

## WITTGENSTEIN AND THE PROBLEM OF MACHINE CONSCIOUSNESS\*

J.C. NYIRI

University of Budapest

In the course of Wittgenstein's repeated attempts at a systematic exposition of his ideas, from the *Blue Book* to the *Philosophical Investigations*, it was invariably the arguments relating to *meaning as use* which set the stage for the rest of the discussion. It was precisely the insight conveyed by these arguments which, in 1930-31, led to an over-all philosophical breakthrough in Wittgenstein's own development — to the antimentalism and anti-essentialism characteristic of his later views. "Thinking is the use of symbols", he puts down in his notebook on June 29, 1930<sup>1</sup>, adding, some pages later: "Thought, insofar as one can speak of it at all, must be something quite pedestrian."<sup>2</sup> Now the philosopher who first introduced the idea that meaning is *reference*, and faced the ensuing difficulties bound up with terms which are meaningful but do not, apparently, refer to anything, was of course Plato. And it is a striking fact that Wittgenstein, whose comments on earlier figures in the history of philosophy are notoriously scarce, often mentions, and indeed quotes, Plato.<sup>3</sup> In the 1930-31

\*I am indebted to the Alexander von Humboldt Stiftung under whose auspices, in 1986-87, most of the research embodied in the present paper was conducted. The argument as here outlined was first presented at the International Summer School "Philosophical Perspectives on Artificial Intelligence" at Bolzano, in July 1988.

1. "Das Denken ist das Gebrauchen von Symbolen", MS 108, p.201. Passages from Wittgenstein's *Nachlaß* are cited according to the numbering given in G.H. von Wright, "The Wittgenstein Papers", in von Wright, *Wittgenstein*, Oxford: Basil Blackwell, 1982.

2. "Der Gedanke, soweit man überhaupt von ihm reden kann, muß etwas ganz hausbackenes sein", *ibid.*, p.216.

3. "Wittgenstein", writes von Wright, "had done no systematic reading in the classics of philosophy. He could read only what he could wholeheartedly

notebooks, in particular, Wittgenstein repeatedly refers to passages in Plato where the connections between a misguided theory of meaning on the one hand, and an extravagant ontology on the other, are especially conspicuous. Thus for example the important entry of July 9, 1931 — “To understand the meaning of a word means to be acquainted with, to understand, a use”<sup>4</sup> — is followed, on the next page, by a quote from the *Cratylus* on the assumed, or desirable, likeness between signs and the objects they stand for<sup>5</sup>, and, on the page after, by a quote from the *Theaetetus* on the particular state of affairs that anything represented must be, somehow, real.<sup>6</sup> The latter elicits a comment from Wittgenstein: “How infinitely simple this problem is! And how strange that it could have been regarded as a problem at all”<sup>7</sup>, followed, on the next day, by the remark: “I do not find in Plato to a question like ‘what is knowledge’ the preliminary answer: Let us try and see how this word is used”.<sup>8</sup> It is as if Wittgenstein were conscious of the fact that he has, in philosophy, to start anew at the

---

assimilate. ... [A]s a young man he read Schopenhauer. From Spinoza, Hume, and Kant he said the he could only get occasional glimpses of understanding. I do not think that he could have enjoyed Aristotle or Leibniz, two great logicians before him. But it is significant that he did read and enjoy Plato.” (“Ludwig Wittgenstein: A Biographical Sketch”, in von Wright, *Wittgenstein*, p.33.)

4. “Die Bedeutung eines Wortes verstehen, heißt, einen Gebrauch kennen, verstehen”, MS 111, p.12.

5. “Unsere Weise von den Wörtern zu reden können wir durch das beleuchten was Sokrates im ‘Kratylos’ sagt. Kratylos: ‘Bei weitem & ohne Frage ist es vorzüglicher, Sokrates, durch ein ähnliches darzustellen, was jemand darstellen will, als durch das erste beste.’ — Sokrates: ‘Wohlgesprochen.’ ”

6. “Sokrates zu Theaitetos: ‘Und wer vorstellt, sollte nicht etwas vorstellen?’ — Th.: ‘Notwendig.’ — Soc.: ‘Und wer etwas vorstellt, nichts Wirkliches?’ — Th.: ‘So scheint es.’ ”

7. “Wie unendlich einfach dieses Problem! Und wie seltsam, daß man es überhaupt als Problem konnte ansehen wollen”, *ibid.*, p.14, entry of July 14, 1931.

8. “Ich finde bei Plato auf eine Frage wie ‘was ist Erkenntnis’ nicht die vorläufige Antwort: Sehen wir einmal nach, wie dieses Wort gebraucht wird”, *ibid.*, pp. 26f.

very juncture where Plato's arguments went astray.<sup>9</sup> "What marvelous people we are", he puts it in 1931, "to have solved these ancient problems! — No, time has changed us & the problems have vanished."<sup>10</sup>

What the precise nature of those relevant changes might be is a question with regard to which Wittgenstein did not, it seems, develop any very clear ideas. But he obviously had the right kind of intuitions. At about the time when he reflected on Plato and on the ways in which linguistic appearances could give rise to a wrong philosophy, he made extensive notes on Frazer's *The Golden Bough* — a book dealing with primitive peoples, with myths, and, as Wittgenstein saw it, with *language*, and in particular with techniques of communication essentially different from our own.<sup>11</sup>

### *Orality and Literacy*

Continuing a line of scholarship that began with the work of Milman Parry on the text of the *Iliad* and the *Odyssey*,<sup>12</sup> Jack Goody in his *The Domestication of the Savage Mind*<sup>13</sup> has plausibly shown how changes

9. As von Wright has formulated it: "The later Wittgenstein has no ancestors in the history of thought. His work signals a radical departure from previously existing paths of philosophy. ... The *Tractatus* belongs in a definite tradition in European philosophy... Wittgenstein's so-called 'later philosophy' ... is quite different. Its *spirit* is unlike anything I know in Western thought and in many ways opposed to aims and methods in traditional philosophy" ("Biographical Sketch", *op. cit.*, p.27).

10. "Was für großartige Menschen wir sind diese alten Probleme gelöst zu haben! — Nein die Zeit hat uns geändert & und die Probleme sind verschwunden", MS 159, p.94.

11. As Wittgenstein comments: "In den alten Riten haben wir den Gebrauch einer äußerst ausgebildeten Gebärdensprache", MS 110, p.256, cf. Ludwig Wittgenstein, "Bemerkungen über Frazer's *The Golden Bough*", *Synthese* 17 (1967), p.242.

12. Another, albeit less obvious, source is the work of the Hungarian historian István Hajnal. His "Le rôle social de l'écriture et l'évolution européenne", *Revue de l'Institut de Sociologie* (Bruxelles), 1934, is referred to in *The Bias of Communication* by Harold A. Innis (University of Toronto Press, 1951). Recently he is mentioned in Walter J. Ong's admirable book *Orality and Literacy: The Technologizing of the Word* (Methuen, 1982).

13. Cambridge: Cambridge University Press, 1977.

usually characterized as shifts from magic to science, or from 'prelogical' to 'rational' states of consciousness, or from Lévy-Strauss's 'savage' mind to domesticated thought, can be far better explained as shifts from an *oral culture* to various stages of *literacy*. For any given society, its particular technology of communication has far-reaching consequences, not merely as regards social organization, but on the epistemic level as well. As Walter J. Ong formulates it, discussing the psychodynamics of cultures untouched by literacy ("primary orality"): "In an oral culture, restriction of words to sound determines not only modes of expression but also thought processes."<sup>14</sup> If one is to remember what one experiences and thinks, one has to *think memorable thoughts*.

In a primary oral culture, to solve effectively the problem of retaining and retrieving carefully articulated thought, you have to do your thinking in mnemonic patterns, shaped for ready oral recurrence. Your thought must come into being in heavily rhythmic, balanced patterns, in repetitions or antitheses, in alliterations and assonances, in epithetic and other formulaic expressions, in standard thematic settings ..., in proverbs which are constantly heard by everyone so that they come to mind readily... Fixed ... expressions ... in oral cultures ... form the substance of thought itself. ... Heavy patterning and communal fixed formulas in oral cultures serve some of the purposes of writing in chirographic cultures, but in doing so they of course determine the kind of thinking that can be done, the way experience is intellectually organized. In an oral culture, experience is intellectualized mnemonically.<sup>15</sup>

With the invention of alphabetic writing a new technology for storing knowledge emerged. The psychological implications were marked. "Writing", as Ong puts it, "introduces division and alienation, but a higher unity as well."<sup>16</sup> Writing so to speak separated the knower from the known, intensified the sense of self, it "freed the mind for more original, more abstract thought".<sup>17</sup>

14. *Orality and Literacy*, p.33.

15. *Ibid.*, pp.34ff.

16. *Ibid.*, p.179.

17. *Ibid.*, p.24. — Literacy as an indispensable, although ambivalent, framework for individuality and creativity is eloquently defended in a recent paper by Elisabeth List, "Schreiben und Lesen. Ein Plädoyer für die literale Kultur vor deren Ende", *DOXA* 14 — *Semiotische Berichte* 12/1-2 (1988), cf. esp. pp.57f.

The Greek alphabet was definitely established by 700 BC, but it was not until Plato's days that the Greeks had at long last effectively assimilated writing.<sup>18</sup> Plato himself expresses serious reservations in the *Phaedrus* and his *Seventh Letter* about writing, "as a mechanical, inhuman way of processing knowledge, unresponsive to questions and destructive of memory".<sup>19</sup> But it was precisely the deeply interiorized experience of writing, of *written language*, that essentially shaped Plato's own philosophy. The ascendancy of Greek analytic thought, generally speaking, seems to be intimately bound up with the Greeks' introduction of vowels into the alphabet — a development creating a new level of abstract, visual coding as regards the world of sound.<sup>20</sup> Plato's theory of knowledge, in particular, shows the direct impact of literacy. As Havelock points out, the term "idea", form, is *visually based*<sup>21</sup> — coming, actually, from the same root as the Latin term "video". Platonic form, as Ong puts it, "was form conceived of by analogy with visible form".<sup>22</sup>

Platonic ideas are linked to the experience of written language in yet another way. As indicated earlier, they provide, or promise to provide, an answer to all those perplexing questions which arise on the soil of a name-theory of meaning — the theory according to which words are typically *names*, and the meaning of a word is the object to which the word *refers*. Now the experience of *spoken* language — of language *alive*<sup>23</sup> — would hardly suggest a name-theory of meaning. In oral

18. "Between Homer and Plato", writes Eric Havelock, "the method of storage began to alter, as the information became alphabetised, and correspondingly the eye supplanted the ear as the chief organ employed for this purpose. The complete results of literacy did not supervene in Greece until the ushering in of the Hellenistic age, when conceptual thought achieved as it were fluency and its vocabulary became more or less standardized. Plato, living in the midst of this revolution, announced it and became its prophet." *Preface to Plato*, Cambridge, Mass.: 1963, p.vii. — Havelock stresses that it took a great deal of time for the new technique to become generally applied; on pp.49ff. he provides a survey of the heated discussions surrounding the question as to when the Greek alphabet was introduced.

19. Ong, *op. cit.*, p.24.

20. Cf. Eric Havelock, *Origins of Western Literacy*, Toronto: 1976.

21. *Preface to Plato*, cf. esp. pp.261f.

22. *Op. cit.*, p.80.

23. "Im lebendigen Gebrauch der Sprache", MS 111, p.7.

discourse words are indeed experienced as instruments, as deeds, as speech acts within broader, non-linguistic situations. Ong points to the "startling paradox" that writing is traditionally, ever since the time when the New Testament was composed, so closely associated with death: "The letter kills but the spirit gives life" (2 Cor. 3,6).<sup>24</sup> Wittgenstein remarks: "Every sign *by itself* seems dead. *What* gives it life? — In *use* it is alive."<sup>25</sup> There is a deceptive appearance giving rise to philosophy, but this appearance is essentially bound up with a kind of *reality* — a reality Wittgenstein was keen to explore<sup>26</sup> — a reality which is now unmasked as *written language*. In developing a use-theory of meaning, Wittgenstein in effect reintroduces into philosophy a feeling for language as oral discourse. And there are traces in his work which show that he was not entirely unaware of what he was doing. His use of the dialogue form is of course a conspicuous stylistic feature — and it is significant that it emerges, in 1930, simultaneously with the development of the use-theory of meaning. In the *Philosophical Investigations* one finds quite a number of allusions to the basically oral character of language.<sup>27</sup> Another important pointer is Wittgenstein's emphasis on everyday language, and indeed on the vernacular. Wittgenstein's *Wörterbuch für Volksschulen*, compiled in the early 1920s in the course of his activity as an elementary school teacher in Lower Austria, deliberately reflects the pupils' own linguistic usage; it does not avoid dialectal expressions, and includes some very common words which are typically used in *speech*, like "geh!" or "hierher". It even utilizes dialectal *pronunciation* in order to bring home some grammatical points.<sup>28</sup>

24. *Op. cit.*, p.81.

25. *Philosophical Investigations*, I, §432.

26. "In der Philosophie werden wir durch einen Schein getäuscht. Aber dieser /ein/ Schein ist auch etwas, & ich muß ihn einmal ganz klar mir vor Augen stellen, ehe ich sagen kann, daß es nur ein Schein ist", MS 110, p.239, entry of June 30, 1931.

27. Cf. e.g. I, §§27 and 134 — but also §167, where Wittgenstein's alertness to the peculiarities of the printed word is quite manifest.

28. "*ihm*, in der Mundart: 'eam', z.B.: 'I hob *eam* g'sogt' — *ihn*, in der Mundart: 'n' oder 'm', z.B.: 'I hob *m* g'sehn' ", *Wörterbuch für Volksschulen* (1926 — repr. Vienna: Hölder-Pichler-Tempsky, 1977, eds. A. Hübner et al.), p.15. — In the (originally unpublished) preface to the dictionary Wittgenstein

The traditional role of dictionaries is of course a very different one: they are instrumental in isolating the national written language, the "grapholect", from its original dialect base. "The lexical richness of grapholects", Ong stresses, "begins with writing, but its fullness is due to print. For the resources of a modern grapholect are available largely through dictionaries."<sup>29</sup> Printing actually introduces a new phase of literacy. The letters used in writing, Ong suggests, do not actually exist prior to the text in which they occur. With print it is different.

Words are made out of units (types) which pre-exist as units before the words which they will constitute. Print suggests that words are things far more than writing ever did. ... it was print, not writing, that effectively reified the word and, with it, noetic activity ... Hearing rather than sight had dominated the older noetic world in significant ways, even long after writing was deeply interiorized. Manuscript culture in the west remained always marginally oral. ... Writing served largely to recycle knowledge back into oral world.<sup>30</sup>

Print, Ong goes on to point out, was "a major factor in the development of the sense of personal privacy that marks modern society".<sup>31</sup> Certainly it was with print that *solo reading* and *silent reading* became the general practice.

It has become a truism that today we are witnessing yet another transformation in the technology of communication; and it is a major achievement of Ong to have turned this truism into a philosophically interesting thesis. We have entered, he writes, "with telephone, radio, television and various kinds of sound tape", an age of "secondary orality".<sup>32</sup> Now it is of course the computer that has become the essential instrument of this age.<sup>33</sup> With the advent of the word

---

makes the remark: "Again and again psychological principles (where will the student look for the word, how does one guard him against confusions in the best possible manner) clash with grammatical ones (base word, derivative) and with the typographical utilization of space, with the well-organized appearance of the printed page, etc.", *ibid.*, p.XXXV, transl. by Elisabeth Leinfellner.

29. *Op. cit.*, p.107.

30. Ong, *op. cit.*, pp.118f.

31. *Ibid.*, p.130.

32. *Ibid.*, p.136.

33. Aaron Sloman speaks of "the computer, and the processes which run

processor, language has once more become "dynamic rather than static, malleable rather than fixed, soft rather than hard, plastic rather than rigid".<sup>34</sup>

This is a picture of language the later Wittgenstein could well have agreed with; and it strikes one that his use-theory of meaning, and indeed his later philosophy as a whole, are, in a sense, very much in harmony with our times. That Wittgenstein detested the present age is of course obvious, and has often been emphasized. He ridiculed the idea that human progress could be fostered by advances in technology; however, he deeply enjoyed engineering problems of all kinds. He was concerned to isolate his philosophizing from the shallow influence of science; however, as S.G. Shanker recently pointed out, he by no means wanted to "consolidate a 'two cultures' division", and "emphasized that the role of modern philosophy will (or at least should) become increasingly bound up with the interpretation of scientific discoveries".<sup>35</sup> He was a traditionalist, yearning for bygone times; however, since there is a strong connection between orality and

---

on it, as a new medium of self-expression, perhaps comparable in importance to the invention of writing". *The Computer Revolution in Philosophy: Philosophy, Science and Models of Mind*. Harvester Press, 1978, p.1.

34. G. Richard Dimler, S.J., "Word Processing and the New Electronic Language", *Thought*, vol.61, no.243 (Dec. 1986), p.463. — More broadly, electronic word processing marks a second stage in the "externalization of knowledge" (cf. Henrik Sinding Larsen, "Information Technology and the Management of Knowledge", *AI & Society*, vol.1, no.2 [Oct.-Dec. 1987], pp.97ff.), a stage in which knowledge, though externalized — or, more precisely: *objectified* — can once again become dynamic, a "social memory both extensive and active", as Alvin Toffler puts it in *The Third Wave*, New York: 1980, p.177.

35. Shanker, "The Decline and Fall of the Mechanist Metaphor", in Rainer Born (ed.), *Artificial Intelligence: The Case Against*. London: Croom Helm, 1987, p.73. — Significant is Wittgenstein's following remark: "Perhaps one day a culture will emerge from this civilization. — When that happens there will be a real history of the inventions of the 18th, 19th and 20th centuries, which will be deeply interesting." Ludwig Wittgenstein, *Culture and Value*, ed. G.H. von Wright, transl. by Peter Winch, Oxford: Basil Blackwell, 1980, p.64e. I have slightly changed the translation.

traditionality<sup>36</sup>, his nostalgia could be plausibly interpreted as a sensitivity to the as yet hidden tendencies of the present, or as a premonition of the imminent future. And he was certainly not immune to the charms of the new media. His favourite distraction, in Cambridge in the 1930s, must have meant, at the same time, an important audio-visual experience: the "flick"<sup>37</sup> — or, to spell it out, the *sound film*.

Wittgenstein died before the first computers gained public attention. But there are reasons to believe that Alan Turing's pioneering paper "On Computable Numbers, with an Application to the Entscheidungsproblem" (1937), containing a first abstract description of the principles of a digital computer, was not independent of Wittgenstein's influence.<sup>38</sup> And a problem to which Wittgenstein, throughout his later philosophy, again and again returned, was that of *machine consciousness*.

### *The Ascription of Intelligence*

The question whether machines can possess consciousness arises for Wittgenstein as an aspect of the general problem what our criteria for attributing mental predicates to some kinds of beings are. "Why can't we imagine", he asks in an early entry, "a machine with memory?" Or

36. Some relevant arguments are summarized in my paper "Tradition' and Related Terms: A Semantic Survey", *DOXA* 14 — *Semiotische Berichte* 12/1-2 (1988), cf. esp. pp.123ff.

37. Cf. Norman Malcolm, *Ludwig Wittgenstein: A Memoir*, London: Oxford University Press, 1958, pp.27f.

38. Turing's paper existed in a draft typescript by April 1936. According to Andrew Hodges he did not make Wittgenstein's acquaintance before 1937 (cf. *Alan Turing: The Enigma*, London: Burnett Books, 1983, p.136); however, in the volume *Ludwig Wittgenstein: Sein Leben in Bildern und Texten* (eds. M. Nedo and M. Ranchetti, Frankfurt/M.: Suhrkamp, 1983, p.358) he is reported to have attended Wittgenstein's class in 1935. Certainly there are some interesting parallels between Wittgenstein's approach to the philosophy of mathematics on the one hand, and the outlook of "Computable Numbers" on the other. I have touched on this issue in my volume *Gefühl und Gefüge: Studien zum Entstehen der Philosophie Wittgensteins*, Amsterdam: Rodopi, 1986, p.203.

one which thinks, or is in pain? It is clear, he continues, that one can imagine a machine which in all relevant respects behaves just like a human being in pain; or, put differently, there is no difficulty in *calling* the body of a human being a machine capable of experiencing pain.<sup>39</sup> The issue, Wittgenstein suggests, is a *grammatical*, not an empirical one. As he puts it in the *Blue Book*:

the problem here arises which could be expressed by the question: 'Is it possible for a machine to think?' ... And the trouble which is expressed in this question is not really that we don't yet know a machine which could do the job. ... The trouble is rather that the sentence, 'A machine thinks (perceives, wishes)': seems somehow nonsensical. It is as though we had asked 'Has the number 3 a colour?'"<sup>40</sup>

In the *Philosophical Investigations* the same argument reappears<sup>41</sup>, but with a different emphasis:

Only of a human being and what resembles (behaves like) a living human

39. "Warum können wir uns keine Maschine mit einem Gedächtnis denken? ... Nun ist das aber ganz so wie wenn man sagt, eine Maschine kann nicht denken. Oder kann keine Schmerzen haben. Und hier kommt es darauf an was man darunter versteht 'Schmerzen zu haben'. Es ist klar daß ich mir eine Maschine denken kann, die sich genau so benimmt (in allen Details) wie ein Mensch der Schmerzen hat. Oder vielmehr: ich kann den Andern eine Maschine nennen die Schmerzen hat[.] D.h.: den andern *Körper* und ebenso natürlich meinen Körper", MS 110, p.35 (Febr. 5, 1931). — For a closely related passage, cf. *Philosophical Grammar*, I, §64.

40. Oxford: Basil Blackwell, 1964, p.47.

41. Drawing on this argument, Anthony Kenny speaks of "the reckless application of human-being predicates to insufficiently human-like objects", taking issue with those "psychologists, physiologists and computer experts" who "take predicates whose normal application is to complete human beings or complete animals and apply them to parts of animals, such as brains, or to electrical systems" ("The Homunculus Fallacy", in Kenny, *The Legacy of Wittgenstein*, Oxford: Basil Blackwell, 1984, p.125). In a similar vein Shanker writes: "we are not in the least concerned with the empirical question of the extent to which machines can be developed which will simulate complex aspects of human behaviour... It is not that computers lack consciousness, it is that the concept of consciousness simply cannot be applied to a machine." ("Introduction: The Nature of Philosophy", in S.G. Shanker, ed., *Ludwig Wittgenstein: Critical Assessments*, vol. 4, London: Croom Helm, 1986, pp.11f.)

being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. ... Could a machine think? — Could it be in pain? — Well, is the human body to be called such a machine? It surely comes as close as possible to being such a machine. — But a machine surely cannot think! — Is that an empirical statement? No. We only say of a human being and what is like one that it thinks. We also say it of dolls and no doubt of spirits too. Look at the word 'to think' as a tool.<sup>42</sup>

Here Wittgenstein is still suggesting that to ascribe mental predicates to machines is plainly a conceptual confusion. However, he broadens the circle of entities with respect to which such ascription is felt to make sense. And by stressing that the word "to think" is but a *tool* he actually gives a new direction to the argument: for the application of tools can *change*. It seems, after all, conceivable that our use of words like "conscious", "intelligent", etc., should become *different*. — That our attitude towards machines might change is of course a topic Wittgenstein had little reason to speculate about; but he did reflect on the conditions under which mental predicates would, possibly, *cease* to be applied to human beings. "Can't I imagine", Wittgenstein asks,

that the people around me are automata, lack consciousness, even though they behave in the same way as usual? ... But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: 'The children over there are mere automata; all their liveliness is mere automatism.'<sup>43</sup>

Now in ordinary circumstances one would "find these words becoming quite meaningless".<sup>44</sup> However, there *are* conditions under which we produce utterances automatically as it were — e.g. when reporting observations.<sup>45</sup> And there are certain people, e.g. *mental defectives*,

42. *PI* I, §§281, 359f.

43. *Ibid.*, §420.

44. *Ibid.* — Cf. *PI* II, iv: " 'I believe that he is not an automaton', just like that, so far makes no sense. — My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul."<sup>3</sup>

45. "The *important* insight is that there is a language-game in which I produce information *automatically*, information which can be treated by other people quite as they treat non-automatic information — only here there will be no question of any 'lying' — information which I may receive like that of a third person." Ludwig Wittgenstein, *Remarks on the Philosophy of Psychology*, I §817, cf. §816.

with respect to whom one might easily feel "as if they talked more automatically than we do".<sup>46</sup> The reason for this feeling, Wittgenstein suggests, is ultimately a *sociological* one: mental defectives are, in important respects, not treated as members of our community; with the consequence that we do not regard them as capable of forming any society.<sup>47</sup>

In an age when computers are rapidly becoming an indispensable part of our *Lebenswelt*, Wittgenstein's argument can be reformulated in the following way. According to the rules of the language-game we normally play, mental predicates apply to human beings, to some extent also to animals, to ghosts if we believe in them, and to dolls and the like when we pretend they are alive. Such predicates do *not* apply to machines, except in a figurative sense. — Now if the environment changes in significant ways, our language-games can lose their point; they can become awkward, or even useless. It is indeed the case that "rule and empirical proposition merge into one another"; that certain events would "put me into a position in which I could not go on with the old game any further", would "completely alter the character of the language-game".<sup>48</sup> Let us assume that computers at some stage will become able to accomplish, in important areas and on a significant level, feats that would require intelligence if accomplished by people; and let us, further, assume that human attitudes and behaviour with respect to computers will at the same time undergo some relevant, appropriate modifications. One expects that under such conditions a change in the use — a shift in the meaning — of certain psychological terms would occur; that there would arise a tendency to ascribe to computers a kind of thinking, a sort of consciousness<sup>49</sup> — perhaps even some emotions and, possibly, pain.

46. *Ibid.*, I, §198.

47. "What would a society all of deaf people look like? Or a society of 'mental defectives'? An *important question!* What, that is, would a society be like, that never played a lot of our ordinary language-games?" (*Ibid.*, §957 — cf. *Zettel* §372: "We just don't see a *society* of such people.")

48. Ludwig Wittgenstein, *On Certainty*, §§309, 617, 646.

49. This is the argument of Otto Neumaier's outstanding paper "A Wittgensteinian View of Artificial Intelligence", in Born, ed., *Artificial Intelligence: The Case Against*. "For Wittgenstein", Neumaier writes, "the main reason why primarily human beings have the 'privilege' of 'having'

In fact even today the ascription of intelligence to computers seems to be, on a more prosaic level, a natural move. Discussing chess-playing computers, Daniel Dennett points out that their behaviour is practically inaccessible to prediction should the latter be based on acquaintance with their design or on knowledge of their actual physical state. "A man's best hope of defeating such a machine in a chess match", writes Dennett,

is to predict its responses by figuring out as best he can what the best or most rational move would be, given the rules and goals of chess. ... Put another way, when one can no longer hope to beat the machine by utilizing one's knowledge of physics or programming to anticipate its responses, one may still be able to avoid defeat by treating the machine rather like an intelligent human opponent.<sup>50</sup>

That is, one predicts the behaviour of the computer by ascribing to it "*the possession of certain information*" and supposing it "*to be directed by certain goals*"; and then "by working out the most reasonable or appropriate action" on the basis of these ascriptions and suppositions.<sup>51</sup> Now one might as well call the information possessed by the computer its "beliefs", and the goals and subgoals it has: "desires".<sup>52</sup> And Dennett stresses that "lingering doubts about whether the chess-playing computer *really* has beliefs and desires are misplaced", for all that has been said is that chess-playing computers are systems the behaviour of which can be explained and predicted by

---

mental states is that 'thinking', 'understanding', 'feeling pain' and other psychological phenomena are inseparably tied to human life and can only be grasped against this background... the ascription of psychological states is necessarily connected with the human *form of life*; already as children we learn to *see* some events as *signs* for human behavior as well as some specific behavior as sign for 'inner' states like 'thinking' or 'being in pain'. ... Wittgenstein in no case *opens* the door for the development of AI. Wittgenstein, however, *leaves open* the door for AI to become a meaningful enterprise; the conditions for this have to be clear, as well: the aims of AI can only be arrived at after a *radical change* in our form of life" (pp.151f., 158).

50. "Intentional Systems", in Daniel C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, Harvester Press, 1981, p.5.

51. *Ibid.*, p.6.

52. *Ibid.*

"*ascribing* beliefs and desires to them"<sup>53</sup>, i.e., in Dennett's terminology, by viewing them as "intentional systems".

The inescapable and interesting fact is that for the best chess-playing computers of today, intentional explanation and prediction of their behavior is not only common, but works when no other sort of prediction of their behavior is manageable. We do quite successfully treat these computers as intentional systems, and we do this independently of any considerations about what substance they are composed of, their origin, their position or lack of position in the community of moral agents, their consciousness or self-consciousness, or the determinacy or indeterminacy of their operations. The decision to adopt the strategy is pragmatic, and is not intrinsically right or wrong.<sup>54</sup>

The intelligence of even the best chess-playing computers is of course severely limited — viz. to playing chess, without the knowledge or understanding of anything that does not, in the strictest sense, belong to the game. A full understanding of chess, one would like to say, involves an understanding of the role chess plays in our life; it even involves an understanding of the *emotions* surrounding chess — and that means *acquaintance* with emotions. It is however difficult to see how acquaintance with emotions should be possible without actually knowing *what it is like to feel them* — i.e., without an experience of *having* them. "Only a being that is non-verbally active in the world", wrote Dennett in his early book *Content and Consciousness*, "could meet our requirements for understanding."<sup>55</sup> Programs whose "only

53. *Ibid.*, p.7.

54. *Ibid.*

55. London: Routledge & Kegan Paul, 1969, here quoted from the 1986 edition, p.182. "A computer whose only input and output was verbal", Dennett goes on to write, "would always be blind to the meaning of what was written. It might 'grasp' all the verbal connections, but it would lack 'acquaintance' with the things the words are about." Fed with a description of the Taj Mahal, the computer might well "respond with an output like 'The Taj Mahal must be very beautiful', but one wants the computer also to produce outputs like 'Take me there; I want to see for myself', and such outputs", Dennett stresses, "would be a hoax if the computer did not have some perceptual apparatus and many other sophisticated capacities." — Or, as Anthony Quinton puts it: "before we can say that machines think in the sense that really interests us we have to equip them with feelings, desires and emotion" ("Machines with Minds", London, BBC Publications, March 1983).

modes of perception and action are linguistic" — Dennett calls them "bedridden programs" — cannot achieve anything like real understanding.<sup>56</sup>

Alan Turing, probably the most important early figure in machine intelligence research, was certainly aware of this. In his opinion, "the 'sure' way of producing a thinking machine" would be to build a system that included "television cameras, microphones, loudspeakers, wheels and 'handling servo-mechanisms' as well as some sort of 'electronic brain' ", a system which, in order to "have a chance of finding things out for itself", should be allowed, as he put it, "to roam the countryside".<sup>57</sup> Turing of course realized that such a project,

56. *The Behavioral and Brain Sciences* 3 (1980), p.429. Such programs, Dennett here writes, "exhibit no 'language-entry' and 'language-exit' transitions, to use Wilfrid Sellars's terms". Language-entry and language-exit transitions are those moves belonging to a language-game which connect non-linguistic to linguistic entities. In my paper "No Place for Semantics" (*Foundations of Language* 7, 1971, pp.56-69) I have tried to show how the approach of Sellars, and of course that of Wittgenstein, makes the distinction between syntactic and semantic rules lose all its naturalness. I take this to be one of the many points where John Searle's ill-reputed Chinese room argument breaks down, cf. his "Minds, Brains, and Programs", *The Behavioral and Brain Sciences* 3 (1980), p.422. — Compare, incidentally, the point made by Y. Wilks in 1975. According to Wittgenstein, Wilks wrote, "it is dangerous to assess understanding other than in terms of actual and possible performances, and, if we take that to mean 'performances with language' we will see that it argues against one sort of criticism that AI workers have sometimes made of each other's systems: that they only 'appeared to understand' but 'didn't really do so'" (Yorick Wilks, "Philosophy of Language", in *Computational Semantics: An Introduction to Artificial Intelligence and Natural Language Comprehension*, eds. E. Charniak and Y. Wilks, Amsterdam: North-Holland, 1976, p.230). Cf. also Klaus C. Obermaier, "Wittgenstein on Language and Artificial Intelligence: The Chinese-Room Thought Experiment Revisited", *Synthese* 56 (1983), p.347: "In the case of a computer program, the understanding is based on the frame of reference which the speech community has with which the program interacts."

57. A.M. Turing, "Intelligent Machinery" (1947), in B. Meltzer - D. Michie, eds., *Machine Intelligence* 5, Edinburgh: Edinburgh University Press, 1969, p.13.

though attractive in principle, was, at the time, unattainable in practice. But he did propose to furnish his machine with “organs of sight, speech, and hearing”<sup>58</sup>; and he continued to emphasize the role of *machine learning* — educating a “child machine” — through “punishment” and “reward”, and indeed through “pain” and “pleasure”, as parts of the teaching process.<sup>59</sup> However, with the idea of ascribing pain to computers the question as to a radically new relationship between man and machine reappears — and indeed becomes inescapable.<sup>60</sup>

### *A New Form of Life*

The problem whether we can, or should, apply mental predicates to machines is, in the last analysis, an ethical one. If human-level machine intelligence is to be achieved, the computers envisaged have to be, so it seems, conscious and emotional creatures with respect to whom the question of rights and obligations cannot be easily dismissed.<sup>61</sup> And

58. *Ibid.*

59. *Ibid.*, p. 17, see also his “Computing Machinery and Intelligence”, *Mind*, Oct. 1950, pp.454ff.

60. Although some kind of a physiological theory of pain would be indispensable for building a machine capable of the appropriate sort of suffering, the issue is, clearly, a much broader one. As Dennett writes: “There can be no denying ... that our concept of pain is inextricably bound up with ... our ethical intuitions, our senses of suffering, obligation, and evil. It will not do to suppose that an assessment of any attempt at robot synthesis of pain can be conducted independently of questions about what our moral obligations to this robot might be. ... our concept of pain is not a pure psychological concept but also ethical, social, and parochial, so that whatever we *put inside* our computer or robot will not avail unless it brings in its train these other considerations...” (“Why You Can’t Make a Computer that Feels Pain”, in *Brainstorms*, pp.197f.)

61. “Artificial consciousness”, as Dan Lloyd puts it, “possesses the rights of natural consciousness, and our obligations toward it are what they are toward our fellows. ... Minds, artificial or otherwise, are entitled (*prima facie*) to continued consciousness, and entitled (*prima facie*) to the furtherance of whatever they undertake, provided their projects do not conflict with similar rights of others. Artificial minds give rise to a few special claims. Here the crucial fact is that these minds are *created*, by which fact the creator acquires

since it is a plausible hypothesis — or perhaps an obvious grammatical remark — that “differently embodied minds have different subjective values”<sup>62</sup>, this line of development appears to lead to a world in which, as Dan Lloyd puts it, “we are likely to preside over the creation of fundamentally alien beings”.<sup>63</sup> The conclusion is hard to avoid (and will certainly not affect the work of those who are involved in the relevant kind of research) that, for ethical reasons, the project of human-level machine intelligence should be abandoned.

Fascinating though this line of argument is, we can not pursue it here. Instead, we have to ask in what way our form of life — in the Wittgensteinian, basically epistemic, sense — will change, or is already changing, as a result of the more commonplace developments pertaining to the emergence of what has become called a computer culture. For even if computers remain essentially what they are today — information-processing machines of ever increasing speed and sophistication, housed in grey boxes — they are obviously altering our psychological language-game. As Sherry Turkle writes,

the machine ... enters into social life and psychological development, the computer ... affects the way we think, especially the way we think about ourselves. ... The question is not what will the computer be like in the future, but instead, what will *we* be like? What kind of people are we becoming?<sup>64</sup>

The computer is “rational, uniform, constrained by logic”, but it is also “an evocative object, an object that fascinates”;<sup>65</sup> it is a “partner in a great diversity of relationships”, “an expressive medium”, “a projection of part of the self, a mirror of the mind”.<sup>66</sup> But of course

computers are more than screens onto which personality is projected. They have already become a part of how a new generation is growing up. ... irrespective of the future of machine intelligence, computers are affecting

---

special responsibilities” (“Frankenstein’s Children: Artificial Intelligence and Human Value”, *Metaphilosophy* vol.16, no.4 [Oct. 1985], p.314).

62. *Ibid.*

63. *Ibid.*, p.317.

64. *The Second Self: Computers and the Human Spirit*, London: Granada, 1984, p.3.

65. *Ibid.*

66. *Ibid.*, p.5.

how today's children think, influencing how they construct such concepts as animate and inanimate, conscious and not conscious.<sup>67</sup>

Children growing up with computers are prone to ascribe some sort of consciousness to them. But it does not seem to be the case that such children will not develop an acute awareness of the differences between man and machine. Small children might "become biased toward seeing the machines as 'sort of' alive" if they are socialized in subcultures where it is considered bad style to "kill" machine processes, to "crash" or to interrupt programs.<sup>68</sup> Older children however develop sophisticated distinctions between the cognitive and the affective (computers "think", but they do not "feel"), between "free will and autonomy as opposed to programming and pre-determination"<sup>69</sup> and, in particular, between consciousness and *life*. Today's children, writes Turkle,

may be the first generation to grow up with such a radical split between the concepts of consciousness and life, the first generation to believe that human beings are not alone as aware intelligences. The child's splitting of consciousness and life may be a case where instead of thinking in terms of adult ideas 'filtering down' to children, it makes sense to think of children's resolutions prefiguring new positions for the computer culture to come.<sup>70</sup>

67. *Ibid.*, p.6.

68. *Ibid.*, p.52.

69. *Ibid.*, p.49. Observations like these ought to lessen the standard worry formulated, e.g., by Shanker: "if you institute a conceptual revolution in the concept of *thought* so that it henceforward becomes intelligible to describe mechanical operations as *thinking*, then conversely there seems little reason why the argument should not proceed in the opposite direction, thereby denying human beings the notions of autonomy and consciousness which underpin our conception of man as a *rule-following creature*. And if this conceptual revolution is allowed to proceed unhindered, the social and political consequences can surely not be far behind" ("The Nature of Philosophy", p.25). Incidentally, Wittgenstein's emphasis was *not* on "man as a rule-following creature" as should be clear from *PII*, §§198f., and from the subsequent text both of that book and of *On Certainty*.

70. *Op. cit.*, pp.50f. The expression "prefiguring" is clearly a reference to the work of the American anthropologist Margaret Mead who, in the early 1960s, has developed the terminology of "postfigurative", "cofigurative", and "prefigurative" cultures. Mead is using the words "*postfigurative*, when

One can get a glimpse of the age thus prefigured by looking at the culture in which, today already, the computer occupies a predominant position: the culture of virtuoso programmers, known as “hackers”. Although amongst hackers, too — naturally — there are people who, as Turkle puts it, are “terrified of being alone, yet afraid of intimacy”, and for whom the computer is “a companion without emotional demands”,<sup>71</sup> the majority of hackers, more importantly, form a *group*,

a group that has developed in conditions favorable to its growth as a culture: it shares a unity of place, lifestyle, and passion. It has developed its own rituals, language, myths, even its own literature. ... Hackers are often described in the media as people whose involvement with computers has drawn them away from involvement with other people; in fact they are drawn away from people who don't belong to their world, but within it they form a tight web of relationships where the computer is the center of an all-embracing way of life.<sup>72</sup>

It seems altogether implausible that the intertwined values of companionship and intellectual autonomy, or of moral solidarity and individual dignity, will be absent from the culture to come. However, certain other typical traits of our outlook are likely to change. In his book *Turing's Man: Western Culture in the Computer Age* J. David Bolter contrasts the Greek taste for “the linear, the superficial, the immediate, and the tactile in mathematics, art, and ethics” on the one hand with, on the other hand, the Western “fascination with depth, a desire to penetrate the surface of reality” and “the Christian pre-occupation with the human soul as something deep and mysterious

---

the future repeats the past, *configurative*, in which the present is guide to future expectations, and *prefigurative* for the kind of culture in which the elders have to learn from the children about experiences which they have never had. ... We are now entering a period, new in history, in which the young are taking on new authority in their prefigurative apprehension of the still unknown future.” (Margaret Mead, *Culture and Commitment*. Rev. and updated ed., New York: Columbia University Press, 1978, p.13.) That the element of prefiguration within any society can never be a *dominant* one was of course amply demonstrated by Wittgenstein — and recognized by an embittered Margaret Mead in the late 1970s.

71. *Op. cit.*, p.320.

72. *Ibid.*, pp.199f.

beneath the facade of human behavior".<sup>73</sup> Spengler, Bolter writes, was right when he "emphasized ... the fact that Faustian men had an appreciation for the idea of infinity, which the ancients did not".<sup>74</sup> And it is precisely the concern for depth and for infinity that "men of the computer age seem destined to lose".<sup>75</sup>

But this again has a very Wittgensteinian ring. Was not Wittgenstein anxious to exclude from the philosophy of mathematics any notion of the "infinite" that was not firmly rooted in the everyday use of that term?<sup>76</sup> Did he not strive to ban any idea of an "essence" *hidden* under the surface?<sup>77</sup> And did he not say that the *depth* philosophical problems seem to possess is an illusion — created by a misinterpretation of our linguistic forms?<sup>78</sup>

73. Chapel Hill: The University of North Carolina Press, 1984, p.217.

74. *Ibid.*, p.218.

75. *Ibid.*, p.220. Bolter quotes Marvin Minsky, a leading AI researcher to the effect that " 'intelligence' seems to denote little more than the complex of performances which we happen to respect, but do not understand. So it is, usually, with the question of 'depth' in mathematics. Once the proof of a theorem is really understood its content seems to become trivial. ... It may be so with *man*, as with *machine*, that, when we understand finally the structure and the program, the feeling of mystery ... will weaken", *ibid.*, p.221.

76. Cf., e.g., his *Remarks on the Foundations of Mathematics*, I, §60.

77. Cf., e.g., *PII*, §92.

78. *PII*, §111.