up in the story, to assume that it really is Dr. McCoy who steps out of the transporter chamber is of little value as a guide to what we really believe about personal identity.) I should point out that Unger is not so naive as to suppose that one's responses to cases of the sort he considers will be always be clear and unambiguous. He is aware that one can be of two minds about what to say when presented with such cases. When our responses are ambiguous, his concern is to uncover our "dominant" response, and to explain why our response is ambiguous and why a particular one of the responses that we feel drawn towards is dominant.

One might raise the question of who "we" are. I once had a Muslim student (he was the salutatorian of his class at Syracuse University, and went on to earn a Ph.D. in nuclear engineering) whose responses to the kinds of examples Unger presents in his book were wildly at odds with the responses Unger ascribes to "us." (For example: A perfect atom-for-atom duplicate of a living human being would not be alive, since the principle of life is an immaterial soul.) No doubt my student's beliefs were a function of his upbringing—and whose are not?—, but that does not mean that they were not his "deepest" beliefs. Unger does not give us any principled way of discovering who "we" are, although it is clear that "we" are supposed to share a certain world-view (the one that the cases we consider are not permitted to be wildly at variance with). I suspect that only a very small percentage of the five billion or so human beings currently living, and an even smaller percentage of the human beings who have ever lived, actually subscribe to this world-view. I suspect that few people would react to the cases and principles and so on that Unger presents as he says "we" do. In the end I think we have to say that "we" are moderately well educated people of European descent (cultural if not biological descent) who have abandoned the religions of their ancestors and who have been exposed to little in the way of philosophical theorizing about the metaphysics of persons. (Religions tend to preach some form of dualism, and philosophical theorizing taints the data.)

Unger's single most important conclusion is that the data one gathers by these methods are best explained by supposing that we accept a "physically-based theory of our survival." His best attempt at a compact statement of the theory we accept is this:

> The person X now is one and the same as the person Y at some time in the future if, and only if, (1) there is a sufficiently continuous physical realization of a core psychology between the physical realizer of X's core psychology and the physical realizer of Y's core psychology, and [probably] (2) [insert here some clause suitable for ruling out unwanted cases of "branching"].

(A person's "core psychology" comprises the dispositional psychology that is common to everyone, or just about everyone: the capacity for consciousness and the ability to

make very rudimentary inferences. The "physical realizer" of a person's core psychology will normally be the person's brain, but this is not a necessary truth; indeed Unger is convinced that a person could start out with a brain as the realizer of his psychology and end up with a mess of electronic circuitry as its realizer—provided only that the transition from the one realizer to the other was "sufficiently continuous.") Thus Unger does justice both to our conviction—"we" are, of course, physicalists—that our survival has something to do with physical continuity, but also has something to do with the persistence of certain of the psychological capacities of the original person. Unger is, however, not entirely satisfied with the above statement as an articulation of our deepest beliefs about ourselves. On the psychological side, he is not entirely satisfied that an adequate articulation of our deepest beliefs about survival can get by without sometimes allowing a person's "distinctive psychology" to have some influence in deciding questions of survival. On the physical side, he thinks that this statement can yield the wrong results (results in conflict with our deepest beliefs about ourselves) if it is applied in cases in which parts of the physical realizer of one's psychology are replaced—as continuously as you like—so fast that there is insufficient time for "new" parts to be "assimilated" before they are replaced in their turn. I shall not discuss these difficulties.

If indeed "our deepest beliefs" are anything like what Unger suggests, our deepest beliefs would appear to have certain important logical and metaphysical consequences, among them these: since one's continued existence requires physical continuity, there can be no temporal gaps in one's existence; since there are terms that occur in the above statement that are inherently vague, we are "gradual beings."

I agree with Unger that we are physical beings, that we cannot exist again if there is a time at which we do not exist, and that we are "gradual"—or at least that we should be gradual in various conceivable circumstances. Nevertheless, I find that I must dissent from much of what he says. For one thing, my reactions to his imaginary cases are usually different from "ours": but this is not worth arguing about, for he will say, and I shall have to agree, that my reactions to his cases are influenced by my own theories. (I doubt whether my religious beliefs have much to do with my reactions to his cases, since they do not tend toward dualism.) I would also record my doubts about whether one *could* take a randomly selected group of, say, a thousand moderately well educated "secular" Americans and extract from them anything like an unequivocal endorsement of what Unger says are "our deepest beliefs about ourselves." But this is a sociological question. (If I understand Unger, he has operated mainly by introspection, perhaps eked out with a little Socratic interrogation of undergraduates.)

What I want to do in the remainder of this review is to raise some logical and metaphysical questions about "our deepest beliefs about ourselves"—that is, about the propositions that Unger says this description applies to, without (much) further discussion of the question whether the description is correct. If I am successful in showing that Unger's attempt at an articulation of our deepest beliefs faces logical and metaphysical difficulties, this will not, of course, show that Unger is wrong about what they are, for perhaps our deepest beliefs *do* face logical and metaphysical difficulties.

Suppose I am alone in a certain room. This room will, of course, contain a living human organism. What is the relation between me and this organism? I would say, and Unger would agree, that the relation is numerical identity. Unger, however, is committed to the view that I can be turned into a thing that is not an organism at all—say, by the gradual replacement of the neurons and other cells that make up my brain with

"functionally equivalent" inanimate gadgets: at the beginning of the replacement process, the realizer of my core psychology is a brain: somewhat later, it is a thing partly composed of brain cells and partly of inanimate gadgets; when the process is complete, my core psychology—indeed, the whole of my psychology— is realized in a thing that is composed entirely of inanimate gadgets. And if the realizer of my psychology can be turned into an inorganic object, so can I; it is necessary only to replace the rest of me with something wholly inorganic, or to eliminate it entirely. But if I am a thing that can be turned into an inorganic object, and if I am identical with a certain living organism, then this organism is a thing that can be turned into an inorganic object. If Unger is right, there are possible situations in which someone points at a certain mass of metal, plastic, glass, and silicon and says truly, "See that machine? It used to be a living flesh-and-blood organism." There are ways to take the sting out of this conclusion. One could, for example, contend that "identity is always relative to a sortal term" and say that, although the machine and the organism are the same person, they are not the same physical object. But Unger does not want to take the sting out. He is perfectly happy with the general conclusion that persons can slip with ease (logical or metaphysical, if not practical or physical ease) from one broad physical-object category to another. A ship must always be a ship, or at least something not entirely unlike a ship, and a worm can never be anything very different from a worm; certainly a ship can never become a mouse, and a worm can never become a computer. You and I, however, can be living organisms at one time and machines (metal and plastic and glass and silicon artifacts) later, simply in virtue of the fact that we have core psychologies whose realizers can be gradually changed from a physical thing of one kind to a physical thing of another kind. (But Unger's statement of our deepest beliefs does not require that the physical realizer of one's core psychology at a given time be a thing that can "change physical categories." It does not require that my brain be potentially a mass of metal and silicon. The statement of our deepest beliefs does not require this because it does not require that the physical realizer of my core psychology when I am a machine be numerically identical with the physical realizer of my core psychology when I am an organism.) This seems to me to be a *reductio* of Unger's position. But, of course, it doesn't seem that way to Unger. That's how philosophy works, I suppose. We are all, as David Lewis likes to say, trying to get our own views into equilibrium, and different philosophers have different points of equilibrium. (Nevertheless, I continue to be suspicious of Unger's claim that *his* point of equilibrium coincides with "our deepest beliefs about ourselves.")

The second logical or metaphysical point I wish to make has to do with the thesis that we are gradual beings. To appreciate this point, let us first consider what Unger calls "the Spectrum of Congenial Decomposition with Reconstruction." The Spectrum is a sequence of possible surgical operations that we might imagine a particular person—you, say—being forced to undergo. The first term of the sequence is the following operation: a single cell is removed from your brain and replaced with a precisely similar cell. The second term consists in the simultaneous removal and replacement of that cell and one other with precisely similar cells. "And so on": the final term in the sequence, of course, consists in the simultaneous removal and replacement of every cell in your brain with precisely similar cells. The Spectrum is called 'congenial' because of the following further requirement: when each term is constructed, the "next" cell, the one that is first mentioned in that term of the sequence, is to be the one whose removal and replacement (along with the cells whose removal and replacement is dictated by the previous term in

the sequence) would be most "congenial" in those circumstances to your survival and continued personhood—if there *is* a particular cell that has this feature, that is; if two or more cells are equally good candidates, by this criterion, for specification in the next term of the sequence, then one of them is to be chosen at random.

Unger believes, and I agree with him, that you would definitely survive the first operation in the Spectrum and that you would definitely not survive the final operation in the Spectrum. Unger believes, and I agree with him, that any two adjacent terms in the Spectrum describe very similar operations. (I would describe them as radically similar or prodigiously similar.) Now one would expect that a philosopher who believed that we were "gradual beings" would deny that there was a pair of adjacent terms in the Spectrum such that if you were to undergo the operation described by the earlier of them you would survive, and such that if you were to undergo the operation described by the latter you would fail to survive. (Call such a pair a "straddle pair.") Curiously enough, however, this is not Unger's position. His position is that there *is* a straddle pair ("the last bare survival case" and the "first near miss case"). There is a straddle pair because we have certain conventions that describe the "boundaries" of a person, and the one case falls just inside these boundaries and the other just outside.

Why does Unger think that such conventions exist? The argument seems to be this: "Since we are gradual beings, the existence of a straddle pair can only be a matter of convention. And there must be a straddle pair." But why must there be a straddle pair? Unger's only argument for the conclusion that there must be a straddle pair is a brief passage on p. 220: *"Given that these beliefs are correct,* it is a matter of mathematics and logic that there is, in the spectrum, a *first* case where it is *not true* that I survive." (Italics in original.) And what are "these beliefs"? I quote, because I am not confident of my ability accurately to paraphrase Unger's words: ". . .as regards there emerging someone with good qualifications for being me, we believe that, in this spectrum of cases, there is a *general decrease*, starting from the top end and proceeding toward the bottom end. Finally, as we believe, there is not only a *finite number* of cases in this spectrum, but there is also a *[greatest] lower bound* on the relevant decrements." Insofar as I can claim to understand these words, they seem to be a way of saying that there has to be a first case—the greatest lower bound—in the Spectrum in which the patient does not survive.

In my view, there are weighty reasons for denying the existence of these conventions. First, it seems conceptually impossible for there to be conventions governing the boundaries of things. Secondly, if this is mistaken, it nevertheless does not seem to be *necessary* that we have conventions of this type, and it is very hard to see why we should bother to establish any. Thirdly, even if such conventions were in principle possible and we for some reason wanted to establish one of them, it would seem that, for practical reasons, it would be impossible for us to do so.

As to the first point, we should note that all that conventions can do is regulate behavior—including, of course, linguistic behavior. Sometimes people say things like this: "The 'end of the Roman Empire' is conventionally said to have occurred in 410 A.D. when Alaric's Visigoths sacked Rome." Well, it may be handy to have a date for the end of the Roman Empire (although I am not sure why), but the fact of the matter is that the Roman Empire was a temporally vague entity, and no convention can change this fact: a convention to the effect that the Roman Empire ended in a certain year can be nothing other than an agreement to pretend, in certain circumstances in which the falsity of what is pretended does not matter, that the Empire was something it was not: a

temporally precise entity. (Or perhaps there are no temporally vague entities; perhaps there are only a lot of temporally precise entities that are all equally good candidates for the office of "Roman Empire." In that case, *that* is the fact of the matter, and no convention can make one of those temporally precise entities the Roman Empire. In that case, a convention to the effect that the Roman Empire ended in a certain year can be nothing other than an agreement to pretend that there was a unique entity that was the Roman Empire.)

Or so I believe.[1] Suppose I'm wrong. Suppose that it is possible in principle to establish the boundaries of the Roman Empire or of a particular human person by convention. Surely there is no necessity that forces us to establish a boundary of any particular degree of sharpness around any entity. We shall, presumably, do so if there is some purpose to be served by doing so, and, if doing so will serve no purpose, then we shall not bother to. If it were possible for us to establish a conventional boundary around Unger (say) that was so precise that an adjacent pair of terms in a Spectrum of Congenial Decomposition lay on opposite sides of this boundary, why should we in fact bother to do such a thing? What human purpose would be served by doing so? Wouldn't it be like dating the end of the Roman Empire to the nearest nanosecond? If people actually underwent operations like those specified in a Spectrum of Congenial Decomposition, then, conceivably, some human purpose would be served by extremely sharp conventional boundaries around human beings. But they don't and none is.

Finally, suppose that it is conceptually possible to draw conventional boundaries around things and there is some reason to draw boundaries as precise as those that Unger says that we have in fact drawn around ourselves. Suppose we haven't yet done this and that now we want to. How shall we go about it? We may think of Spectra of Congenial Decomposition as set-theoretic constructs on a particular person's brain-cells. There are obviously a vast number of such constructs on *your* brain-cells. What Unger asserts the existence of, if I understand him, is a convention that assigns a straddle pair to every one of them. Well, always assuming that conventional boundary-drawing is possible at all, there would be no difficulty in establishing a convention that satisfied this condition. One could, for example, lay it down that in every Spectrum, the straddle pair consisted of the ten-billionth and the ten-billion-and-first terms. But that would be a pretty silly convention. Even if we assume that the congeniality requirement entails that, in any two Spectra, the ten-billionth and ten-billion-and-first terms involve cells whose function is of about the same relevance to questions about personal identity, we must still ask what principle, if any, led us to choose that particular pair of numbers. I am not raising the question why we chose that pair of numbers rather than ten-billion-and-one and ten-billion-and-two. *That* question could be legitimately answered by saying that the choice was an arbitrary one. But what principle led us to choose those two numbers rather than 20, 087, 510, 338, and 20, 087, 510, 339? I can see no possible answer to this question. And is it even faintly plausible to suppose that there exists a number n such that it would be reasonable to establish the convention that the straddle pair for *every* Spectrum of Congenial Decomposition consists of the nth and $n + 1$st terms in the Spectrum? It is hard to believe that the answer to this question is Yes. It seems to me that any convention that assigned a straddle pair in a non-silly way to every one of your Spectra would have to be based on physiological knowledge that no human being actually possesses and that it would in any case be far too complicated for any human being to learn—or even to write down. (A hard copy would probably look something like the Encyclopedia Britannica.)

When you put these three difficulties together, it seems to be very hard to believe that we are conventionally demarcated entities, particularly if the demarcation is supposed to be so sharp that the removal of a single neuron can make the difference between one's existence and one's non-existence. It seems to me to be much more realistic to suppose that we are naturally demarcated entities and that the natural demarcation is vague. It should, however, be clearly understood that if I am right about this, it does not have any consequences for Unger's theses about our deepest beliefs about personal identity across time, for these are logically independent of his thesis that we are entities whose boundaries in time and space (and in logical space) have been established, and made very sharp, by convention.

Whatever objections anyone may want to raise against particular theses and arguments of Unger's book, everyone who cares about the problem of personal identity should attend carefully to it. No one who accepts a "psychological continuity and connectedness" theory of personal identity, or who (like me) accepts a purely physical account of personal identity should avoid coming to terms with Unger's powerful criticisms of theories of those types. And there does seem to be something awfully attractive about Unger's attempt to incorporate both psychological and physical continuity into his account of personal identity. Someone who, like me, has excluded any psychological component from his theory of personal identity can none the less feel very strongly the force of Unger's presentation of the thesis that psychology has to be regarded as relevant to questions of personal identity.

This is clearly a book to be reckoned with.[2]

Notes

[1]For further arguments for the thesis that it is conceptually impossible to establish the boundaries of things by convention, see my book *Material Beings* (Cornell University Press: Ithaca NY, 1990), pp. 6-12.

[2]I am grateful to Peter Unger and Gregory Ganssle for help with this review.