UDC: 007.52
Review article

LANGUAGE, COMPUTATION, AND INTELLIGENCE

Mario Radovan
University of Rijeka, Faculty of Philosophy, Rijeka, Croatia
E-mail: mradovan@oliver.efpu.hr

Theoretical discussions about the possibilities and limitations of Artificial Intelligence could have a major impact on the future development of computer technologies. However, such discussions are often based on vague concepts and assumptions. This paper proposes a conceptual framework for the appropriate presentation of knowledge and problems concerned with human cognitive abilities as well as the capabilities of computational machines. Within this context we hold it necessary to observe the distinction between authentic and functional cognitive abilities. Although computation is not a plausible way of gaining authentic intelligence, computational systems do offer virtually unlimited possibilities to replicate and surpass human cognitive abilities on the functional level.

Keywords: computation, subjectivity, understanding, thinking, intelligence, three worlds, care thesis, background thesis.

1. INTRODUCTION

Discussions about the possibilities and limitations of Artificial Intelligence (AI) can shed additional light on the plausibility of different approaches to computation, and with that, they can have "major consequences" for the future development of computer technologies [17, p. vi]. However, it seems that many well-known (and much discussed) arguments lack the clarity and precision needed to transcend this level of rhetoric and to become a necessary complement for empirical analyses and technological research. The victory of the chess-playing system *Deep Blue* over the world's chess champion offers an illustrative example of the tone which dominates in such arguments. The media presented the event as a *threat to human dignity*, as if the designers and creators of *Deep Blue* were not human beings. On the other hand, Searle offers a reassuring explanation: "The computer knows nothing of chess ... It just manipulates meaningless formal symbols according to the instructions we give it" [15, p. 59]. An explanation such as this of the machine's abilities implicitly assumes that a chess master's skill consists of something radically different than the ability to "manipulate meaningless symbols". But Searle gives no explanation about the nature of the chess master's "meaningful" way of reasoning (or computation) for his/her next move. Hence, it is not obvious why the claim stated for the computer - that it only "manipulates meaningless formal symbols" - could not be equally stated about the world chess champion himself.

Concepts such as *understanding* and *intelligence* - although much used in various arguments - do not have a precisely defined meaning. Hence, we argue that the discussion of the relationship between human cognition and machine computation should start with the analysis of the genesis and essence of the *human* understanding and intelligence. Namely, the basic questions that we must face are: Where does understanding *come from*, and what does it *mean* to understand and to act intelligently? In Section 2 we deal with the problem of *subjectivity* which is the central problem at the intersection of the various fields of scientific interest. In Section 3 we propose the idea of the *three worlds - the physical world, the world of subjective states*, and *the world created by humans* - which offers a conceptual framework in which the problems concerning cognition and computation can be expressed and discussed in an appropriate way. In Section 4 we discuss the basic problems and limitations of AI in terms of this three-world ontology. In this context, it is necessary to distinguish between *authentic* and *functional* cognitive abilities; we claim that authentic intelligence is out of the reach of purely computational systems, though on the level of functional intelligence, the performances of computer systems shall be more and more difficult for humans to keep up with. It must be noted that we will not discuss the *symbolic* and the *connectionist approach* to AI separately, since we believe that both approaches face the same basic problems (see e.g. [3], [11] and [12]).

2. THE PROBLEM OF SUBJECTIVITY

The standard approach to science assumes that reality is *objective* in the sense that neither its existence nor its structure depend on the viewpoint of a particular observer. Scientific knowledge deals with things which could exist without being known, and is expressed in a way which is equally accessible to every (sufficiently qualified) human being. We can say that science speaks of phenomena from the *neutral* (or third-person's) point of view. On the other hand, subjective mental states seem to be a different kind of phenomena in an ontological and epistemic sense. Namely, mental states can only exist as *someone's* state, and they are really known only to the subject whose states they are. Neural activities that manifest themselves as pain could (in principle) exist and be observed without being experienced, but the *pain itself* could not: it exists only *if* and *as* experienced. Nagel claims that the subjective is "an irreducible feature of reality", and that "it must occupy as fundamental a place in any credible world view as matter, energy, space, time and numbers" [7, pp. 7-8]. Churchland stresses that the taxonomy of physical science has some limits, and that it reaches these limits "at the subjective character of the contents of consciousness" [2, p. 196]. That does not mean that mental states must be something non-physical, but only that the present scientific taxonomy *cannot express* what these states are *like* from the unique perspective of the subject that has them. And as *subjective* states, they do *not have* any other way of existence. In this context, positions about the nature of mental states are divided into two basic views: *physicalism* and *property dualism*. There are a few variants within each of these two basic views, but none of them really solves the problem. There are also authors who reject both views, but do not offer a coherent view of their own.

2.1. Physicalism and property dualism

Physicalism claims that mental states can be ontologically reduced to physical states of the brain, and thus expressed in the objective (third-person) language of science - firstly, in the language of neurology, and then in the language of physics. Copeland admits that physicalism is "supported only by faith", or more precisely, by the fact that other theories seem less plausible or at least less "comfortable". Copeland says, "an irreducibly mental dimension that sticks out of an otherwise physical universe like a sore thumb - is unpalatable to us. It offends against our expectation that nature is a harmonious, integrated affair" [3, p. 179]. In essence, physicalism excludes the subjective from the scope of the discussion, but in doing so it does not solve the problem of how to deal with the "offensive" fact that the universe *does* contain subjective states of people like you and me. On the other hand, property dualism claims that subjective mental states can be neither ontologically reduced to the physical nor expressed by the taxonomy of physics. To know everything about the physical processes going on in the brain does not mean that you know anything about the *experience* of pain or joy, or about the *taste* of a lemon. However, by accepting the irreducibility of conscious mental states, property dualism faces the (seemingly unsolvable) problem of the *relationship* between the physical and the mental "dimensions" of reality.

Physicalism does not solve the problem of the subjective states by their reductive elimination; and property dualism does not solve the problem either, since it cannot explain the relationship between subjective states and their objective sources. To solve the problem of the relationship between the mental and the physical we need a radically new understanding of the basic phenomena in the world, and a radically new taxonomy. Penrose claims that there is simply no room for conscious mental states within the current scientific world-picture. He holds that "we must look to a new physics" which would make it possible to find a scientific explanation of consciousness [9, p. 183]. Penrose makes such an attempt by trying to discern the source of the mental at some physical level lower than the neural level. However, his attempt remains within the realm of physicalism. Shimony claims that a new theory of mental phenomena should attribute "mentalistic properties" to the *most primitive* entities in the universe: that is, to a kind of *monads*, whatever they may be [16, p. 153]; Chalmers [1] also speculates along the same lines. However, the basic problem of how to speak about the subjective dimension of reality by means of the objective language of science remains unresolved. Hence, the primary question is not whether machines can think – and therefore *be*; or, whether *to be* means to be *a computer* - but is instead the question of *how to speak about subjective phenomena in a scientific fashion at all, since the scientific taxonomy is intrinsically objective*. The mystery of the subjective dimension of reality is primarily of a conceptual nature, and we do not really know how to solve it.

2.2. Searle and the mind-body problem

Searle claims that a taxonomy which speaks of mental and physical properties as radically different kinds of properties is wrong and misleading. He claims that subjective states are ontologically irreducible to brain states (as property dualists do), but he insists that his view is not a property dualism. He claims that there are *a lot* of properties in the world, and that it is a mistake to count them, because "we live in one world, not two or three or twenty-seven" [15, p. 88]. However, Searle's "one world" is not the one of physicalists, since he insists that conscious mental states are *not* ontologically reducible to physical states of the brain. It seems that Searle believes that his one-world view can avoid the difficulties of both views, i.e. of physicalism and of property dualism. However, it seems to me that his position is incoherent.

Searle claims that "consciousness has a first-person or subjective ontology and so cannot be reduced to anything that has third-person or objective ontology". He says that it is not possible to reduce "feelings to neuron firings" or "neuron firings to feelings", because in the first case we would leave out the subjectivity and in the second case the objectivity of the phenomenon [15, p. 212]. However, Searle holds that such an ontological irreducibility does not mean that there is some "really important" difference between the properties of this one world. Moreover, he "hopes" that he has "made it clear" that we can accept the "irreducibility of consciousness" without accepting that there are "two ontologically different sorts of realms", or at least "two different sorts of properties in the world" [15, p. 195]. In short, Searle's peculiar position can be reduced to the following: (1) "consciousness has a first-person ontology", and (2) it "cannot be reduced to anything that has third-person ontology"; however, Searle cannot accept (3) that there are "two different sorts of properties in the world", because he would then satisfy the definition of property dualism. And that is exactly what he does not want to do.

Searle justifies his position by the fact that consciousness - although irreducible to biology - nevertheless *is* a biological (i.e. physical) phenomenon: hence, its ontological irreducibility, by itself, need not imply that consciousness is a property of a "different sort". Searle says: "consciousness is a mental, and therefore physical, property of the brain in the sense in which liquidity is a property of systems of molecules" [13, p. 14]. Of course, with such definitions ("mental, and therefore physical"), everything can be made of the same "sort". But it is not clear what such a *solution by definition* could accomplish, since it does not solve but only hides the problem of the relationship between the mental and the physical. Indeed, the perennial problem of the relationship between the properties of "different sorts" for property dualists, for Searle became the problem of the relationship between the two *subclasses* of properties of the same sort. But the problem is the same. Let us say here, that it is not formally inconsistent to proclaim that all the phenomena of the world are of the same "sort". However, if *mutual ontological irreducibility* between disjunctive subclasses of phenomena is not enough to say that these phenomena are of "different sorts", then one must ask what such an ontology is good for.

Finally, Searle does accept that there is some "really important" difference between the entities of the one world. However, he holds that the really important difference is not the one between the mental and the physical but the one between those entities which exist *independently of the observers* and those which are *dependent on observers*. Observer-relative entities, such as social institutions, functions and values, depend on consciousness for their existence, because they are created by humans and they exist only if recognised by humans. Consciousness itself is not observer-relative, since it is "a real and intrinsic feature of certain biological systems" [15, p. 211]. However, although Searle emphasises this "really important" ontological difference between the two classes of phenomena, he nevertheless sticks to his one-world view, in which everything is of the same (physical) "sort". For example, he says: "There is no physical-chemical description adequate to define 'restaurant', 'waiter', 'sentence of French', 'money', or even 'chair' and 'table', even though all restaurants, waiters, sentences of French, money, and chairs and tables are physical phenomena" [14, p. 3]. In short, everything is physical but it cannot be described adequately by the language of physics. Searle says he "dislikes the old Cartesian terminology", and hence, instead of the old terminology, introduces the distinction between "*nonmental brute physical facts* and *mental facts*" ("mental" intended here as *physical*, of course). However, he admits that these two basic categories may not "exhaust all the kinds of facts"; namely, "if there are mathematical facts, for example, they would not be included in this taxonomy" [14, p. 122]. In Section 3, we discuss a *three-world* ontological framework, which does not solve the problem of subjectivity itself, but does allow a more appropriate way to speak about subjective mental states as well as about the facts of mathematics.

3. THE THREE-WORLD FRAMEWORK

A minimal requirement which an ontology should satisfy is to provide a conceptual framework within which we can speak appropriately about known (or, in principle, verifiable) phenomena. In this section we argue that an adoption of the three-world ontological framework makes it possible to express and discuss the problems of human cognition and computation in the most appropriate way. This ontological framework could be regarded as both realistic and conceptual, since there are factual and cognitive reasons for the assumption that there exist three basic and mutually irreducible kinds of phenomena.

We propose to classify all phenomena into three ontologically different realms or worlds: *the physical world*, *the world of subjective states*, and *the world created by humans*. The idea of the "three worlds" is attributed to Popper, although he says that this terminology "stems from Frege" [10, p. 105]. Popper's basic ideas about the three worlds are clear, but he treats many problems only scantily or unsuitably, so they must be considered open. We assume that to the physical world (*world1*) belong all natural phenomena up to - but not including - the phenomena of consciousness as a boundary case. Although consciousness is a natural phenomenon, it is that peculiar and unique natural phenomenon which at the same time also comprises a new world for itself

(*world2*): the world of subjective states, such as pain, desire, love or anger. The world created by humans (*world3*) contains everything that has been created by conscious human beings (although not all the entities from *world3* have to be created consciously). *World3* contains abstract entities such as numbers and theorems as well as *all forms* created by humans. Forms have an autonomous existence; however, they are not eternal (in the Platonic sense), but are created by humans and (often) imposed on the physical world. For example, a desk conceived of as a piece of wood belongs to the physical world; however, if it is conceived of as a *desk*, in a functional or aesthetic sense, it is a creation of *world2* (the conscious mind) and belongs to *world3*, the world created by humans. The same holds for a computer: the computer on my desk is a physical object, but what makes that "heap of atoms" on my desk a computer is the form imposed on it by humans.

In the context of the unsolved problem of subjectivity, the question of the relationships between the three worlds does not have a simple solution. According to Popper, consciousness "plays the main role in the causal chains" that lead from *world3* to *world1*. He says: "The abstract part of *world3* has so far never exerted a *direct* influence upon *world1*. ... The link is always forged by consciousness, by *world2* (Perhaps this will be different one day.)" [15, p. 24]. Popper's arguments are not precise enough to be simply accepted or to allow a clear confutation. In any case, the tree-world ontological framework does not, by itself, solve the problem of "links" between the three worlds, nor does it seem that this problem has some obvious solution. We hold that entities of *world3 cannot* have a "direct influence" on *world1* since "the abstract" cannot do anything by itself. Furthermore, Popper does not pay enough attention to the problem of delineation between the three worlds; he says: "reality consists of three worlds, which are interconnected and act upon each other *in some way*, and also *partially overlap* each other" [10, p. 8; my italic]. A lack of precision such as this creates various difficulties for him. Penrose is one of the few authors that accepts the three worlds ontological frame; he interprets the basic idea in a peculiar way, but he does not solve any of the problems which are immanent to such an ontological framework. He holds that the three worlds are mutually "profoundly dependent", but that the connections between them are "mysterious": "There is something distinctly mysterious about the way that these three worlds inter-relate with one another" [9, p. 139]. But instead of dealing with the drawbacks of Popper's and Penrose's versions of the idea of these three worlds, let us express our basic positions in a more precise way.

Firstly, we assume that the three worlds are disjunctive; they do not overlap, nor are they mutually ontologically reducible. Furthermore, we assume that the realm of the physical is *causally closed*; and we know no other causal relations than the physical ones. Hence, it is possible to speak coherently of causal relations only on the level of the physical world. It seems that such a position implies that, for example, music cannot affect my physical behaviour; or perhaps that my conscious deliberation about my next move cannot have any effect on my choice. However, the proposed position does not imply such consequences; all it claims is that *causal* effects take part only in the *physical* world. Music can have an impact on behaviour, but only when instantiated

by means of the physical world, and only through the physical impact of its instantiation on the neural system (through listening to it). Being conscious of some situation does impact a decision; however, the conscious state *itself* emerges from some neural processes, which were triggered and are influenced only by physical causes. And from these processes new mental states are constantly emerging. Therefore, all causal impacts take part inside *world1*.

3.1. The role of consciousness

The proposed three-world ontological/conceptual framework can be considered as an extension of the *epiphenomenalism.* According to epiphenomenalism, mental states are products and manifestations of brain activities, but mental states (as subjective phenomena) do not have any causal impact on these activities. There are claims that whether epiphenomenalism is true or not is an empirical question, although "it is notoriously hard to think of decisive tests for or against it" [5, p. 136]. Flanagan discusses the notorious experiments of Benjamin Libet, which show that there are electrical activities taking place in the brain about a second before the time when the subject believes he or she made a conscious decision to perform an act. In other words, the specific neural activities which lead to an act notably *precede* the agent's being *conscious* of the fact that he or she was going to perform that act. In the context of these experiments, "if you imagine that consciousness is something which does something", says Penrose, "then you are presented with almost a paradox" [9, p. 135]. However, Flanagan insists that it is "extremely implausible" that subjective awareness "plays no significant causal role", although he admits that it is hard to say "exactly what role it plays". Flanagan concludes that all disputes about epiphenomenalism "are matters to be settled in empirical court" [5, p. 151]. However, we believe that before entering the "empirical court" we should know exactly what we are contesting. Namely, the current scientific taxonomy does not allow us to speak about any "causal role" of an entity that is not physical. Hence, if conscious mental states are not reducible to anything physical (e.g. to brain states), then no "empirical court" can say anything substantial about the "causal role" of these conscious mental states.

The position advocated here claims that a conscious mental state (as subjective and irreducible to the physical) cannot be said to cause anything by itself; what causes something is the *brain state* whose features and manifestation are the conscious state itself. Therefore, when we say that a conscious mind "creates" or that it "imposes a structure", we wish to say that the causal impacts take place on a physical level. In other words, it is not consciousness as a subjective state that creates by itself, but the physical system *must be* conscious to be *able* to fear, desire, think, understand and create.

3.2. Discovering and creating

Creation is a result of the human capacity for abstraction and generalisation (which is based on observed phenomena from *world1*), as well as of the attempts to express and articulate specific subjective states. We claim that scientific theories are creations of the human mind; we create theories, and we then evaluate them on the basis of the

effects we manage to obtain from these theories. To *world3* belong not only mathematics, science, and technological devices, but everything that has been created by humans by means of some *system of representation*, as well as these systems of representation themselves. Examples of these types of things are social structures, customs, works of art and social symbols, which are created by humans and exist on the basis of some system of representation. We claim that entities of *world3* cannot create by themselves; when instantiated by *world1*, they only explicate that which has already been implicitly created by their very creation. For example, an implemented program can produce symbols, shapes, colours or sounds; but its outputs were implicitly created by the very creation (and implementation) of the program itself, although many of the outputs may not be created with a specific intention or a purpose.

Contrary to this position, Penrose holds that *world3*, or at least mathematics, is not a human creation, but exists by itself; consequently, we can only *discover* entities of this world and not create them. Furthermore, Penrose conceives of the physical world as something "emerging out of the ('timeless') world of mathematics" [9, p. 2]. Concerning scientific theories, Penrose says that "Einstein revealed something that was there", because "the mathematical structure is just there in Nature, the theory really is out there in space". But Penrose also claims that Einstein's discovery of the General Theory of Relativity "was not motivated by any observational need but by various aesthetic, geometric and physical desiderata" [9, p. 25]. Hence, it seems that Penrose's position is not coherent, unless he presupposes that there is some strange coincidence between Einstein's aesthetic "desiderata" and the "structure" of the physical world.

Popper holds that the two opposite positions can be reconciled, since "they are both right". He claims that while, for example, the infinite series of natural numbers is our creation, the problems concerning prime numbers are not our creation: we discover such problems in an objective world which we have created, but which became "objectified, detached from their creators and independent of their will" [10, p. 26]. We hold that Popper's attempt at reconciling the two positions was not successful. Created entities also have features which were not created on purpose, which are often not obvious, and which could be discovered later, or perhaps will remain undiscovered forever. But all those features which are intrinsic to some created entity were created with the very creation of the entity itself. Hence, if we accept, as Popper does, that the infinite series of natural numbers is a human creation, then we should also accept that all features of the series, as well as all the possible problems concerned with it, are human creations. For nothing would be a problem if nobody tried or *desired* to solve it.

4. UNDERSTANDING AS PROBLEM SOLVING

The question of the relation between *functional capacities* of computational systems and *authentic features* of human beings cannot be resolved in terms of outside effects, but we must take into account internal motivations, since no behaviour can be said to be *intelligent* without taking into account its *motivations*. Popper stresses that all living beings are constantly preoccupied with solving problems; they are trying to improve their situation, or at least to avoid its deterioration. Popper claims that

consciousness was a *problem-solving feature* of the living system right from the start: "the original task of consciousness was to anticipate success and failure in problem-solving and to signal to the organism in the form of pleasure and pain whether it was on the right or wrong path to the solution of the problem" [10, p. 17]. Conscious beings encounter and create: they encounter success and failure, satisfaction and frustration, joy and sorrow; they create myths, religions, arts, and sciences in order to explain, resolve or celebrate the mystery of existence. A conscious being creates in order to reduce fears, produce pleasure, communicate feelings, overcome loneliness and obtain wholeness. It is *internal tension* that triggers acting and creating - the tension which springs from the subject's awareness of his own finitude, loneliness, anxiety and meaninglessness, as well as from the desire for integration, meaning, and harmony [8, p. 171]. Authentic intelligence springs from the *internal necessity of the system* to solve the problems it encounters; it has been driven by the impulses to avoid the painful and achieve happiness.

Hawking claims that consciousness is not something that can be measured "from the outside"; hence, he holds, we should leave consciousness aside, and deal with the intelligence itself, "which is a quality that can be measured from the outside" [6, p. 171]. It really does seem more feasible to "measure" behaviour rather than a mental state. However, intelligent behaviour without awareness, observed "from the outside", is essentially different from intelligence that is aware. The existence of intelligence without awareness depends - in an ontological and epistemic sense - on intelligence that *is* aware, as the only *source and measure* of every other quality. Because of these very reasons, we cannot speak of the creativity of intelligence, since such intelligence does not and cannot have any motivation to create: it does not have any aim or any *criterion* on the basis of which it could evaluate its own behaviour *as* creating.

4.1. The Care thesis

The motivational approach to cognition that we propose, argues that thoughts and intelligent behaviour are not independent of subjective states (moods), neither in the generative nor in the evaluative sense. A mood opens a specific way for phenomena to show themselves, and along with that it opens different ways of *understanding* these phenomena. This means that aware-less cognitive abilities are qualitatively limited and that they exist only as interpretations made by an aware cognitive system. There are many successful expert systems, and we argue that there is no obvious limits to the further improvement of machine abilities and skills within the realm of functional replication of human cognitive abilities. However, aware-less machines cannot reach authentic intelligence since for such systems there exist neither the problems nor the motivation to create. Indeed, we could say that the cognitive abilities of artificial (aware-less) systems are limited by the fact that they *do not have problems*. Artificial intelligence is a human creation, and as such it belongs to *world3*. On the other hand, human intelligence is a feature of *world2*: a feature which could neither exist nor be evaluated without the existence of the *other* features of *world2*, such as desire, fear, love and anger. Let us sum up the above in the following *Care thesis*: Human beings, as aware systems, are essentially determined by anxiety and desires: by *care*; human

cognitive abilities spring from care and are shaped by it, and they cannot be fully replicated by a *care-less* (aware-less) system. On the other hand, computers, as symbol manipulating devices, can be said (although not proved) to be *care-less* systems; hence, such systems cannot understand, be intelligent and creative in the sense in which humans are. There are claims that to reach human-like intelligence, an artificial system should have a complete model of the world. The notion 'complete model of the world' is rather vague; however, as long as a system does not have any *problem* (care), no purely descriptive model of the world seems to be sufficient for authentic understanding and intelligence.

4.2. The Background thesis

The Care thesis says that a care-less system cannot understand, and be authentically intelligent and creative. In this context, the question arises, what are the limits (if any) of the *functional* replication of human cognitive abilities? It is generally thought that for successful functional replication of human cognitive abilities, a system should have a substantial amount of *common-sense knowledge*. But Copeland emphasises that the development of a system with common-sense knowledge faces such huge difficulties on the level of ontology, epistemology, knowledge organisation, and inference that such an enterprise seems to be "the hardest that computer science has ever faced" [3, p. 107]. Copeland does not claim that such a goal could not be reached at all, but he maintains that the idea of creating a system with common-sense knowledge is not at present within our grasp.

Concerning the possibility of a formal description of human knowledge and behaviour, Copeland says that it is "certainly true" that his preparing an omelette "can be *described* by means of if-then sentences". However, he emphasises that from this fact it does not follow that his actions are *produced* by some device in the brain "scanning through lists of if-then rules of this sort" [3, p. 101]. But that is not relevant for our purpose, or at the very least not decisive enough. Namely, a behaviour which can be *described* in a computable way, *can* be replicated by computation, regardless of how the *original* behaviour was produced. Copeland's criticism of the *CYC* project is based primarily on the immense complexity of such an enterprise which aims to build a knowledge base with relevant (or complete) human common-sense knowledge. However, sceptical comments concerning the possibility of building such a knowledge base are based primarily on the *Background thesis* which claims that it is *not possible at all* to give a full formal description (e.g. in terms of if-then rules) of a common-sense capacity such as the preparation of an omelette. Computers, as symbol manipulating systems, can acquire only that knowledge which *can* be expressed in some symbolic language. (Let us mention here that *connectionist systems* are mainly *software simulations* implemented on digital/symbolic computers.) Hence, if the Background thesis is true, then computers could neither reach authentic intelligence (as the Care thesis says) nor replicate an essential part of the common-sense human behaviour on a functional level.

The Background thesis says that a sentence (or a phenomenon) can assume

meaning only in relation to some *background* (context) in which it appears, and in relation to some *background capacities* of the subject. A meaning is an *implicit interpretation* (mainly unconscious) of perceptions, concepts, sentences, beliefs, desires, and experiences. Such an interpretation takes place within some background or other and on the basis of a set of background capacities, which are not themselves explicitly interpreted. In other words, the Background thesis says that *represented* knowledge can exist and function only in the context of some background capacities which are not and *cannot* themselves be explicitly represented. Heidegger calls these (implicit) capacities the "understanding of being", and Dreyfus says that "only if we draw on the background familiarity that is not in the mind but in the shared practices can we understand how assertions can pick out objects" [4, p. 269].

The Background thesis implies that there is no way to express (and represent) human common-sense knowledge by means of a computational system, because such a system does not have the background capacities on the basis of which it could *interpret* the inputs and act (behave) in an appropriate way. However, we believe that the arguments in favour of the Background thesis are not conclusive, and that the common-sense knowledge problem, even if the most difficult one, should be considered open. So, although it would be very difficult to "teach" a machine *all* the background knowledge and capacities which humans need and use (mainly unconsciously) in the course of an action such as "having dinner in a restaurant", it seems that *each individual "atom"* of the background knowledge and capacities which are needed in such a situation, *can* be described in a computable form and then also replicated. Such descriptions could either take the form of sentences (facts and rules), or artificially realised mechanical skills, or simply "demonstratives" (e.g., "*that* is a glass"). Demonstrative knowledge could be based on a set of patterns for all the things which humans encounter in a given situation. Such patterns could be stored in a figurative form in the knowledge base of the system. Our background knowledge and capacities are a kind of unconscious map of our surroundings and skills. It would be difficult to create such a background map of human expectations and skills *for* something and *with* something so different from humans as computers are. However, we hold that there is no obvious reason to believe that something like this would be impossible. After all, our background knowledge and capacities were *learned* mainly by the method of *trial and error*. And our learning starts with some (rather humble) predisposition, but with nearly no knowledge or skills.

Arguing for the Background thesis, Searle says that the literal meaning of the sentence "She gave him her key and he opened the door" does not "block" the following interpretation of the sentence: "He opened the door with her key by bashing the door down with the key; the key weighted two hundred pounds and was in the shape of an axe" [14, p. 131]. It is true that such an interpretation of the sentence cannot be "blocked" by the literal content of the sentence, but only by the fact that we have "a certain sort of knowledge about how the world works" and "a certain set of abilities for coping with the world" [14, p. 131]. But we claim that there is no evidence that such abilities could not be simulated by an artificial system. After all, why do *we* ourselves not expect somebody to open the door with the key "by bashing the door

down with the key"? Probably only because we *learned* that keys are usually used in a different way. Finally, the fact that humans know how to behave in a given situation without consciously following some explicit rules of behaviour, does not mean that human behaviour could not be *described* by a set of rules and replicated by an artificial system. And such a system should then be able to "cope" in the given situation in the same way humans do, at least at the level of behaviour or functioning.

5. SUMMARY

The problem of the subjective dimension of reality is primarily a conceptual problem. We do not know how to express (represent) subjective phenomena in a scientific fashion, as scientific taxonomy is intrinsically objective. None of the present positions about the ontological status of conscious mental states can really solve this problem. In this context, we argue that the three-world ontological framework offers a basic conceptual framework in which the problems concerned with cognition and computation can be expressed and discussed in the most appropriate way, although the solutions to some of these problems seem to lie beyond the horizon of our current knowledge. Discussions concerning intelligence and understanding are usually limited to the level of behaviour or functioning of the system; contrary to that, we claim that these phenomena should be observed and defined in terms of their origins and motivations. Human authentic cognitive abilities originate from care and are permanently shaped by it. Hence, these abilities cannot be fully replicated by care-less (aware-less) systems. On the other hand, the arguments in favour of the Background thesis are not conclusive, and there is no obvious limit to further improvements of machine capacities within the realm of functional replication of human cognitive abilities. There is actually no reason to expect that computational systems could acquire some new qualities which would be radically different from those that such systems have had from their very beginning. However, we must expect that computational systems will continue to exceed human functional capacities in various new fields of human activity. In this context, the claims that machine capacities are "meaningless" tell us rather little as well as the claims that these capacities represent a "threat to human dignity".

REFERENCES

[1] J. D. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, Oxford, 1996.

[2] P. M. Churchland. *The Engine of Reason, the Seat of the Soul*. The MIT Press, Cambridge, Massachusetts, 1995.

[3] J. Copeland. *Artificial Intelligence: A Philosophical Introduction*. Blackwell, Oxford, 1993.

[4] H. L. Dreyfus. *Being-in-the-World: A Commentary on Heidegger's 'Being and Time', Division I*. The MIT Press, Cambridge, Massachusetts, 1991.

[5] O. Flanagan. *Consciousness Reconsidered.* The MIT Press, Cambridge, Massachusetts, 1992.

[6] S. Hawking. The objection of an unashamed reductionist. In R. Penrose et al. *The Large, the Small and the Human Mind.* Cambridge University Press, Cambridge, 1997, pp. 169-172.

[7] T. Nagel. *The View from Nowhere.* Oxford University Press, Oxford, 1986.

[8] T. Organ. *The Self in Its Worlds: East and West.* Associated University Press, London and Toronto, 1987.

[9] R. Penrose et al. *The Large, the Small and the Human Mind.* Cambridge University Press, Cambridge, 1997, pp. 1-143 and 173-185.

[10] K. R. Popper. *In Search of a Better World.* Routledge, London, 1992.

[11] M. Radovan. Intelligent systems: Approaches and limitations. *Informatica*, Vol. 20, No. 3, 1996, pp. 319-330.

[12] M. Radovan. Computation and understanding. In M. Gams, M. Paprzycki, and X. Wu, eds. *Mind Versus Computer.* IOS Press / Ohmsha Ltd., Amsterdam and Tokyo, 1997, pp. 211-223.

[13] J. R. Searle. *The Rediscovery of the Mind.* The MIT Press, Cambridge, Massachusetts, 1992.

[14] J. R. Searle. *The Construction of Social Reality.* The Free Press, New York, 1995.

[15] J. R. Searle. *The Mystery of Consciousness.* The New York Review of Books, New York, 1997.

[16] A. Shimony. On mentality, quantum mechanic and the actualisation of potentialities. In R. Penrose et al. *The Large, the Small and the Human Mind.* Cambridge University Press, Cambridge, 1997, pp. 144-160.

[17] T. Winograd. Preface. In M. Gams, M. Paprzycki, and X. Wu, eds. *Mind Versus Computer.* IOS Press / Ohmsha Ltd., Amsterdam and Tokyo, 1997, pp. v-vii.

Received: 15 April 1999
Accepted: 30 November 1999

Mario Radovan

JEZIK, RAČUNANJE I INTELIGENCIJA

Sažetak

Teorijske rasprave o mogućnostima i ograničenjima umjetne inteligencije mogle bi imati bitan utjecaj na budući razvoj računalnih tehnologija. Međutim, takve rasprave često se zasnivaju na nejasno definiranim pojmovima i pretpostavkama. Članak predlaže jedan konceptualni okvir koji otvara mogućnosti za primjereno predstavljanje znanja i problema vezanih uz čovjekove kognitivne sposobnosti, kao i mogućih dosega računalnih strojeva. U tom kontekstu držimo

nužnim razlikovati autentične i funkcionalne kognitivne sposobnosti. Iako samo računanje ne izgleda kao put koji bi mogao dovesti do autentične inteligencije, računalni strojevi pružaju praktički neograničene mogućnosti reproduciranja i nadmašivanja ljudskih kognitivnih sposobnosti na funkcionalnoj razini.

Ključne riječi: računanje, subjektivnost, razumijevanje, mišljenje, inteligencija, tri svijeta, teza brige, teza pozadine/okruženja.