ABSTRACT. Much of the literature on the question "Is a human essentially distinct from every possible machine?" proceeds on the assumption that we know what a man essentially is, namely a living body with such attributes as consciousness, freedom, feeling and linguistic competence. Is a man essentially that? The paper contrasts that picture of man with Kierkegaard's account of man as essentially self. Hard limits of machine subjectivity begin to appear in the failure of certain everyday concepts involving 'self' to engage at all with the concept 'machine'.

The question to be decided is whether every possible machine is essentially distinct from a man.[1]

The question does not offer much traction except on the assumption that we have a satisfactory account of what a man essentially is, since the phrase 'every possible machine' is as open-ended and elastic as our imaginations, which produced The Stepford Wives, Westworld and many another saga of soft-skinned robots. However, the disputants I have read in this area express little puzzlement about what essentially a human is. But what then am I? A language-user, of course, alive, conscious, thinking, free and not unfeeling. These tend to be the familiar features of man that figure centrally in writings on this topic, but do these features plus a human body constitute a man essentially? Too quick a Yes is dangerous, if only because important words such as thinking, consciousness, freedom, life and feeling have so often been lifted out of what Wittgenstein calls 'their natural homes' and dealt with strangely by philosophers and psychologists, with the result that we can't be entirely sure of what we're getting when we package them into an account of man's es-
HARRY A. NIELSEN

sence. Sentences like the following are currently in print: "... I now believe that it is possible so to construct a super-computer as to make it unreasonable to deny that it had feelings."[2]

Though Scriven's 'question to be decided' is not formed well or fully because of the faint question mark hanging over what essentially a man is, we need it, or something like it, to position ourselves toward remarks of this type: "... there is no reason to suppose that machines have any limitations not shared by man."[3] We also need, I think, to address the question of what essentially a man is, and because I feel less than supremely confident about doing that, I want to borrow a few sentences from the opening paragraph of Kierkegaard's The Sickness Unto Death: "Man is spirit. But what is spirit? Spirit is the self. But what is the self? The self is a relation which relates itself to itself, or it is that in the relation (which accounts for it) that the relation relates itself to its own self; the self is not the relation but (consists in the fact) that the relation relates itself to its own self..."[4]

Labyrinthine and not exactly blue-ribbon prose, the last sentence can be unpacked, I believe, by linking up the 'relation' constitutive of the self to certain complexities of an individual's self-dialoguing.[5] More directly important for today's purpose is what Kierkegaard goes on to say about the human self which, for him, is essentially man. As it relates itself to itself, my self over a period of time comes to a more or less hard-finished judgment or position concerning what I think it is worth to be the one I am, or what I really think of the one I am.

Bear in mind that this judgment, whether it finds expression in words or some other medium, deals not with casual matters or headlines, but strictly with what I think of myself as one of the family of man. The individual's judgment itself may be yea-saying: I can be utterly thankful for being the one I am. It can be negative, even to the point of fiercely resenting having to be this. Or conditional: "I find it tolerable to be the one I am, provided I get generous helpings of certain treats from life's buffet." An individual's judgment along these lines may have various expressions and maskings. If I read Kierkegaard correctly, this firming up of 'what I really think' is a distinctive feature of the human self, and something which each of us, given enough time, brings himself around to. So far it would appear that the individual can check this out for himself; if indeed I harbor an evaluative opinion of being the one I am, I should be able to find it in some corner of the knapsack, though this will scarcely count as evidence that everyone else has such an opinion.

At this point we can imagine someone saying, "But technology can build that kind of wrinkle into an Artificial
Intelligence unit! I see no contradiction.” Maybe there is none, but is contradiction the only way a notion can misfire? Let us consider this by steps. Technicians can make a machine print out anything they want it to, and perhaps build one that ‘dialogues’ with itself in a pretty intricate way, for instance like consulting partners playing chess against a stronger player. All the same we can notice that numerous expressions we apply to persons sound very sticky when applied to parcels of hardware. I list a few:

', ... puts up with slights and insults'
', ... lacks common decency'
', ... holds a grudge'
', ... is terribly secretive'.

The fact that we use such expressions to speak of aspects of another human's subjectivity, but do not apply them to machines, falls some distance short of answering Scriven's question about 'every possible machine', though some philosophers talk as if that grammatical fact closed the issue.[6]

What would it be like to construct a unit which, in addition to playing difficult games, predicting elections, extracting cube roots, and whatever other tasks we engineer into it, also forms the kind of judgment about itself that we spoke of earlier? What would it be like to program into that unit a tendency to attach itself--gradually and without messing up its other functions--to an unshakeable opinion of itself? For example, what would it mean to program a unit that would come to think it an outrage to be the unit it is? Or a unit that is sick of having to be the one it is, and yearns to be rid of itself? A unit whose silences are charged with resentment? Or, for that matter, a unit that wouldn't trade places with any other in the world? These are expressions we apply only to ourselves and one another.

Here it may be enticing to imagine we can knife through these difficulties by saying, "Well, all we need do is to build language into a machine. That will make it as complex as you wish." (What would make someone think complexity is the difference between a man and a machine?[7]) And if complexity does figure in some important way, why should it be the kind of complexity we associate with language or grammar? Or the kind we associate with the microstructure of the brain and nervous system?[8] Why not the kind of complexity (if that is what it is) that occasions our finely shaded talk about the various happy and unhappy states that human subjects get into? The idea of relative complexity can be an alluring convenience for characterizing the difference between man and machine in the absence of a clear view of what man essentially is.
Seen in terms of Kierkegaard's account, a human is an organism that is constrained to form a judgment of its individual worth. A modification of the person occurs through his coming to be of this or that opinion of himself. This much, it seems to me, could serve as the basis of a wholly secular psychology in addition to the purpose Kierkegaard gives it. What strikes me as valuable is his recognition of conflict-of-interest in the spectacle of a man telling us what man essentially is by listing some obvious human features and letting it go at that, thus making an end-run around the pathos of being an individual, hence indefinable. Kierkegaard seeks to answer the question "What is man?" but in the sense of "What is it essentially to be in the estate of man?" or "What is it not only to be one of these but to have no choice but to be this one?"

Do these reflections enable us to specify some limits of machine subjectivity or selfhood that are not shared by man? Others may know better, but I for one am flatly unable to make a conceptual connection between a machine and certain elements in Kierkegaard's account of the self. (Perhaps one reason is that his account allows for inward modification of a subject with no necessary behavioral expression, no necessary changes in outward demeanor or habits.) Kierkegaard deals with the self in terms of the tonalities of despair, which may or may not fight their way to the visible surface of life, but the same point can be made about the tonalities of joy, happiness, peace, resignation and the like, all of them concepts to which we have given no meaning in connection with machines. Can the space-age make any sense of:

(a) designing a machine that will function as intended and will also be at peace with itself?

(b) designing one that will turn out to be its own worst enemy?

(c) designing one that will resign itself to its inadequacies?

(d) designing one that will not only do its job but be grateful for the chance?

(e) designing one that can put up with prolonged solitude?

In our way of looking at it, 'self' rather than 'man' is the concept that fails most immediately to engage with 'machine'. For the most part the literature on this topic stays away from areas of the language of 'self' such as those in (a) - (e), working instead with grander--in a sense grosser--concepts such as consciousness, freedom, feeling, life, language, and even self-awareness, though in a most
COMPUTER SUBJECTIVITY

gingerly manner.[9] This is one reason, I believe, why the already rather bloated literature of Yea, Nay and distinguo continues to swell. This reason plus the fact that the literature contains a certain detectable amount of pretending to know what man essentially is, as defined by those larger concepts, makes it useful to turn to someone like Kierkegaard who ventures a head-on address to the question of what man is. Even if he wrong, his effort can remind us that maybe nobody has got it quite right.

FOOTNOTES

* This paper was read at "Apollo Agonistes", an international symposium sponsored by the Institute for Humanistic Studies at SUNY (Albany) in April, 1979. Permission granted.


2. Scriven, op. cit., 42.


8. For an account of difficulties in the language-programming of computers, see H. Dreyfus, What Computers Can't Do: A Critique of Artificial Reasoning (New York, 1972), 44-57, 126-33. Knowing how to build certain linguistic elements into an AI unit, e.g., programming an automaton that can scan the records of chess games and move the pieces accordingly—is this the same as knowing how to build into a machine any designated use of language? In the same vein, what does it take to make a silence a-pregnant one?

9. For example A. M. Turing, speaking to the claim that a machine cannot be the subject of its own thought, writes "... It may be used to help in making up its own thought, or to predict the effects of alterations in its own structure. By observing the results of its own behavior it can modify its own programs so as to achieve some purpose more effectively ...." Computing Machinery and Intelligence in Anderson, op. cit., 20.